

HORIZON

**A Collection of Research Based Articles on
Multi-Disciplinary Science**



**Chief Editor
Nilim Jyoti Senapati**

**Editor
Satyajyoti Gogoi**



Tinsukia College Publication

HORIZON

A Collection of Research Based Articles on Multi-Disciplinary Science



Chief Editor

Nilim Jyoti Senapati

Editor

Satyajyoti Gogoi



Tinsukia College Publication

HORIZON

A Collection of Research Based Articles on Multi-Disciplinary Science

First Published in 2023

Published by : Tinsukia College
Tinsukia-786125, Assam, India
Website: <https://tinsukiacollege.in>

© Tinsukia College

Publication, distribution and promotion rights reserved by the Tinsukia College
Publication.

No part of this publication may be reproduced or transmitted in any form or by any
means without the prior written permission of the Publisher.

*Any plagiarism related issue will not be the concern of editorial board. The authors are
solely responsible for this.*

Printed at : The Assam Computers
Sector A1, 1st Bylane, Industrial Estate,
Borguri, Tinsukia-786125, Assam, India.

ISBN : 978-81-19322-07-7

Birds Photography by Dr. S. J. Gogoi
Location: Ramapur, Dibrugarh-786008

Price : Rs.350.00
(Rupees Three Hundred and Fifty only)

HORIZON

A Collection of Research Based Articles on Multi-Disciplinary Science

ISBN : 978-81-19322-07-7



Editorial Board

Chief Advisor : Dr.Surjya Chutia, M.A.,Ph.D.
Principal, Tinsukia College

Advisors : Dr.Bulbul Gogoi, M.Sc.,Ph.D., Coordinator, IQAC.
Rajen Saikia, M.Sc.
Dr.Ratna Gopal Gogoi, M.A.,Ph.D.
Dr.Rupali Gogoi, M.Sc.,Ph.D.

Chief Editor : Dr.Nilim Jyoti Senapati, M.A., Ph.D.

Editor : Dr.Satyajyoti Gogoi, M.Sc.,M.Phil.,Ph.D.

Members : Dr.Indira Baruah, M.Sc.,Ph.D.
Dr.Saurav Bhattacharjee, M.Sc.,Ph.D.
Dr.Daimalu Baro, M.Sc.Ph.D.,DISE.

HORIZON

ASSAM, INDIA

LIBRARY



1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57

1956-57





Foreword

The process of research is a testament to human curiosity, resilience, and determination. It involves hours of meticulous investigation, critical thinking, and the courage to challenge established norms. In an era where information flows abundantly like a vast river, discerning wisdom has become a rare gem nestled within the stream. As we navigate through a sea of data, facts, and opinions, the ability to extract true wisdom from this deluge has emerged as a distinct and precious skill.

In this regard, I am delighted to introduce this multi-disciplinary research based scientific articles edited volume '*Horizon*', published by the Tinsukia college, a evidence to the intellectual prowess and scholarly dedication that thrive within the academic community. As we embark on a journey through the pages of this research-based compendium, we delve into a world of insights and innovative ideas that reflect the very essence of our commitment to advancing knowledge.

For this, I extend my heartfelt appreciation to the editors and contributors who have diligently curated and crafted this volume. To the readers, I invite you to delve into the richness of this volume and to contemplate the implications of the research presented here. May these pages spark new ideas, fuel intellectual discourse, and foster a sense of intellectual exploration at the boundless possibilities that the pursuit of knowledge offers.

Surjya Chutia
Principal
Tinsukia College
01/09/2023



Preamble from IQAC desk

The purpose of doing research in undergraduate educational institution is to improve mainly the teaching and learning process by taking a collaborative, reflective and problem solving approach of investigation. It can also help to find out the solutions on problems arising in the classrooms connecting to the sources of informations, thereby enhance society by advancing knowledge, through the development of scientific concepts and ideas. The research works are not always towards the new discoveries or inventions but it may be the new interpretations of previous observations through careful consideration regarding to particular problem using scientific standpoint. As a part of the noble approach of the publication of newly edited research book entitled "*Horizon*" from Tinsukia College Publishing, Tinsukia encompassing a wide range of scientific research articles of micro to macro world will definitely provide sufficient academic informations to the teaching and student community for enhancing in the field of quality teaching and researches.

Bulbul Gogoi
Coordinator
Internal Quality Assurance Cell (IQAC),
Tinsukia College
01/09/2023

Chief Editor's Note

We are glad to present this volume of compilation of research articles on multi-disciplinary science entitled '*Horizon*' before you. This is our second endeavour of publishing edited books under the project undertaken by the IQAC, Tinsukia College with the objective of encouraging research activities and boasting the academic pursuit among the faculty members of the college. Accordingly we have published two edited books covering the disciplines of Arts, Commerce and Science last year and this book is a continuation of the process. However, for the present volume, we have invited research based papers/articles from some distinguished scholars, researchers and academicians from other institutions and this our pleasure that they have obliged us with their valuable contribution. Here I record our gratitude to all our esteemed contributors for their kindness. I also offer my thankfulness to the principal and coordinator, IQAC of Tinsukia College for putting trust on us and entrusted with the responsibility of bringing out the book.

I must appreciate the sincere efforts of the editor as well as the members of the editorial board to make the book possible. Despite all the care taken by our editors some errors may occur which is highly regretted.

In this volume, we have attempted to include articles on different and diverse branches of Natural Sciences contributed by Scholars/researchers and experts in the concerned fields of study so that it helps the readers look through the infinite Horizon of knowledge and get a glimpse of advanced studies and researches in various branches of Natural Sciences in different parts of the globe.

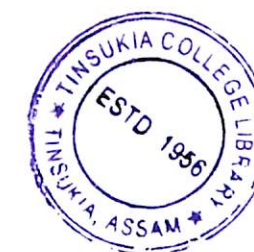
Finally, the edited book is before our esteemed readers and opens for public review. Every kind of feedback and suggestions from our esteemed readers are most welcome and highly appreciated. Thank you.

Nilim Jyoti Senapati
Chief Editor, Horizon
01/09/2023

HORIZON

A Collection of Research Based Articles on Multi-Disciplinary Science

ISBN : 978-81-19322-07-7



I have always felt that science can only flower when there is an internal urge. It cannot thrive under external pressures. I strongly believe that fundamental science cannot be driven by instructional, industrial, Governmental or military pressures.

Sir C. V. Raman, Noble Laureate

(From A. Jayaraman "C. V. Raman-A Memoir", Page No.90. Affiliated East-West Press, 1989, ISBN: 8185336245)

Contents



List of Contributors

Preface

1. Simulation: Methods, Applications, and Future Directions
P. P. Borthakur, D. Das 1-21
2. Darwin's Legacy and the Complexity of Evolutionary Battle against Pathogens
N. Borah 22-28
3. Silibinin loaded xanthan gum nanoparticles and its effect as an *in-vitro* anticancer agent
P. Gogoi 29-50
4. Phase transition driven structural transformation in colloidal membranes
L. Saikia 51-73
5. Obesity and Its Treatment with Herbal Remedies
K. Pathak, P. Kalita, A. Das, R. Saikia 74-91
6. Space-filling curves: A path tracing towards efficient miniature antenna
A. Das 92-102
7. Galaxy distribution observations predict a smaller Hubble constant
D. J. Gogoi, J. Bora 103-111
8. Influence of Mimas on Saturn's Ring System
S. Bora, D. J. Gogoi 112-121
9. Variations of NDVI and Its Association with Rainfall over North East India
A. Borah, N. Gogoi, R. Gautam, B. Pathak 122-129
10. Dust Coupling Strength in presence of Asymmetric Ion Flow
S. Bhattacharjee 130-141

11. A Discussion on Nonlinear Energy Exchange Process through Kinetic theory in Vlasov Ionospheric Plasma
S. J. Gogoi 142-152
12. Nurseries of Celestial Light: Exploring Star Formation in Young Open Clusters
A. Hussain 153-169
13. On the pulsar-like compact stars: Strangeon stars
J. Bora, D. J. Gogoi, S. Bora 170-176
14. The Theory of Black Hole
R. Karmakar 177-183
15. Propagation of UHE Cosmic Rays and its Signatures
S. P. Sarmah 184-190
16. Transition metal mediated oxidation of alcohols using environmentally benign oxidant
B. Banik 191-198
17. Synthesis of some *natural* and *synthetic* adhesives and their applications
S. Phookan, S. C. Das 199-225
18. A Review of *Six Sigma* as a Quality Improvement Program
P. Dutta 226-250
19. From Certainty to Uncertainty-A paradigm shift
D. Bhattacharjee 251-254
20. A Brief Note on *Pi*
R. Gogoi 255-257
21. A Study on the Method of Finding Rank of Elliptic Curves
D. Buragohain 258-270
22. Investigation of Phytochemicals in Some Common Vegetables Sold in the Market of Tinsukia Town
P. Saikia, P. Prasad 271-280

List of Contributors

Partha Protim Borthakur, M.Tech(VNIT Nagpur)
Assistant Professor, Department of Mechanical Engineering,
Dibrugarh University, Dibrugarh, India
parthaborthakur22@gmail.com

Dipankar Das, M.Tech.(IIT G), Ph.D.(IIT G)
Assistant Professor, Department of Mechanical Engineering,
Dibrugarh University, Dibrugarh, India
ursdipankardas@gmail.com

Nabajyoti Borah, M.Tech
Research Scholar
Institute of Microbiology of the Czech Academy of Sciences,
Prague, Czech Republic-14220
nabajyoti.borah@biomed.cas.cz

Plabita Gogoi, M.Sc., Ph.D.(TU)
Department of Chemical Sciences,
Tezpur University, Assam, India
plabita41@gmail.com

Lachit Saikia, M.Sc., Ph.D.(IISc)
Assistant Professor,
Department of Physics,
Dibru College, Dibrugarh, Assam, India
lachit.iisc@gmail.com

Kalyani Pathak, M.Pharm (DU), PhD (DU)
Woman Scientist (DHR-ICMR), Dept of Pharmaceutical sciences,
Dibrugarh University, Dibrugarh, India
kalyakster@gmail.com

Parimita Kalita, M.Pharm(DU)
Research Scholar. School of Pharmacy,
The Assam Kaziranga University, Jorhat, India

Aparoop Das, M.Pharm(BIT, Ranchi), PhD(DU)
Professor, Dept of Pharmaceutical sciences
Dibrugarh University, Dibrugarh, India
aparoopdas@dibru.ac.in

Riya Saikia, M.Pharm(DU)
Research Scholar, Dept of Pharmaceutical sciences,
Dibrugarh University, Dibrugarh, India

Abhijeet Das, M.Sc., Ph.D(RGU)
Center for Advanced Research, Department of Physics,
Rajiv Gandhi University, Arunachal Pradesh - 791112, India
abhijeet.das@rgu.ac.in

DhrubaJyoti Gogoi, M.Sc., Ph.D.(DU)
Assistant Professor, Department of Physics
Dibrugarh University, Dibrugarh, India
molojdhruba@gmail.com

Jyatsnasree Bora, M.Sc., Ph.D.(DU)
Department of Physics
Dibrugarh University, Dibrugarh, India

Shyamalee Bora, M.Sc.
Research Scholar, Department of Physics,
Dibrugarh University, Dibrugarh, India
swamaleebora@gmail.com

Anindita Borah, M.Sc.
Research scholar, Department of Physics,
Dibrugarh University, Dibrugarh, India
borahanindita97@gmail.com

Niki Gogoi, M.Sc.
Research scholar, Department of Physics,
Dibrugarh University, Dibrugarh, India

Rohit Gautam, M.Sc.
Research scholar, Department of Physics,
Dibrugarh University, Dibrugarh, India

Binita Pathak, M.Sc., M.Tech. (Andhra U), Ph.D. (DU)
Assistant Professor, Centre for Atmospheric Studies,
Department of Physics,
Dibrugarh University, Dibrugarh, India

Saurav Bhattacharjee, M.Sc., Ph.D.(TU)
Assistant Professor, Department of Physics,
Tinsukia College, Tinsukia, Assam, India
sauravtsk.bhattacharjee@gmail.com

Satyajyoti Gogoi, M.Sc., M.Phil.(VMU), Ph.D.(DU)
Assistant Professor, Department of Physics,
Tinsukia College, Tinsukia, India
satyajyotitsk@gmail.com

Arshad Hussain, M.Sc.
Research Scholar, Department of Physics,
Dibrugarh University, Dibrugarh, India
arshad007h@gmail.com

Ronit Karmakar, M.Sc.
Research Scholar, Department of Physics,
Dibrugarh University, Dibrugarh, India
ronit.karmakar622@gmail.com

Swaraj Pratim Sarmah, M.Sc.
Research Scholar, Department of Physics,
Dibrugarh University, Dibrugarh, India
swarajpratimsarmah16@gmail.com

Biplab Banik, M.Sc., Ph.D.(DU)
Assistant Professor
Department of Chemistry,
Tinsukia College, Tinsukia, Assam, India

Saswati Phookan, M.Sc, Ph..(DU),
Assistant Professor,
Department of Chemistry,
Tinsukia College, Tinsukia, Assam
saswatiphookan1973@gmail.com

Shauvik Chandra Das, M.Sc, B.Ed.
shauvikchdasmrg@gmail.com

Porinita Dutta, M.Sc., Ph.D.(DU)
Assistant professor
Department of Statistics
Tinsukia College, Tinsukia, Assam
dutta.porinita1@gmail.com

Deepika Bhattacharjee, M.Sc., Ph.D.(DU)
Associate Professor
Department of Mathematics
Tinsukia College, Tinsukia, Assam

Ranjan Kumar Gogoi, M.Sc., M.Phil.(DU)
Associate Professor
Department of Mathematics
Tinsukia College, Tinsukia, Assam

Debakanta Buragohain, M.Sc.
Assistant professor
Department of Mathematics
Tinsukia College, Tinsukia, Assam

Pankaj Saikia, M.Sc, Ph..(GU),
Assistant Professor,
Department of Chemistry,
Tinsukia College, Tinsukia, Assam



Preface

Curiosity is the architect of scientific researches and the innate human inclination to question, explore, and seek understanding about the natural world and the phenomena that shape it. It is this curiosity that sets the wheels of scientific research in motion and leads to the pursuit of knowledge and discovery. Curiosity is the driving force behind the entire scientific endeavour. It propels researchers to explore the unknown, seek answers, and uncover the mysteries of the universe. From asking fundamental questions to pushing the boundaries of human understanding, curiosity remains an essential and inspiring aspect of scientific research.

Scientific research ignites curiosity, fosters critical thinking, and nurtures a deeper understanding of the world around us. Scientific research is of paramount importance for a multitude of reasons, contributing to advancements in knowledge, its applications and our understanding of the natural world. Scientific research drives human progress, enriches our knowledge, and shapes the course of our civilization. It is an investment in the future that benefits both individuals and society as a whole.

Scientific research has a profound and multifaceted impact on students, shaping their educational experiences, personal development, and future career trajectories. It equips them with skills, insights, and experiences that enrich their education, shape their character, and position them as future leaders and innovators in a rapidly evolving world.

Thus, scientific researcher articles are published in the form of Journals/edited books to spread the new innovations and researches among academicians, intellectuals, students and public to foster research culture and motivate young minds. The publication

of research articles in edited books by institutes holds crucial significance in the academic realm. It serves as a platform for scholars to share their expertise and insights with a broader perspective. These collaborative volumes foster interdisciplinary connections, providing comprehensive explorations of topics while showcasing an institute's commitment to advancing knowledge and leaving a lasting imprint on their fields of study. In this regard, Tinsukia College have a taken a major step from last year to promote the same by publishing edited books and present volume is a collection of research based articles on multi-disciplinary science.

Last but not the least, I would like to offer my heartfelt gratitude to Dr.Surjya Chutia,Principal,Tinsukia College for his initiative, IQAC,Tinsukia College for financial support,entire editorial team for their cooperation and advice,to the all authors for their valuable write up and all staffs of 'Assam Computers' efforts to make "*Horizon*".

Satyajyoti Gogoi
Editor,*Horizon*
01/09/2023





Simulation: Methods, Applications, and Future Directions

**Partha Protim Borthakur⁽¹⁾,
Dipankar Das⁽²⁾**

*^{(1), (2)} Department of Mechanical Engineering,
Dibrugarh University, Dibrugarh, India
Mail: parthaborthakur22@gmail.com*

Abstract

This book chapter encompasses the diverse field of simulation by analyzing its historical development, methodologies, applications, and future prospects. Simulation, the act of mimicking real-world processes, systems, or phenomenon through computational models, plays a significant role in various disciplines such as engineering, healthcare, economics, education, and entertainment. Building upon the methodology section, the text delves into a wide range of current and emerging applications of simulation. With technological advancements and increasing computational power, the horizons of simulation are continually expanding. These applications span numerous sectors such as healthcare, manufacturing, transportation, military, and social sciences. Each sector-specific application of simulation is presented through case studies, highlighting the practical implications and value derived from simulation. The review speculates on the future directions of simulation. Here, the authors argue that the convergence of advanced computational technologies such as AI, machine learning, and quantum computing will bring forth a new era of simulation, significantly enhancing its capability and scope. They predict that simulation will play an increasingly crucial role in decision making and strategy formulation across various sectors, thanks to its enhanced predictive accuracy and risk evaluation capability. The chapter concludes by highlighting the necessity for developing novel simulation techniques and their integration with advanced Overall, "Simulation: Methods, Applications, and Future Directions" provides readers with a solid understanding of the simulation landscape, its real-world applications, and the trajectory it's likely to take in the coming years, positioning it as a fundamental tool for strategizing and decision-making in an increasingly complex world.

P.P.Borthakur,D.Das. Simulation: Methods, Applications, and Future Directions, Horizon(Ed.Book).1-21(2023),copyright:Tinsukia College Publishing.ISBN: 978-81-19322-07-7

1. Introduction

Simulation has been an essential tool for researchers, practitioners, and educators for decades. It offers a safe, cost-effective, and efficient means to study and analyze complex systems that might be infeasible or dangerous to experiment within the real world. Simulation methods represent a critical set of tools in modern science, engineering, and various other fields, enabling us to model and study complex systems by imitating their operation over time. These techniques facilitate the analysis of real-world scenarios, providing invaluable insights that assist decision-making and strategy formulation processes. Simulation methods allow us to create 'virtual laboratories', where we can run experiments, evaluate different strategies, and understand the behavior of systems without incurring the high costs, time, or potential risks associated with physical experimentation. The fundamental goal of simulation methods is to mimic the behavior of an actual system in a virtual environment to study its characteristics and evaluate its performance under various conditions [1,2]. Different types of simulation methodologies have evolved to suit various kinds of systems and problems. These include Discrete Event Simulation (DES), which is often used for queuing and logistical problems where events occur at specific points in time. System Dynamics (SD) is a methodology that focuses on understanding the non-linear behavior of complex systems over time using flows, accumulations, and feedback loops. Agent-Based Simulation (ABS) allows for the modeling of systems composed of autonomous, interacting agents, which is useful for understanding emergent behavior in complex systems. Each of these methodologies has unique strengths and is suited to different kinds of problems. DES is well suited to scenarios where events are distinct and affect the state of the system, such as customer arrivals in a queuing system. SD is most useful for systems where the focus is on overall dynamics and not individual events, such as the growth of a population. ABS excels in scenarios where the interaction between individual elements can lead to emergent system-level behavior, such as the spread of disease in a population. This chapter gives a brief history of simulation, its importance, Methods, Applications, presently available software and Future Directions on Simulation study [3,4].

2. Important simulation methods [5,6,7,8]

- 2.1 *Monte Carlo Simulation*: The Monte Carlo method employs random sampling techniques to solve complex mathematical problems. It has been widely used in fields such as finance, physics, and engineering.
- 2.2 *Discrete Event Simulation (DES)*: DES is utilized for modeling systems where changes occur at discrete points in time. This method is particularly applicable in areas like manufacturing, logistics, and healthcare.
- 2.3 *Continuous simulation*: In contrast to Discrete Event Simulation (DES) and Agent-Based Simulation (ABS), which deal with systems that change at specific points in time or in response to individual behaviors, continuous simulation concerns systems where changes occur continuously over time. Continuous simulation is typically used for systems

described by differential equations or differential-algebraic equations, which express how system variables evolve continuously with time. This makes it especially suitable for physical systems and processes such as fluid dynamics, heat transfer, electrical circuits, and population dynamics, where system states change in a continuous manner. An example of continuous simulation could be the modeling of a lake's temperature over the course of a year. Factors such as the air temperature, solar radiation, wind, and water flow would all continuously influence the lake's temperature, making this a continuous system. In a continuous simulation, time advances in very small increments, and the state variables are updated at each increment using mathematical models, typically the solutions of differential equations. This allows for a detailed representation of the system's behavior over time. While continuous simulation can provide a high level of detail and accuracy, it can also be computationally intensive, especially for large or complex systems. However, modern computational tools and techniques have made continuous simulation a feasible and valuable tool for analyzing and optimizing a wide variety of systems. Continuous simulation serves as a powerful method for modeling and studying systems that evolve continuously over time, providing a high-resolution view of system behavior and facilitating effective decision-making and system optimization.

- 2.4 *Agent-Based Simulation*: This approach simulates the actions and interactions of autonomous agents within a given environment. It is used in social sciences, ecology, and economics.
- 2.5 *System Dynamics*: It focuses on the modeling of dynamic systems and their feedback loops. It is extensively used in policy analysis, business strategy, and environmental studies.
- 2.6 *Finite Element Analysis (FEA)*: FEA is used for solving complex problems in structural mechanics, fluid dynamics, and other fields requiring spatial discretization.
- 2.7 *Molecular Dynamics Simulation*: This method simulates the physical movements of atoms and molecules and is heavily utilized in chemistry and material science.

3. Applications

Simulation techniques have been applied extensively across a wide range of domains due to their capability to imitate real-world systems and predict outcomes under various scenarios. Here are some of the most common applications [9,10]:

1. *Engineering*: Simulation tools such as FEA are employed to analyze stress and strain on structures and materials.
2. *Healthcare*: Simulation is used for training medical professionals through simulators and for analyzing healthcare systems via DES.
3. *Economics*: Economists employ agent-based simulations to study market dynamics and economic systems.

4. *Education*: Simulation tools offer interactive learning experiences, allowing students to understand complex concepts.
5. *Entertainment*: The gaming and film industry extensively use simulations to create realistic environments and character movements.
6. *Environmental Science*: Climate modeling and environmental impact analysis are conducted through simulation tools to study and predict climate change and its effects.
4. **Monte Carlo Simulation** [11,12,13]

4.1 *Introduction*: Monte Carlo Simulation (MCS) is a powerful computational technique used to model and analyze systems under uncertainty. It provides a probabilistic approach to decision-making by generating random samples from input distributions and simulating the system's behavior. This chapter explores the fundamentals of Monte Carlo Simulation, its applications, and the steps involved in conducting a Monte Carlo study.

4.2 *Basic Concepts of Monte Carlo Simulation*: Monte Carlo Simulation derives its name from the famous Monte Carlo Casino and the element of chance involved in the simulation process. It relies on random sampling and statistical analysis to estimate the behavior of complex systems.

4.3 *Random Sampling*: Random sampling is a fundamental component of Monte Carlo Simulation. It involves generating random numbers from probability distributions to represent uncertain inputs in the model. These random samples are then used to drive the simulation and generate output results.

4.4 *Probability Distributions*: Probability distributions describe the likelihood of different outcomes for uncertain variables. Common distributions used in Monte Carlo Simulation include normal (Gaussian), uniform, exponential, and triangular distributions. These distributions characterize the variability and uncertainty associated with model inputs.

4.5 *Steps in Conducting a Monte Carlo Study*: A Monte Carlo study typically involves several key steps, each of which plays a crucial role in obtaining accurate and meaningful results.

4.6 *Problem Formulation*: The first step is to clearly define the problem and the system to be simulated. This includes identifying the variables of interest, their relationships, and the objectives of the simulation study.

4.7 *Input Modeling*: Input modeling involves determining the probability distributions that best represent the uncertainty in the system's input variables. This requires analyzing historical data, expert opinions, or using available information to estimate the parameters of the distributions.

4.8 *Random Sampling*: Random sampling is used to generate values from the input distributions. Various sampling techniques, such as Latin Hypercube Sampling or random number generators, can be employed to ensure representative and efficient sampling.

4.9 *Simulation Execution*: Simulation execution involves running the model multiple times using the generated random samples as inputs. The model's behavior is simulated for each sample, and the corresponding output values are recorded.

4.10 *Statistical Analysis*: Statistical analysis is performed on the output data collected from the simulation runs. Descriptive statistics, such as mean, standard deviation, and percentiles, are computed to summarize the output distribution. Confidence intervals and hypothesis tests can also be used to make inferences about the system's behavior.

4.11 *Applications of Monte Carlo Simulation*: Monte Carlo Simulation finds applications in a wide range of fields and decision-making scenarios.

4.12 *Finance and Risk Analysis*: Monte Carlo Simulation is extensively used in finance for option pricing, portfolio optimization, and risk assessment. It allows for the evaluation of different investment strategies and provides insights into the potential range of financial outcomes.

4.13 *Engineering and Reliability Analysis*: In engineering, Monte Carlo Simulation helps analyze the reliability of systems by considering uncertainties in component parameters, environmental conditions, and operational factors. It aids in identifying critical components and optimizing system performance.

4.14 *Project Management*: Monte Carlo Simulation is employed in project management to estimate project schedules, costs, and resource utilization. By considering uncertainties in task durations and resource availability, it enables project managers to assess project risks and develop contingency plans.

4.15 *Quality Control and Manufacturing*: Monte Carlo Simulation assists in quality control by evaluating process variations and identifying potential defects. It helps optimize manufacturing processes by determining the impact of parameter variations on product quality and performance.

4.2 **Popular Monte Carlo Simulation Software:**

4.2.1 *@RISK*: @RISK is a widely used Monte Carlo Simulation software package integrated with Microsoft Excel. It offers a comprehensive suite of tools for uncertainty modeling, simulation execution, and result analysis. @RISK supports various distribution types, provides advanced statistical features, and offers integration with other analytical software.

4.2.2 *Crystal Ball*: Crystal Ball is another popular Monte Carlo Simulation software that integrates with Microsoft Excel. It offers a range of features for risk analysis, optimization, and decision-making. Crystal Ball includes a user-friendly interface, extensive statistical analysis capabilities, and advanced sensitivity analysis tools.

4.2.3 *Simul8*: Simul8 is simulation software that supports Monte Carlo Simulation and discrete event simulation. It offers a visual modeling environment and facilitates modeling complex systems using flowcharts and process diagrams. Simul8 provides features for input modeling, simulation execution, and result analysis.

4.2.4 *AnyLogic*: AnyLogic is a versatile simulation software that supports various modeling approaches, including Monte Carlo Simulation. It offers a graphical modeling environment and supports both continuous and discrete event simulations. AnyLogic provides extensive analysis and visualization capabilities.

5. Discrete Event Simulation: *Modeling and Analysis*[2,14,15]

Introduction: Discrete Event Simulation (DES) is a powerful tool used in various fields, including engineering, computer science, operations research, and many others. DES enables the representation and analysis of complex systems by modeling the flow of discrete events over time. This chapter explores the fundamental concepts, modeling techniques, and analysis methods employed in discrete event simulation.

5.1 *Basic Concepts of Discrete Event Simulation*: Discrete event simulation involves the simulation of a system as a sequence of discrete events that occur at specific points in time. These events drive the system's behavior and may include arrivals, departures, state changes, or other significant occurrences. Key concepts in DES include entities, attributes, events, and queues.

5.2 *Entities*: Entities represent the objects or entities of interest within the simulated system. They can be physical objects, such as customers in a queue, or abstract entities, such as packets in a computer network. Each entity typically possesses attributes that define its characteristics or state.

5.3 *Attributes*: Attributes describe the properties or characteristics of entities. They can be static or dynamic, influencing the behavior of the system. Examples of attributes include arrival time, service time, priority, or queue length.

5.4 *Events*: Events represent specific moments in time when significant changes occur within the system. These changes can include entity arrivals, departures, state transitions, or changes in resource availability. Events are scheduled and processed in chronological order.

5.5 *Queues*: Queues are data structures used to hold entities awaiting processing or service. Entities enter the queue upon arrival and are processed according to the system's rules. Queuing disciplines, such as First-In-First-Out (FIFO) or priority-based, determine the order in which entities are served.

5.6 *Modeling Techniques*: Effective simulation models require appropriate abstractions and simplifications of the real-world system. Several modeling techniques are commonly used in DES.

5.7 *System Dynamics Modeling*: System dynamics modeling focuses on capturing the dynamic behavior of a system using stock and flow diagrams. It emphasizes the interrelationships between different components and the feedback loops that drive the system's behavior.

5.8 *Process-Oriented Modeling*: Process-oriented modeling focuses on the sequence

of activities or processes that entities undergo within the system. It employs flowcharts or activity diagrams to represent the flow of entities and their interactions with resources.

5.9 *Event Graphs*: Event graphs provide a graphical representation of the events and their relationships in a DES model. They illustrate the order in which events occur, allowing analysts to understand the flow of entities through the system.

5.10 *Simulation Execution and Event-Driven Simulation*: To execute a discrete event simulation, an event-driven approach is commonly employed. 6.11 *Event Scheduling*: Event scheduling involves determining the next event to process and its associated time. This can be achieved using event lists or priority queues, where events are sorted by their scheduled time.

5.11 *Event Processing*: Event processing involves handling the occurrence of an event by updating the system state, generating new events, and performing relevant actions. Entities may move through different states, and resources may become available or busy depending on the event being processed.

5.12 *Analysis and Output Evaluation*: Once the simulation execution is complete, the obtained results need to be analyzed and evaluated to gain insights into system performance and behavior.

5.13 *Performance Metrics*: Performance metrics quantify the performance of the simulated system and provide a basis for comparison. Common metrics include throughput, response time, utilization, and waiting time.

5.14 *Experimental Design and Variance Reduction*: To obtain reliable estimates, simulation experiments must be carefully designed. Techniques like random number seeding, replication, and variance reduction are used to reduce statistical variability and improve the accuracy of results.

5.15 *Sensitivity Analysis*: Sensitivity analysis explores how changes in model parameters or assumptions affect system performance. It helps identify critical factors and understand the impact of uncertainties on the system's behavior.

DES software plays a crucial role in various fields, including manufacturing, logistics, healthcare, transportation, and more. Here are a few notable examples of discrete event simulation software:

i. *AnyLogic*: AnyLogic is a versatile simulation software that supports multiple modeling paradigms, including discrete event simulation. It offers a user-friendly interface and allows users to create complex simulations using a combination of discrete events, system dynamics, and agent-based modeling. AnyLogic supports a wide range of industries and provides powerful analysis tools.

ii. *Simio*: Simio is a simulation software that combines discrete and continuous modeling approaches. It offers a drag-and-drop interface, making it easy to

build models and define the logic behind events. Simio includes advanced features like 3D visualization, optimization, and the ability to interface with other software tools. It finds applications in manufacturing, healthcare, transportation, and other domains.

iii. *Arena*: Arena is a popular simulation software developed by Rockwell Automation. It provides a comprehensive set of tools for modeling and analyzing complex systems using discrete event simulation. Arena offers a visual interface and a wide range of built-in modules, making it suitable for various industries, including manufacturing, logistics, and supply chain management.

iv. *SIMUL8*: SIMUL8 is a user-friendly simulation software that focuses on discrete event simulation. It allows users to create models using a flowchart-based approach, making it accessible to both beginners and experienced users. SIMUL8 provides a range of analysis tools and supports integration with external data sources, making it suitable for diverse applications.

v. *ExtendSim*: ExtendSim is a simulation software that supports both discrete event and continuous modeling. It offers a graphical interface and a vast library of pre-built modeling blocks, simplifying the creation of complex simulations. ExtendSim is widely used in various fields, including manufacturing, supply chain management, and healthcare.

Discrete event simulation is a valuable tool for modeling and analyzing complex systems. By representing the flow of discrete events and capturing the behavior of entities and resources, DES enables the exploration of system dynamics, optimization of processes, and evaluation of performance. Understanding the basic concepts, modeling techniques, and analysis methods described in this chapter provides a solid foundation for leveraging the power of discrete event simulation in diverse domains.

6. Continuous Simulation [15,16,17,18]

Introduction: In the world of computer simulations, continuous simulation is a powerful technique used to model dynamic systems that evolve continuously over time. Unlike discrete simulations, which model systems that change in discrete steps, continuous simulations provide a more detailed representation of real-world phenomena by capturing continuous changes. This chapter explores the concepts, techniques, and applications of continuous simulation.

6.1 *Understanding Continuous Simulation*: Continuous simulation involves the modeling of systems where the state variables change continuously and smoothly over time. These systems can be described by differential equations or difference equations, which capture the relationships between the variables and their rates of change. By numerically solving these equations, continuous simulation approximates the behavior of the system and enables the analysis of complex dynamic processes.

6.2 *Mathematical Modeling*: To perform continuous simulation, it is necessary to develop

mathematical models that describe the system under consideration. These models can take the form of ordinary differential equations (ODEs) or partial differential equations (PDEs) depending on the nature of the system. The choice of equations and their parameters depends on the specific problem at hand and the phenomena being simulated.

6.3 *Numerical Integration Methods*: Numerical integration methods play a crucial role in continuous simulation by approximating the solutions to the differential equations. These methods convert the continuous equations into discrete steps, allowing for the calculation of the system's state at specific time intervals. Some commonly used numerical integration methods include Euler's method, Runge-Kutta methods, and the finite element method for PDEs. These methods ensure accurate and efficient computation of the system's behavior.

6.4 *Simulation Time*: In continuous simulation, time is a fundamental parameter. It is divided into small intervals called time steps, and the system's state is updated at the end of each time step based on the integration method used. The choice of time step size is critical to balance the trade-off between accuracy and computational efficiency. A small time step provides more accurate results but requires more computational resources, while a large time step can lead to inaccuracies in the simulation.

6.5 *Continuous Simulation Environments*: Various software tools and programming languages are available to implement continuous simulations. These environments provide libraries, functions, and frameworks specifically designed for numerical computations and simulation purposes. Examples of popular simulation environments include MATLAB, Python with libraries like SciPy and NumPy, and specialized simulation software like Simulink or COMSOL.

6.6 *Applications of Continuous Simulation*: Continuous simulation finds applications in a wide range of fields, including physics, engineering, finance, biology, and climate science. Some common examples include modeling the behavior of electrical circuits, predicting the spread of diseases, simulating fluid dynamics, analyzing financial markets, and studying climate change patterns. Continuous simulation enables researchers, engineers, and scientists to gain insights, make predictions, and optimize systems without the need for expensive and time-consuming physical experiments.

6.7 *Challenges and Limitations*: While continuous simulation is a powerful tool, it also faces certain challenges and limitations. The accuracy of the simulation heavily relies on the mathematical models used and the numerical integration methods employed. Complex systems with nonlinear behavior can be particularly challenging to model accurately. Additionally, the computational resources required for continuous simulation can be significant, especially for large-scale and long-duration simulations.

6.8 *Future Directions*: Continuous simulation continues to evolve with advancements in computational power, mathematical modeling techniques, and simulation environments. Researchers are working on developing more efficient numerical integration methods,

improving modeling approaches, and exploring new applications in emerging fields. The integration of machine learning and artificial intelligence techniques with continuous simulation is also an area of active research, opening up new possibilities for enhanced modeling and analysis.

Continuous simulation software is widely used in fields such as process engineering, physics, economics, and environmental modeling. Here are a few notable examples of continuous simulation software:

- i. *Aspen Plus*: Aspen Plus is a process simulation software widely used in chemical engineering and related industries. It allows engineers to model and simulate complex chemical processes involving multiple components and reactions. Aspen Plus employs rigorous mathematical models to simulate mass and energy balances, phase equilibria, and other phenomena. It offers a comprehensive set of tools for process optimization, sensitivity analysis, and design.
- ii. *COMSOL Multiphysics*: COMSOL Multiphysics is a powerful simulation software that enables the modeling and simulation of coupled physical phenomena in various scientific and engineering disciplines. It supports the simulation of a wide range of physics, such as fluid dynamics, structural mechanics, heat transfer, electromagnetics, and more. COMSOL Multiphysics provides a flexible and customizable modeling environment and allows users to create and solve their own systems of differential equations.
- iii. *MATLAB/Simulink*: MATLAB and Simulink are widely used software packages for mathematical modeling, simulation, and analysis. Simulink, a graphical programming environment, is particularly well-suited for continuous system simulation. It offers a block diagram interface where users can build models by connecting different blocks representing mathematical equations, inputs, outputs, and other components. MATLAB provides powerful mathematical functions and tools to support simulation and data analysis.
- iv. *Dymola*: Dymola is a modeling and simulation environment specifically designed for modeling and simulating complex physical systems. It supports the modeling of multidomain systems involving mechanical, electrical, thermal, and other physical phenomena. Dymola uses the Modelica language to describe system behavior and equations. It offers a range of analysis tools, including parameter optimization and sensitivity analysis.
- v. *Simio*: While Simio was previously mentioned as a discrete event simulation software, it also supports continuous simulation. In addition to discrete event modeling, Simio allows users to model and simulate continuous systems with a focus on dynamic behavior. It offers tools to model continuous flows, resource allocation, and real-time constraints, making it suitable for various applications, including manufacturing, logistics, and transportation.

Continuous simulation is a valuable technique for modeling and analyzing dynamic systems that evolve continuously over time. By capturing the continuous changes in variables, it provides a more detailed.

7. Agent-Based Simulation: Modeling Complex Systems and Emergent Behavior[19,20,21]

Introduction: Agent-Based Simulation (ABS) is a modeling technique that focuses on the behavior and interactions of individual autonomous entities, known as agents, within a system. ABS allows for the exploration of complex systems by simulating the actions and decisions of individual agents and observing the emergent behavior that arises from their interactions. This chapter explores the principles, applications, and methodologies of Agent-Based Simulation.

- 7.1 *Basic Concepts of Agent-Based Simulation:* Agent-Based Simulation is rooted in the idea that complex systems can be better understood by examining the behavior of their constituent parts. The following concepts form the foundation of ABS:
- 7.2 *Agents:* Agents are autonomous entities that have their own goals, knowledge, and decision-making capabilities. They can represent individuals, organizations, animals, or even abstract entities. Agents interact with each other and their environment, and their actions collectively shape the behavior of the system.
- 7.3 *Environment:* The environment represents the space or context in which agents operate. It includes physical elements, such as geography or resources, as well as social and cultural factors. The environment provides the context for agent interactions and influences their decision-making processes.
- 7.4 *Interactions:* Interactions occur when agents communicate, exchange information, or influence each other's behavior. These interactions can be direct (face-to-face) or indirect (through the environment). The nature and rules of interactions govern the dynamics of the system and can lead to emergent phenomena.
- 7.5 *Emergent Behavior:* Emergent behavior refers to the collective behavior or patterns that emerge from the interactions of individual agents. It is often characterized by properties or phenomena that are not explicitly programmed but arise as a result of agent interactions. Emergent behavior is a key aspect of ABS, allowing for the exploration of system-level dynamics.
- 7.6 *Agent-Based Modeling and Methodologies:* Agent-Based Modeling (ABM) involves the construction and analysis of agent-based simulation models. Several methodologies guide the development of ABM models:
- 7.7 *Agent Design:* Agent design involves defining the characteristics and attributes of agents, including their goals, decision-making processes, and rules of behavior. Agents should be designed to capture the relevant aspects of the system being modeled while simplifying or abstracting unnecessary details.

- 7.8 *Environment Specification*: The environment specification entails defining the context and spatial representation of the simulation. This includes specifying the geography, resources, and constraints within which agents interact. The environment should reflect the key features and dynamics of the real-world system.
- 7.9 *Interaction Modeling*: Interaction modeling focuses on capturing the rules and mechanisms that govern agent interactions. This involves defining communication protocols, social networks, or other forms of interaction channels. Interaction models should capture the dynamics of information exchange and influence between agents.
- 7.10 *Validation and Verification*: Validation and verification are critical steps in ABM. Validation involves comparing the simulation outputs with real-world observations or data to ensure the model accurately represents the system. Verification ensures that the simulation model is implemented correctly, with agents and their interactions functioning as intended.
- 7.11 *Applications of Agent-Based Simulation*: Agent-Based Simulation has found applications in various domains, allowing researchers and decision-makers to study complex systems and phenomena.
- 7.12 *Social Sciences and Economics*: ABS has been extensively used in social sciences and economics to model human behavior, social dynamics, and economic systems. It enables the study of phenomena such as market dynamics, social network effects, urban planning, and the spread of diseases.
- 7.13 *Ecology and Environmental Science*: ABS provides a powerful tool for studying ecological systems, including the interactions between species, ecosystem dynamics, and the impact of environmental changes. It aids in understanding phenomena like predator-prey relationships, population dynamics, and the effects of climate change.
- 7.14 *Transportation and Traffic Simulation*: ABS has been applied to transportation and traffic simulation to model the behavior of drivers, traffic flow, and the impact of transportation policies. It helps in evaluating congestion, optimizing traffic management strategies, and studying the effects of infrastructure changes.
- 7.15 *Healthcare and Epidemiology*: ABS is used in healthcare and epidemiology to model the spread of diseases, assess healthcare policies, and study the effects of interventions. It allows for the exploration of factors influencing disease transmission, healthcare access, and the effectiveness of control measures.
- 7.16 *Advancements and Challenges*: ABS has witnessed advancements in recent years, thanks to improved computational power and data availability. However, challenges remain in areas such as scaling up simulations, incorporating realistic cognitive processes into agents, and validating complex models. Interdisciplinary collaboration and the integration of real-time data sources hold promise for addressing these challenges. Agent-Based Simulation offers a unique perspective on modeling complex systems by

focusing on the behavior and interactions of individual agents. Through the study of emergent behavior, ABS provides insights into system-level dynamics that may not be apparent from traditional modeling approaches. By understanding the principles, methodologies, and applications of ABS, researchers and decision-makers can leverage this powerful tool to gain deeper insights into complex systems and inform policy and decision-making processes.

There are several popular agent-based simulation software packages available today, each with its own unique features and capabilities. Let's explore a few notable examples:

- i. *NetLogo*: NetLogo is a widely-used, open-source software that provides an intuitive environment for creating agent-based simulations. It features a simple programming language and a graphical interface, making it accessible to users with diverse backgrounds. NetLogo supports the modeling of various domains, such as social sciences, biology, and economics, and offers a range of built-in agent behaviors and analysis tools.
- ii. *Repast*: Repast (Recursive Porous Agent Simulation Toolkit) is another popular open-source software framework for building agent-based models. It provides a comprehensive set of tools for defining agents, their behaviors, and the environment in which they operate. Repast supports both discrete-event and continuous-time modeling paradigms, making it suitable for a wide range of applications, including social dynamics, transportation systems, and ecology.
- iii. *AnyLogic*: AnyLogic is a commercial multi-method simulation software that includes agent-based modeling capabilities along with system dynamics and discrete-event simulation. It offers a visual modeling environment with drag-and-drop components, allowing users to construct complex agent-based models quickly. AnyLogic also provides extensive data analysis and visualization features, enabling users to gain insights from simulation results effectively.
- iv. *GAMA*: GAMA (Generic Architecture for Modular Agent-based Modeling and Simulation) is an open-source platform designed to support the development of agent-based simulations. It offers a flexible modeling language and an integrated development environment for building complex simulations. GAMA provides a wide range of modeling features and supports the integration of external libraries, making it suitable for various domains, including urban planning, ecology, and economics.
- v. *MASON*: MASON (Multi-Agent Simulator of Neighborhoods) is a Java-based simulation toolkit specifically designed for developing large-scale agent-based models. It focuses on the efficient execution of simulations with a large number of agents, leveraging parallel computing and spatial indexing techniques. MASON provides a high degree of flexibility, allowing users to customize and extend the simulation framework to meet their specific requirements.

These software packages, along with many others, empower researchers and



practitioners to explore and analyze complex systems in a simulated environment. By studying the emergent behavior of agents and their interactions, users can gain valuable insights into the dynamics of real-world systems, inform decision-making processes, and explore hypothetical scenarios. Agent-based simulation software plays a crucial role in understanding and addressing complex problems across various fields, including social sciences, economics, biology, transportation, and urban planning.

8. System Dynamics Simulation[22,23,24]

In a world where complexity governs various systems from traffic to economies, understanding these systems becomes paramount. System Dynamics is a robust methodology that aims to comprehend and analyze the behavior of complex systems over time. This chapter delves into System Dynamics Simulation, elucidating its principles, elements, and applications. Moreover, it portrays how this technique is imperative for decision-making and solving real-world problems.

8.1 *Background*: System Dynamics was developed in the 1950s by Jay W. Forrester, who was intrigued by the complex behavior of industrial systems. It is primarily a methodology for studying and managing complex feedback systems, such as one finds in business and other social systems. In System Dynamics, the emphasis is on the understanding of the underlying structure and behavior of a system.

8.2. Elements of System Dynamics Simulation

8.2.1 *Stocks*: Stocks represent the accumulations or state variables in the system. Examples of stocks could include the amount of money in a bank account, the population of a city, or the inventory of a store.

8.2.2 *Flows*: Flows represent the rates of change or activities that cause stocks to increase or decrease over time. For example, deposits and withdrawals are flows that affect the money stock in a bank account.

8.2.3 *Feedback Loops*: Feedback loops are cycles that exist when changes in a stock affect the flow into or out of that same stock. There are two types of feedback loops: positive feedback, which is self-reinforcing, and negative feedback, which is self-correcting.

8.2.4 *Time Delays*: Time delays occur when actions do not yield immediate results. This is common in real-world systems, where there can be a lag between the change in a variable and its effect on the system.

8.2.5 *Parameters*: Parameters are constants that affect the behavior of the system but do not change over time. For example, the interest rate on a bank account would be a parameter.

8.3 Building a System Dynamics Model

8.3.1 *Problem Definition*: The first step in creating a System Dynamics model is to define the problem or system you want to analyze. This involves identifying the boundaries of the system, the time horizon, and the key variables.

8.3.2 *Conceptualization*: Conceptualization involves creating a causal loop diagram to visually represent the structure of the system. This includes identifying stocks, flows, feedback loops, and other elements.

8.3.3 *Formulation*: In this phase, the conceptual model is translated into a mathematical model by defining equations that describe the relationships between variables.

8.3.4 *Simulation*: Simulation involves running the model over time to analyze how the system evolves. This can be done using specialized software such as Stella, Vensim, or AnyLogic.

8.3.5 *Analysis and Policy Design*: After simulating the model, the results must be analyzed to understand the behavior of the system. Based on the analysis, policies or decisions can be designed to achieve desired outcomes.

8.3.6 *Validation*: Model validation is a crucial step where the model's output is compared with historical data or used in experiments to ascertain its reliability.

8.4 Applications of System Dynamics Simulation

System Dynamics Simulation has applications across various fields, including:

- Business Management*: For strategy planning, understanding market dynamics, supply chain management, etc.
- Public Policy*: To evaluate the impact of policies on social systems, such as healthcare, education, or transportation.
- Environmental Studies*: For resource management, understanding ecological systems, or evaluating environmental policies.
- Engineering*: In design and analysis of complex engineering systems.
- Economics*: For macroeconomic modeling, understanding economic cycles, and policy analysis.

System Dynamics Simulation serves as a potent tool in understanding and managing complex systems. It is not only used for analysis but also for decision-making and policy design. By comprehending the components and the process of System Dynamics, individuals and organizations can efficiently tackle the multifarious challenges posed by intricate systems in our contemporary world. Some popular System Dynamics Simulation software options:

- Vensim*: Vensim is a widely used software package for System Dynamics modeling. It offers a user-friendly interface and a range of modeling capabilities, including stock and flow diagrams, equations, and sensitivity analysis. Vensim supports both continuous and discrete simulation and provides advanced features such as optimization and Monte Carlo simulation.
- AnyLogic*: AnyLogic is a versatile simulation software that supports various modeling methodologies, including System Dynamics. It offers a graphical modeling environment where users can build dynamic models using a combination of stock and flow diagrams,

statecharts, and agent-based modeling. AnyLogic also integrates with other modeling paradigms, allowing for hybrid models that capture different aspects of a system.

- iii. *Stella Architect*: Stella Architect is a user-friendly System Dynamics modeling software with a focus on ease of use and visual representation. It provides a drag-and-drop interface for building causal loop diagrams and stock and flow models. Stella Architect offers simulation capabilities, sensitivity analysis, and an extensive library of pre-built model components to facilitate the modeling process.
 - iv. *Powersim Studio*: Powersim Studio is a comprehensive simulation software that supports both System Dynamics and discrete event modeling. It provides a visual modeling environment with a range of modeling elements, including stocks, flows, converters, and delays. Powersim Studio offers advanced features such as scenario analysis, optimization, and statistical analysis.
 - v. *Simulink*: Simulink, developed by Math Works, is a popular simulation and modeling environment used in various domains, including System Dynamics. It offers a block diagram interface where users can build dynamic models by connecting different blocks representing system components. Simulink supports continuous and discrete simulation and provides extensive analysis and visualization capabilities.
- These software programs are widely used in academic, research, and industrial settings to study complex systems such as supply chains, economic systems, environmental systems, and social systems. They enable users to explore the behavior of these systems, evaluate policy interventions, and make informed decisions based on the simulation results.

9. Finite Element Analysis Simulation

Finite Element Analysis (FEA) is a powerful numerical method used to simulate the behavior of structures and systems under various loading conditions. It has become an essential tool in engineering design and analysis, allowing engineers to predict and understand the response of complex systems. This chapter explores the principles and applications of FEA simulation and its role in modern engineering practices [25,26,27].

- i. *Fundamentals of Finite Element Analysis 2.1*: Mathematical Basis FEA is based on the concept of discretization, where a complex geometry is divided into a finite number of smaller elements, typically triangles or quadrilaterals in two dimensions, or tetrahedra or hexahedra in three dimensions. The behavior of each element is then described using mathematical equations, resulting in a system of equations that represents the entire structure.

Different element types are used in FEA, each with its own set of shape functions that describe the displacement field within the element. Common element types include linear and quadratic elements, which are capable of representing various deformation patterns. The choice of element type depends on the complexity of the problem and the desired accuracy.

Accuracy Achieving convergence in FEA simulations is crucial to obtain accurate results.

Engineers should monitor convergence criteria and refine the mesh or adjust solver settings if necessary to ensure stable and reliable solutions.

Some notable FEA software packages:

- i. *ANSYS*: ANSYS is a leading commercial FEA software widely used in industries such as aerospace, automotive, and energy. It offers a comprehensive suite of tools for structural, thermal, fluid, and electromagnetic analysis. ANSYS provides a user-friendly interface, robust solver capabilities, and advanced features like nonlinear analysis, optimization, and multiphysics simulations.
- ii. *Abaqus*: Abaqus, developed by Dassault Systèmes, is another popular FEA software used for a wide range of engineering applications. It offers advanced modeling capabilities for linear and nonlinear analysis, including structural, thermal, and coupled multiphysics simulations. Abaqus provides a powerful scripting interface and integration with other software, making it highly customizable and suitable for complex analysis requirements.
- iii. *COMSOL Multiphysics*: COMSOL Multiphysics is a versatile FEA software that combines finite element analysis with other modeling techniques like finite volume and boundary element methods. It allows for the simulation of various physical phenomena, including structural mechanics, fluid dynamics, heat transfer, and electromagnetics. COMSOL Multiphysics offers a user-friendly interface and extensive built-in libraries, enabling the modeling of multiphysics phenomena.
- iv. *MSC Nastran*: MSC Nastran is a widely used FEA software known for its capabilities in structural analysis. It offers robust solvers for linear and nonlinear static, dynamic, and transient analyses. MSC Nastran provides advanced features such as optimization, fatigue analysis, and composites modeling. It is often used in industries such as aerospace, automotive, and manufacturing.
- v. *LS-DYNA*: LS-DYNA is a specialized FEA software primarily focused on nonlinear and dynamic analysis. It is commonly used in industries like automotive, aerospace, and defense to simulate crashworthiness, impact, and other high-speed events. LS-DYNA offers extensive material models, contact algorithms, and multi-physics capabilities, making it suitable for complex simulations involving large deformations and transient phenomena.
- vi. *Femap*: Femap is a pre- and post-processing software package for FEA simulations. It provides a user-friendly environment for creating, modifying, and analyzing finite element models. Femap supports a wide range of solvers, including NX Nastran, Abaqus, and ANSYS, making it compatible with various analysis requirements. It offers advanced meshing tools, visualization capabilities, and optimization features.
- vii. *Open-source FEA Software*: Several open-source FEA software packages are available, providing cost-effective alternatives with extensive capabilities. Examples include CalculiX, Code_Aster, and Elmer. These software packages offer various

analysis functionalities, and their open-source nature allows for customization and community-driven development.

10. Molecular Dynamics Simulation

Simulation Molecular Dynamics (MD) simulation is a computational technique used to study the behavior of atoms and molecules over time. It provides insights into the dynamics, thermodynamics, and properties of materials at the atomic scale. This chapter explores the principles, methodologies, and applications of MD simulation in various fields of science and engineering[28,29,30].

10.1 Theoretical Basis of Molecular Dynamics

10.1.1 *Newton's Equations of Motion MD simulation* is based on Newton's laws of motion, which describe the relationship between the forces acting on particles and their resulting motion. By numerically integrating these equations, the positions and velocities of atoms are calculated at each time step.

10.1.2 *Force Fields and Potential Energy MD simulations* utilize force fields, which are mathematical functions that describe the interatomic interactions and energy of the system. Force field consist of bonded and non-bonded terms, accounting for covalent bonds, angles, dihedral angles, van der Waals forces, and electrostatic interactions.

10.2 Simulation Methodology

10.2.1 *System Initialization MD simulations* begin by creating an initial configuration of atoms or molecules, often obtained from experimental data or computational methods such as molecular docking or quantum mechanics. The system is then equilibrated to remove any initial biases and relax into a stable state.

10.2.2 *Integration Algorithms* Numerical integration algorithms, such as the Verlet algorithm or the leapfrog method, are employed to update the positions and velocities of particles at each time step. These algorithms ensure the conservation of energy and preserve the stability of the simulation.

10.2.3 *Periodic Boundary Conditions* Periodic boundary conditions are utilized to simulate an infinite system with a finite number of particles. This technique allows for the simulation of bulk materials and avoids surface effects. Particles exiting one side of the simulation box re-enter from the opposite side.

10.3 Applications of Molecular Dynamics Simulation:

10.3.1 *Materials Science MD simulation* plays a crucial role in materials science, allowing for the study of material properties, phase transitions, diffusion, and mechanical behavior. It aids in the design of novel materials with desired characteristics and provides understanding at the atomic level.

10.3.2 *Biochemistry and Drug Design In biochemistry, MD simulation helps elucidate* protein folding, binding mechanisms, and drug-protein interactions. It aids in rational

drug design, virtual screening, and optimization of drug candidates by simulating their interactions with target proteins.

10.3.3 *Nanotechnology MD simulation* is essential in nanotechnology, where it provides insights into the behavior of nanoparticles, nanoscale structures, and self-assembled systems. It aids in the understanding and design of nanomaterials for various applications, including electronics, energy storage, and drug delivery.

10.3.4 *Fluid Dynamics and Chemical Reactions MD simulation* enables the study of fluid dynamics, including hydrodynamics, transport properties, and flow behavior. It also allows the investigation of chemical reactions at the atomic level, providing information on reaction rates, mechanisms, and energy profiles.

11. Conclusion

Simulation techniques have fundamentally transformed our ability to understand, analyze, and predict the behavior of complex systems in virtually every field. They offer a powerful means to replicate real-world scenarios and conduct experiments in a controlled, virtual environment, thereby reducing costs, saving time, and mitigating risks associated with real-world testing. From the manufacturing floor to the corridors of healthcare, from the front lines of military conflict to the hustle of urban planning, from the intricacies of financial markets to the complexities of social sciences, simulation techniques have proved to be indispensable. They have not only enabled operational efficiency and strategic decision-making but also facilitated advancements in our scientific understanding of various phenomena. With the proliferation of computational power, the integration of AI, machine learning, and the advent of quantum computing, the capabilities of simulation are set to expand even further. This convergence of technologies promises to enhance the accuracy, scope, and depth of simulations, making them an even more integral part of our decision-making process. In conclusion, simulation techniques serve as a cornerstone of modern system analysis and prediction, playing a critical role in numerous sectors and disciplines. As we navigate the complexities of the future, the importance of simulation as a tool for understanding and addressing real-world challenges will continue to grow.

References

- [1] M. Treiber and A. Kesting, "An open-source microscopic traffic simulator," *IEEE Intel Transp. Syst. Mag.*, vol. 2, no. 3, pp. 6–13, 2010, DOI: 10.1109/MITS.2010.939208.
- [2] M. Saidallah, A. El Fergougui, and A. E. Elalaoui, "A comparative study of urban road traffic simulators," *MATEC Web Conf.*, vol. 81, 2016, doi: 10.1051/mateconf/20168105002.
- [3] C. J. Liu, Z. Liu, Y. J. Chai, and T. T. Liu, "Review of Virtual Traffic Simulation and Its Applications," *J. Adv. Transp.*, vol. 2020, 2020, DOI: 10.1155/2020/8237649.

- [4] G. Kotusevski and K. A. Hawick, "A Review of Traffic Simulation Software," *Res. Lett. Inf. Math. Sci.*, vol. 13, pp. 35–54, 2009.
- [5] P. A. Lopez *et al.*, "Microscopic Traffic Simulation using SUMO," *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC*, vol. 2018–November, pp. 2575–2582, 2018, DOI: 10.1109/ITSC.2018.8569938.
- [6] *et al.*, "A SUMO Based Simulation Framework for Intelligent Traffic Management System," *J. Traffic Logist. Eng.*, vol. 8, no. 1, pp. 1–5, 2020, DOI: 10.18178/jtle.8.1.1-5.
- [7] X. Ma, X. Hu, T. Weber, and D. Schramm, "applied sciences," 2021.
- [8] Law, A. M. (2014). *Simulation modeling and analysis* (5th ed.). New York: McGraw-Hill Education.
- [9] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent Development and Applications of {SUMO - Simulation of Urban MObility}," *Int. J. Adv. Syst. Meas.*, vol. 5, no. 3, pp. 128–138, 2012, [Online]. Available: <http://elib.dlr.de/80483/>.
- [10] J. Nguyen, S. T. Powers, N. Urquhart, T. Farrenkopf, and M. Guckert, "An Overview of Agent-based Traffic Simulators," pp. 1–22, 2021, [Online]. Available: <http://arxiv.org/abs/2102.07505>.
- [11] Robert, C., & Casella, G. (2010). *Introducing Monte Carlo Methods with R*. New York: Springer.
- [12] Fishman, G. S. (1996). *Monte Carlo: Concepts, algorithms, and applications*. Springer Science & Business Media.
- [13] Rubinstein, R. Y., & Kroese, D. P. (2016). *Simulation and the Monte Carlo method*. John Wiley & Sons.
- [14] Chwif, L., & Medina, A. (2015). *Discrete Event Simulation Model Design and Execution: Building, Validating, and Experimenting*. CreateSpace Independent Publishing Platform.
- [15] Cassandras, C. G., & Lafortune, S. (2009). *Introduction to Discrete Event Systems*. Springer.
- [16] Cellier, F. E., & Kofman, E. (2006). *Continuous System Simulation*. Springer Science & Business Media.
- [17] Zeigler, B. P., Praehofer, H., & Kim, T. G. (2000). *Theory of Modeling and Simulation: Integrating Discrete Event and Continuous Complex Dynamic Systems*. Academic Press.
- [18] Law, A. M., & Kelton, W. D. (2000). *Simulation Modeling and Analysis*. McGraw-Hill.
- [19] Bonabeau, E. (2002). *Agent-based modeling: Methods and techniques for simulating human systems*. Proceedings of the National Academy of Sciences, 99(Supplement 3), 7280–7287.
- [20] Gilbert, N., & Troitzsch, K. (2005). *Simulation for the Social Scientist*. Open University Press.
- [21] Macal, C., & North, M. (2010). Tutorial on agent-based modeling and simulation. *Journal of Simulation*, 4(3), 151–162.
- [22] Richardson, G. P. (2011). *Reflections on the Foundations of System Dynamics*. System Dynamics Review, 27(3), 219–243.
- [23] Meadows, D. H., Meadows, D. L., Randers, J., & Behrens III, W. W. (1972). *The Limits to Growth*. Universe Books.
- [24] Morecroft, J. (2007). *Strategic Modelling and Business Dynamics: A feedback systems approach*. Wiley.
- [25] Zienkiewicz, O. C., & Taylor, R. L. (2000). *The Finite Element Method: The Basis*. Butterworth-Heinemann.
- [26] Cook, R. D., Malkus, D. S., Plesha, M. E., & Witt, R. J. (2001). *Concepts and Applications of Finite Element Analysis*. John Wiley & Sons.
- [27] Hughes, T. J. R. (2000). *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*. Dover Publications.
- [28] Rapaport, D. C. (2004). *The Art of Molecular Dynamics Simulation*. Cambridge University Press.
- [29] Allen, M. P., & Tildesley, D. J. (1989). *Computer Simulation of Liquids*. Oxford University Press.
- [30] Tuckerman, M. (2010). *Statistical Mechanics: Theory and Molecular Simulation*. Oxford University Press.

Darwin's Legacy and the Complexity of Evolutionary Battle against Pathogens

Nabajyoti Borah

Institute of Microbiology of the Czech Academy of Sciences,
Prague, Czech Republic-14220
Email: nabajyoti.borah@biomed.cas.cz

Abstract

Charles Darwin revolutionized the scientific landscape with his ground-breaking theory of evolution. His ideas regarding the origin of life also had a profound impact on the understanding of creation within religious circles. Despite the immense influence of Darwin's theories, they have faced significant challenges both within and outside the scientific community since their inception. One notable limitation in Darwin's theory pertains to its inability to account for horizontal gene transfer, a critical mechanism underlying the development of antimicrobial resistance. As the global threat of antimicrobial resistance looms, the scientific community is increasingly concerned, necessitating a re-evaluation of Darwin's evolutionary concepts within this context. Rather than dismissing Darwin's theory outright, this article tries to argue for an inclusive approach that embraces new insights to address this pressing issue effectively.

Darwin's Impact: The Controversial Perspectives

The publication of Darwin's *The Origin of Species* in 1859 revolutionized biology by introducing the concept of evolution and challenging the supernatural theories of species origin. This, along with Mendel's work on inheritance and the discovery of DNA's role in heredity, forms the foundation of modern biology. Darwin's evolutionary perspective recognized humans as one species among many, interconnected through a shared ancestry and engaged in reproduction, adaptation, survival, and change¹. It is this liberating vision that continues to honour Darwin two centuries after his birth. However, despite the enduring significance of Darwin's contributions in biology, they have not been without controversy. The debate surrounding his ideas is not limited to religious circles but has also extended to the scientific community itself. This has led to questions regarding the disparity in recognition between

N. Borah, "Darwin's Legacy and the Complexity of Evolutionary Battle against Pathogens", Horizon (Ed. Book), 22-28(2023), copyright: Tinsukia College Publishing, ISBN: 978-81-19322-07-7

Darwin and other scientific pioneers, such as Newton, despite Darwin's comparable significance in the field of science.

When ordinary textbooks state about Darwinian evolution and how different species are connected through a common ancestry, they often use the metaphor of a tree, known as the tree of life. In this metaphor, all organisms can be traced back to a single trunk, representing the common ancestor, and then they diverge into different lineages, comparable to major limbs, branches, and twigs. The vast variety of living species we observe today is depicted by canopy of leaves. Darwin described it as "*the great Tree of Life, which adorns the Earth's surface with its continuously branching and captivating ramifications, filling it with the remnants of its past and present*" (Figure 1A)³. This metaphor beautifully conveys the interconnectedness and diversity of life on our planet.

However, this traditional view of evolution was challenged when Scientists found ample evidence of horizontal gene transfer (HGT), where genes could be exchanged even between organisms⁴. Since the discovery of penicillin, HGT has contributed to the spread of bacterial resistance genes against *aminoglycosides*, *macrolides*, *glycopeptides*, *tetracyclines*, β -lactams and *fluoroquinolones*. Bacteria employ various methods of HGT to acquire resistance genes collectively, with specific species often specializing in particular mechanisms (Figure 1B)⁵. For instance, within the *Enterobacteriaceae* family, resistance genes are commonly transferred via conjugation, which involves the direct contact-mediated exchange of genetic material. On the other hand, among *Streptococci*, *Neisseriaceae*, and *Haemophilus*, resistance is predominantly disseminated through natural transformation, which involves the uptake of genetic material from the external environment. In *Staphylococci*, resistance genes can be spread through both transduction (genetic material transfer by viruses) and conjugation. This phenomenon challenges the Darwinian concept of strictly vertical gene flow, which postulates that genes are transmitted solely from parents to offspring and cannot be laterally transferred between different species. Science writer David Quammen aptly described this by saying that it "has overturned the traditional certitude that genes flow only vertically, from parents to offspring, and can't be traded sideways across species boundaries⁶." As a result, the tree metaphor, though still valuable, is incomplete. It fails to represent the complexity of genetic exchanges that can occur between species. The evolutionary relationship among organisms is a tangled network - the branches may connect through horizontal gene transfer, where genes can jump from one species to another. Biologists such as Richard Dawkins and Jerry Coyne assert that while the concept of the tree of life is not fundamentally flawed, it is now recognized to be more intricate and multifaceted than previously comprehended, especially with the advancements in molecular genetics⁷.

Antimicrobial Resistance and Avoidance of "E" word

In recent years, the threat posed by microbial resistance has become increasingly concerning. Developed countries' public health authorities are particularly worried about infections caused by drug-resistant strains of common pathogens like *Staphylococcus aureus* and *Mycobacterium tuberculosis*. These strains have evolved in response to the selective

pressures created using antibiotics (Figure 2A). Hospitals, unfortunately, have become breeding grounds for these multi drug resistant bacteria, making hospitalization a potential risk for acquiring difficult-to-treat infections⁸.

The rise and spread of these resistant microbes can also be seen from the Darwinian natural selection (survival of the fittest) in the face of environmental challenges. Bacteria could develop resistance to antibiotics through genetic mutations that change the targets of these drugs within their cells, or by acquiring specialized resistance genes from other bacteria. Once a bacterial population become resistant, they can thrive and multiply in the presence of antibiotics, effectively neutralizing or eliminating their susceptible counterparts. Nonetheless, this resistant population of bacteria qualifies the basic principles of Darwinian natural selection. Hence by understanding the principles of evolution and natural selection, we can better appreciate the formidable challenges posed by microbial resistance and work towards finding sustainable solutions to protect human health.

However, in the light of horizontal gene transfer and the ongoing debates surrounding Darwinian evolution, the use of the term "evolution" within the medical community remains somewhat tenuous. In an essay titled "*Evolution by Any Other Name: Antibiotic Resistance and Avoidance of the E-word*," published in PLoS Biology, the preference for alternative terms like "emerging," "arising," and "spreading" when discussing antibiotic-resistant microbes was highlighted¹¹. Despite being one of the most extensively researched examples of evolution, the antibacterial resistance observed in pathogens is often described as "emergence" in almost all the prestigious biomedical journals. Despite the recognition of antibiotic resistance evolution, only a limited number of individuals in the biomedical community recognizes the competition among bacteria that has influenced the resistance factors, sparking an ongoing arms race against antibiotics (Figure 2B).

The Darwin Dilemma

Genetic mutations play the main role in the development of antibacterial resistance. According to neo-Darwinian theory, natural selection and genetic mutations work together as evolutionary mechanisms. Certain mutations can have profound effects, leading to significant changes in essential proteins that can severely impair an organism's ability to survive. On the other hand, some mutations do not have functional consequences. These mutations are considered adaptively neutral, as they neither improve nor worsen the native gene form. However, there are mutations that confer clear advantages to bacteria, and unfortunately these advantageous mutations are the primary drivers of antibacterial resistance. Upon considering the concept of horizontal flow of genes, it raises the basic question that does horizontal gene transfer really challenges the Darwinian evolution in light of antimicrobial resistance? Does the acquisition of resistance through horizontal gene transfer alone provide insights into the underlying mechanisms of antibiotic resistance?

The emergence of bacterial resistance to antibiotics is not driven by the organisms' immediate adaptive requirements. Mutations that confer resistance are not purposefully directed towards fulfilling the microbe's immediate adaptive needs. Instead, these mutations have underlying causes

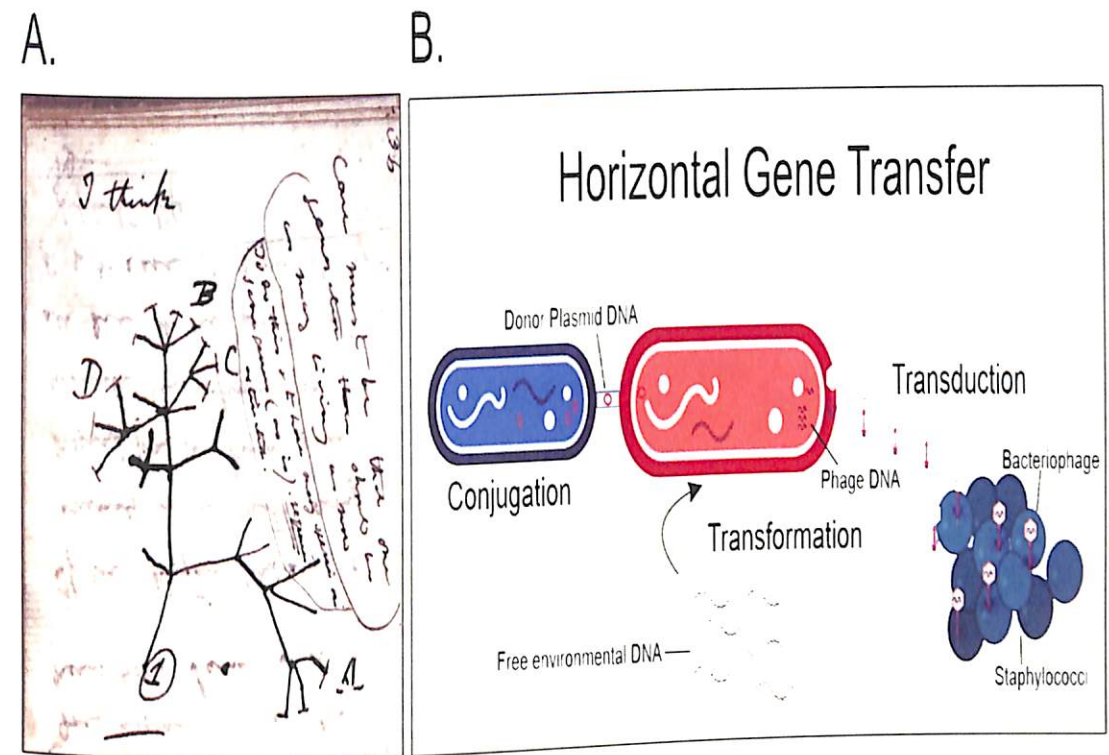


Figure 1- Darwin's tree of life was incomplete but not false: A) Darwin's preliminary sketch from his notebook titled "Transmutation of Species" created between 1837-1838². This early depiction shows a rough and irregular presentation compared to the refined and structured version depicted in "On the Origin of Species" (1859). B) Representational figure on how bacteria acquires genetic material through gene transfer.

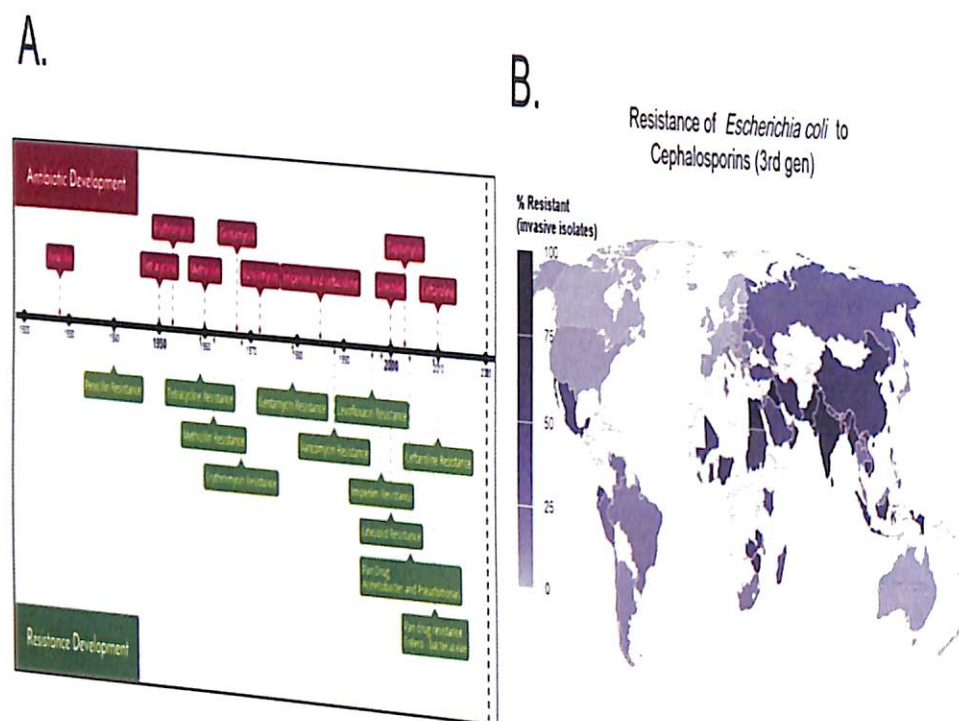


Figure 2- Rise of antibiotic resistance: A) Chronological presentation of antibiotic discovery (highlighted in red boxes) and the development of antibiotic resistance (highlighted in green boxes) over time. The figure is adapted from Zainab et al.⁹ B) Global overview of *Escherichia coli* resistance against the antibiotic cephalosporin as of 2020¹⁰.

unrelated to the species' need to adapt. In fact, the presence of resistance-conferring mutations in bacterial populations predates the discovery and use of antibiotics. Lederberg's 1952 experiments on streptomycin-resistant bacteria demonstrated that the mutations responsible for resistance were already present in bacteria that had never been exposed to antibiotics¹². Furthermore, Bowden in 1991 noted that bacterial cultures frozen 140 years ago exhibited resistance to antibiotics that were developed 100 years later¹³. This suggests that the natural occurrence of the specific chemical need for resistance was inherent in bacteria and did not arise through mutational events. Hence, bacteria do not necessarily mutate to become resistant to antibiotics, and non-resistant varieties do not acquire resistance only through mutations.

Furthermore, pre-existing mutations can confer antibiotic resistance while simultaneously decreasing an organism's viability. As an example, surviving bacterial strains that develop resistance are often less virulent and exhibit reduced metabolism, resulting in slower growth. It is important to note that the acquisition of a specific trait through mutation does not necessarily imply a complete transformation of the entire organism. A well-known case is sickle cell anaemia in humans, where individuals carrying this mutation are resistant to malaria, which may initially seem like a beneficial adaptation. However, the story doesn't end there. While these individuals may be resistant to malaria, they lack vitality and have shortened lifespans, which is the cost of such resistance. Similarly, bacteria can develop resistance to antibiotics, but this resistance comes at a price. In the broader context, acquiring resistance does not automatically lead to the emergence of new species or types of organisms.

Regardless of the mechanism through which bacteria acquire antibiotic resistance (whether it be transformation, conjugation, transduction) they remain the same bacteria they were before and after obtaining this "resistant" characteristic. Evolution, in this context, is not characterized by vertical macroevolution but rather by horizontal microevolution, which encompasses adaptation. Simply put, these bacteria "still belong to the same bacterial type and are only a variant that differs from the norm due to their antibiotic resistance. No new species have been generated. Bacteria are not capable of arbitrarily creating significant changes such as the development of chlorophyll or flagella. The key characteristics of bacteria are exclusive to a well-defined group, and they remain same, as is observed in other groups of living organisms.

These arguments, taken collectively, suggest that the existence of horizontal gene transfer alone is insufficient to invalidate the principles of Darwinian natural selection.

Conclusion

Science continually challenges our existing knowledge and compels us to reassess the foundational assumptions upon which it is built. Microbiology, like any other scientific field, often necessitates the revision of previously established information. This dynamic process embodies the essence of knowledge advancement. The relationship between Darwinian evolution and antimicrobial resistance is no exception to this notion. While philosophical debates persist regarding the merits and criticisms of Darwinian evolution, the pertinent question is scientific in nature: how can we effectively combat antibiotic resistance? While it is possible to

reconfigure the *Tree of Life* proposed by Darwin, the emergence of highly resistant bacteria poses the main challenges in terms global health. Insights from Darwin's ideas, as well as other sources, should be embraced and explored as we strive to address this pressing issue. By fostering a multidisciplinary approach, we can pool together various perspectives and knowledge to develop comprehensive strategies to combat antibiotic resistance effectively.

References

1. Kutschera, U. Charles Darwin's Origin of Species, directional selection, and the evolutionary sciences today. *Naturwissenschaften* 96, 1247–1263 (2009).
2. Kilian, P. Of Trees and Genealogies: A Foucauldian Commentary on Franco Moretti. *Le foucaldien* 2, (2016).
3. Lawton, G. Why Darwin was wrong about the tree of life. *New Sci.* 2692, 34–39 (2009).
4. Wiedenbeck, J. & Cohan, F. M. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol. Rev.* 35, 957–976 (2011).
5. Barlow, M. What antimicrobial resistance has taught us about horizontal gene transfer. *Methods Mol. Biol.* 532, 397–411 (2009).
6. Dietrich, R. F. The Tangled Tree: A Radical New History of Life. (2020).
7. Truman, R. The problem of information for the theory of evolution. *system* 2, 3 (2001).
8. Bumham, C. A. D., Leeds, J., Nordmann, P., O'Grady, J. & Patel, J. Diagnosing antimicrobial resistance. *Nat. Rev. Microbiol.* 2017 151115, 697–703 (2017).
9. Zainab, S. M., Junaid, M., Xu, N. & Malik, R. N. Antibiotics and antibiotic resistant genes (ARGs) in groundwater: A global review on dissemination, sources, interactions, environmental and human health risks. *Water Res.* 187, 116455 (2020).
10. ResistanceMap - Antibiotic Resistance. <https://resistancemap.onehealthtrust.org/AntibioticResistance.php>.
11. Antonovics, J. et al. Evolution by Any Other Name: Antibiotic Resistance and Avoidance of the E-Word. *PLOS Biol.* 5, e30 (2007).
12. Microbial Threats, I. of M. (US) F. on. Antibiotic Resistance: Origins and Countermeasures. (2009).
13. Aminov, R. I. A brief history of the antibiotic era: Lessons learned and challenges for the future. *Front. Microbiol.* 1, 8040 (2010).

Silibinin loaded xanthan gum nanoparticles and its effect as an *in-vitro* anticancer agent

Plabita Gogoi

Department of Chemical Sciences,
Tezpur University, Assam, India
Email: plabita41@gmail.com

Abstract

Silibinin loaded xanthan gum coated nanoparticles were prepared by ionic gelation method using calcium chloride (CaCl_2) as a precipitating agent. Phase separation behavior of xanthan gum was carried out as a function of CaCl_2 . The prepared nanoparticles were further chemically crosslinked with genipin. Effect of amount of surfactant (*Tween80*) and crosslinker (*Genipin*) on different properties of the nanoparticles were investigated. The nanoparticles were characterized by Fourier Transform Infrared spectroscopy (*FTIR*), X-ray diffractometry (*XRD*), Scanning Electron Microscopy (*SEM*) and Transmission Electron Microscopy (*TEM*). The average particle size of the nanoparticles was studied by Dynamic Light Scattering (*DLS*) method. Swelling of nanoparticles and release rate of silibinin were determined at pH 5 and intestinal pH 7.4 and were found to be better in intestinal pH. The drug release profile was studied during different time periods using a UV-visible spectrophotometer. The anticancer effect of the nanoparticles were finally evaluated by MTT [*3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide*] assay, Cell count assay and Clonogenic cell survival assay on MDA-MB-231 breast cancer cell line.

1. Introduction

Cancer is a dominating health issue and is leading the cause of death around the globe. Different drugs are being used in the treatment to cure cancer. Among those, reports have been found to use silibinin (*Bhattacharya2010*), an active constituent of silymarin, extracted from the medicinal plant *silybum marianum* an anticancer agent in the treatment of various types of cancer like breast, lung, colon, brain, prostate, skin, gastric, ovarian (*Singh et al. 2006*;

P. Gogoi, "Silibinin loaded xanthan gum nanoparticles and its effect as an *in-vitro* anticancer agent", Horizon (Ed. Book). 29-50(2023), copyright: Tinsukia College Publishing. ISBN: 978-81-19322-07-7

Singh et al. 2008; Kim et al. 2009; Singh et al. 2004; Deep & Agarwal, 2010; Gao et al. 1995). People from the age old days discovered its anti hepatotoxic nature and also used it as a medicine in the treatment of jaundice, haemorrhage, bronchitis, inflammation and oxidative stress (Xu et al. 2013). The main disadvantages of silibinin are its hydrophobicity and low bioavailability. These properties can be improved by encapsulating it within polymer. Xanthan gum, a natural hetero polysaccharide secreted by the bacterium *Xanthomonas campestris* was recently used to encapsulate varieties of drug (García-Ochoa et al. 2000). It consisted of repeated penta saccharide units formed by two glucose units, two mannose units and one glucuronic acid unit, in the molar ratio 2:2:1 (Fig. 1) (Bueno et al, 2013).

Its application has been observed in various industries such as food, agriculture, oil, paper, paint, pharmaceuticals and cosmetics (Sharma et al. 2006; Flickinger & Draw 1999; Katzbauer 1998; Rosalam & England 2006). In the present scenario of pharmaceutical sciences, the use of natural polymers is flourishing at a massive rate due to its negligible virulence, cell biocompatibility, easy and cheap availability (Mohd Zobir et al. 2001). Reports have been found to prepare xanthan gum stabilised gold nanoparticles (Pooja et al. 2014). Conventional chemotherapy, a path for cancer treatments, has considerable side effects (Yoo et al. 2005) as it affects the cancer tumours as well the normal healthy lymphocytes. When a drug is encapsulated or distributed in a polymer, its toxicity is reduced to some extent as the unified amount of drug is not released at a time. Nanotechnology is being used in drug delivery applications due to its various advantages (Mudshinge et al. 2001). Nanoparticles are prepared by various techniques like coacervation (Kizilay et al. 2011), emulsification (Yuan et al. 2007), desolvation (Teng et al. 2013), ionic gelation (Mi et al. 2003) and spray drying (Broadhead et al. 1992). Ionic gelation method was picked to prepare silibinin loaded xanthan gum nanoparticles as silibinin is poorly soluble in water. To govern the distinct behaviours of the nanoparticles like swelling, release and particle size, different excipients are used in the preparation of polymeric nanoparticles. Genipin, a natural cross linker originated by hydrolysis of geniposide and obtained from the extraction of gardenia fruit, was used as a chemical cross linker as cited in literature (Muhamad et al. 2011; Yuan et al. 2007). The use of CaCl_2 as a physical cross linker was also reported (Banik et al. 2014). Tween 80, a non-ionic surfactant was used to reduce the agglomeration and size of the nanoparticles (Gogoi et al. 2018).

There are few reports regarding the use of xanthan gum for controlled release applications. Thus there is a wide scope to do further work in this area. Keeping this in mind, the present work has been undertaken. The aim of this work is to prepare silibinin loaded xanthan gum nanoparticles by using both physical and chemical crosslinker, surfactant. The effect of variation of crosslinker and surfactant on various properties of prepared nanoparticles has also been studied and reported. An optimised formulation was favoured further to study the anticancer activities.

2. Experimental

2.1. Materials

Polymeric matrix Xantan Gum was purchased from Hi Media (Hi Media Laboratories Pvt. Ltd, Maharashtra, India). Silibinin (MW: 482.44g/mol) was purchased from Sigma Aldrich, U.S.A. Genipin (MW: 226.23g/mol), was bought from CBC, Taiwan. Tween 80, ethanol, calcium chloride were purchased from E-Merck, India. Phosphate buffer saline (PBS), RPMI 1640 and fetal bovine serum (FBS) were procured from HiMedia Laboratories (Mumbai, India). The rest of the chemicals used were of analytical grade.

2.2. Methods

2.2.1. Phase separation behaviour of xanthan gum

To optimise the phase separation behaviour of xanthan gum, a series of experiments were carried as a function of CaCl_2 concentration. Beaker containing 1% (w/v) aqueous solution of xanthan gum was stirred at room temperature for 2 h. A predetermined amount of aqueous CaCl_2 solution was added drop wise to the polymer solution under stirring condition followed by sonication. The optimised ratio at which phase separation occurred was determined by varying the ratio of polymer to CaCl_2 and also by varying the concentration of CaCl_2 (Table 1).

2.2.2. Preparation of silibinin loaded xanthan gum nanoparticles

Nanoparticles were prepared using xanthan gum following a simple coacervation method as reported in the literature by Tavakoli et al. (Tavakoli et al. 2013) with few modifications. A 1% (w/v) aqueous solution of xanthan gum was stirred for 2 h at room temperature in a beaker. Tween 80 was added as surfactant and the solution was sonicated for 20 min. Silibinin (0.01g) was dissolved in 1:1 ethanol water mixture and added into the solution. After 2 h, CaCl_2 (10% w/v) was added drop wise as precipitating agent to the solution under stirring condition and sonicated for 30 min. Then the reaction mixture was cooled to 0-10 °C in order to harden the nanoparticles and genipin was added to the reaction mixture to crosslink the nanoparticles. The reaction temperature was gradually increased to 35-40 °C and it was kept for 3 h in stirring condition till the colour of the reaction mixture changes from whitish to light pinkish colour. The precipitate was separated by centrifuging the reaction mixture at 8000 rpm for 20 min at 4 °C and washed with distilled water. The washing process was repeated several times so as to completely wash off the salt. The precipitate was freeze dried and kept in glass vial for further use. A series of 5 sets were prepared by varying the concentration of the surfactant (1%, 3% and 5%) and cross-linker (3%, 5% and 7%) as tabulated (Table 2) (Fig. 2).

2.3. Process yield

Process yield was calculated using the following equation (Gupta & Kumar 2000).

$$\text{Process yield (\%)} = \frac{\text{Weight of the nanoparticles}}{\text{Weight of (Drug + Polymer)}} \times 100$$

Table 1. Ratio of Xanthan gum and CaCl_2

Xanthan gum % (w/v)	CaCl_2 ratio w.r.t. xanthan gum (CaCl_2 % w/v)	Yield %
1	5(5)	71.83(± 0.08)
1	10(5)	68.26(± 0.52)
1	15(5)	63.42(± 0.36)
1	10(10)	76.49(± 0.74)
1	15(10)	69.57(± 0.66)

Table 2. Preparation and effect of silibinin loaded xanthan gum nanoparticles with the variation in concentration of surfactant and crosslinker.

[Xanthan gum % (w/v) (amount in gm in 50 mL water): 1(0.5), Silibinin (g): 0.01, CaCl_2 : 5g in 50 mL]

Sample code	Tween 80% (v/w) w.r.t. polymer (amount in mL)	Genipin % (w/w) w.r.t. polymer (amount in g)	Process Yield (%)	Drug Loading Efficiency ^a (%)	Particle size ^b (nm)	Zeta Potential ^c (mV)
A0	1(0.005)	0(0.000)	56.84(± 0.07)	9.06(± 0.03)	108.9(± 7)	-44.60(± 0.01)
A1	1(0.005)	3(0.015)	83.21(± 0.03)	33.21(± 0.02)	54.16(± 6)	-37.82(± 0.03)
A2	3(0.015)	3(0.015)	82.63(± 0.05)	35.50(± 0.02)	52.45(± 14)	-36.40(± 0.03)
A3	5(0.025)	3(0.015)	83.46(± 0.02)	41.74(± 0.01)	51.33(± 10)	-34.91(± 0.04)
A4	5(0.025)	5(0.025)	83.57(± 0.02)	43.02(± 0.03)	47.27(± 11)	-32.37(± 0.03)
A5	5(0.025)	7(0.035)	85.83(± 0.03)	43.95(± 0.01)	43.04(± 6)	-29.23(± 0.06)
BF	5(0.025)	3(0.015)	76.03(± 0.04)	-	53.61(± 4)	-36.72(± 0.03)

Notes: Silibinin: 0.01 g in 10 mL solvent, CaCl_2 : 5g in 50 mL.

^a, ^b, ^c: The value represents average of three readings, standard deviation in parenthesis.

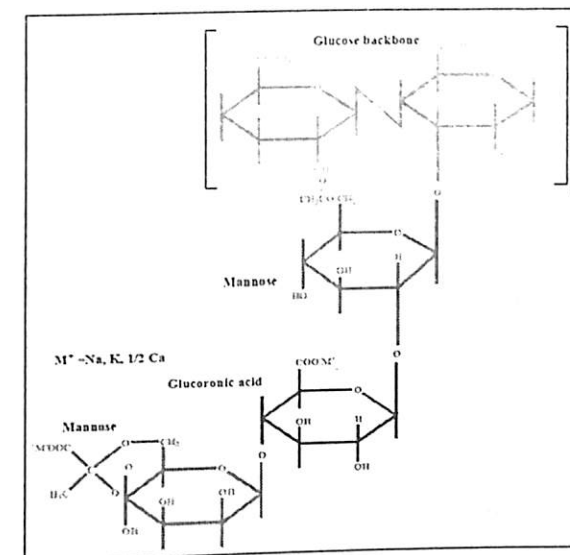


Fig. 1: Chemical structure of xanthan gum.

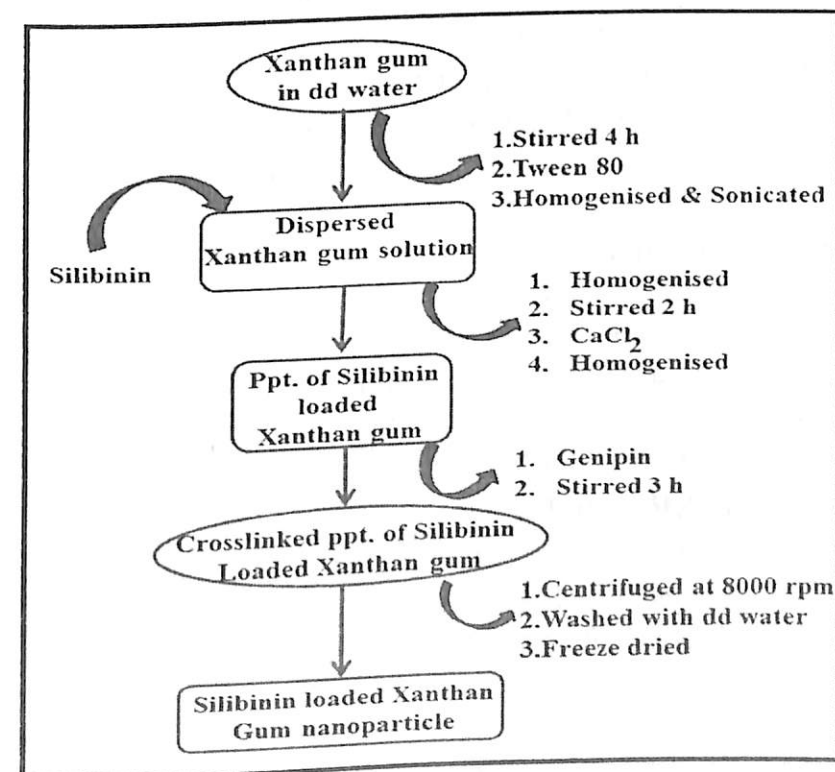


Fig. 2: Preparation of silibinin loaded xanthan gum nanoparticles.

2.4. Calibration curve of silibinin

A calibration curve is essential to estimate the release rate of drug from nanoparticles in the suitable solvent medium.

A known concentration of silibinin was scanned in the range 200-400 nm by using UV-Vis spectro-photometer (UV-2001 Hitachi, Tokyo, Japan). A calibration curve was determined with UV (Selvaraj *et al.* 2010).

From this calibration curve, the unknown concentration of silibinin was obtained from the absorbance value at 285 nm.

2.5. Calculation of drug loading efficiency of the nanoparticles

In order to study the drug loading efficiency and encapsulation efficiency of nanoparticles with different formulations, samples were immersed in ethanol : water (1:1) mixture and sonicated for 2 min. Ultra centrifugation of the samples was done at 8000 rpm at 4°C. The amount of free silibinin was determined by noting the absorbance value of the supernatant liquid at 285 nm using a UV-Vis spectrophotometer. The drug loading efficiency (EE) of the nanoparticles was calculated using the following equations (Devi & Maji 2010).

$$\text{Drug Loading efficiency (EE)\%} = \frac{(\text{Total amount of drug} - \text{Free amount of drug})}{\text{Total amount of drug}} \times 100$$

2.6. Fourier Transform infrared spectroscopy (FTIR) study

FTIR analysis was done in order to study the chemical interaction between xanthan gum and silibinin. FTIR spectra of the prepared nanoparticles were taken in a Nicolet (model Impact-410, USA) spectrophotometer. The nanoparticles were grounded, mixed with KBr and compressed, and the spectra were recorded in the range of 4000-400 cm^{-1} .

2.7. Particle size determination

By using a DLS analyser (model DLS- Nano ZS, Zetasizer, Nanoseries, Malvern Instruments, UK), particle size of the prepared nanoparticles was determined. The experiment was performed in distilled water using a quartz cell at 25 °C with a detection angle of 90°. The suspension of the nanoparticle measurements was repeated thrice for a single nano formulation and the results were the average of the three measurements.

2.8. X-Ray diffraction (XRD) study

In a Rigaku X-ray diffractometer (Miniflex, UK) using $\text{CuK}\alpha$ ($\lambda=0.154$ nm), X-ray diffraction (XRD) study was carried out at a scanning rate of 1°/min with an angle ranging from 10° to 70° of 2 θ to determine the crystallinity distribution of xanthan gum and silibinin in the polymer matrix.

2.9. Thermal analysis

Differential Scanning Calorimetry (DSC) of silibinin, xanthan gum and polymeric nanoparticles were performed using a DSC-50 (Shimadzu DSC-50, Japan).

2.10. Scanning electron microscope (SEM) study

To study the surface morphology of the prepared nanoparticles, the samples were spread in a brass holder and sputtered with platinum. The scanning was carried out in an instrument (JEOL JSM – 6390, Japan) at an accelerated voltage of 20 kV.

2.11. Transmission electron microscope (TEM) study

TEM was completed using a JEOL JEM- 2100 system operating at 100 kV for determination of the size and dispersion of the nanoparticles.

2.12. Swelling studies

Swelling studies were performed in two different pH solution having pH 5 and pH 7.4. Nanoparticles (0.05 g) were taken in a pouch made of nylon cloth. The empty pouch was first conditioned by immersing it in any of the two pH mediums for different time periods [1-48 h]. The pouch containing the nanoparticles was immersed in a similar way in any of the three pH medium for similar time periods. The weights of wet nanoparticles after a definite time period were determined by deducting the respective conditioned weight of the empty nylon pouch from this. Swelling (%) was determined by measuring the change in the weight of the nanoparticles. The percentage of swelling for each sample determined at time 't' was calculated using the following equation (Cassano *et al.* 2012).

$$\text{Swelling (\%)} = \left[\frac{w_2 - w_1}{w_1} \right] \times 100$$

Where ' w_1 ' is the initial weight of the nanoparticles before swelling and ' w_2 ' is the final weight of nanoparticles after swelling for a predetermined time 't'.

The experiments were performed for five times and are represented as a mean value.

2.13. In vitro drug release studies

To study the release profile of the silibinin loaded xanthan gum nanoparticles, dried samples were weighed and immersed in beakers containing different pH values, viz. 5.0 and 7.4, and magnetically stirred continuously. At a specific time interval, 5 mL solution was withdrawn, filtered and assayed spectro-photometrically at 285 nm using a UV-Vis spectrophotometer for the determination of the cumulative amount of drug release upon a time t . Each determination was carried out five times. To maintain a constant volume, 5 mL of the fresh buffer solution having the same pH was returned to the beaker (Das *et al.* 2008).

2.14. Comparative cytotoxicity analysis of silibinin, silibinin loaded nanoparticles and blank formulation

The cytotoxicity study was done on human lymphocytes and MDA-MB-231 breast cancer cell line.

Human blood cells were anti-coagulated using Histopaque (1.077 g/mL) and lymphocytes were delicately isolated from it. Aliquots of 200 mL of the isolated cells were further cultured

in RPMI supplemented with 10% heat inactivated FBS. At first cells were maintained in RPMI for 4 h without FBS at 37 °C in 5% CO₂ in an incubator. They were then treated with the drug and nanoparticles and maintained in presence of FBS for 12 h.

MDA-MB-231 cell line was grown in DMEM supplemented with 10% FBS, 100 mg/mL penicillin-streptomycin antibiotics. Cell line was maintained at 37 °C without CO₂, in a humidified incubator and was sub cultured when they reached 80–90% confluence.

The effect of nanoparticles on the proliferation of cell lines were tested in 96-well microtiter plates, using tetrazolium based semiautomated colorimetric [*MTT*- (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide)] assay. Cells at the logarithmic growth phase were harvested and seeded at a density of 1×10^3 cells per well in 96-well plates and incubated for proper adherence. Subsequently, the drug-free medium was replaced with fully supplemented fresh medium, containing different concentrations of prepared nanoparticles (25, 50 and 100 µg/mL). Concentration of raw silibinin (10, 20 and 40 µg/mL) was compared with respect to loading efficiency of the nanoparticle. The cells were incubated for 48 hours, while maintaining the physiological conditions at 37°C temperature, in a humidified atmosphere. At the end of the treatment period (12 h for normal human lymphocytes, 48 h for cancer cell lines), 10 µL of MTT solution was added to each well. After incubation for 4 h at 37°C, 200 µL of solubilizing buffer (DMSO) was added to each well. The viable cells produced a dark blue/purple formazan product, whereas no such staining was formed in the dead cells. The optical densities (OD) of cells treated at different concentrations were measured at 570 nm/690 nm using a microplate reader and cell viability was calculated as a measure of the optical density (OD) of treated cells relative to the optical density of the untreated controls. A minimum of three independent experiments was performed for the nanoparticles. Dose-dependent response curve was plotted for the samples.

2.15. Comparative study of Trypan blue exclusion assay of silibinin, silibinin loaded nanoparticles and blank formulation

To determine the antiproliferative effect of the samples, the MDA-MB-231 cells were first seeded (1×10^6 cells in L-15 media) in 60 mm tissue culture plates. After incubation for 48 h to allow cell attachment, the exponentially growing cells were exposed to the samples at concentrations described above. The plates were then incubated at 37 °C in the absence of CO₂ for 48 hours. After incubation, the medium was aspirated; the plates were washed with cold phosphate-buffered saline to remove dead cells, and replenished with 1 mL of 0.05% (2 mg/mL) *trypsin-ethylenediaminetetraacetic acid*. The plates were then incubated at 37°C for 3–5 minutes, until the majority of the cells detached as microscopically confirmed. The cells were harvested and the cell suspension was centrifuged at 2000 rpm for 5 minutes and the supernatant discarded. The 100 µL cell suspension was then mixed with 100 µL of a 0.4% trypan blue solution, the cells were subsequently resuspended and dye-excluding viable cells were microscopically counted using a hemocytometer chamber.

2.16. Clonogenic cell survival assay

The sensitivity of breast cancer cell line MDA-MB-231 to prepared nanoparticles was measured using the clonogenic assay, the method of choice used to determine the effectiveness of cytotoxic agents such as chemotherapy and determine cell reproductive death after treatment. Clonogenic assay or colony formation assay is an *in vitro* cell survival assay based on the ability of a single cell to grow into a colony. Cells were allowed to adhere overnight at 37°C and treated with different concentrations (25, 50 and 100 µg/mL) for 9–14 days. Colonies were fixed and stained crystal violet (0.05% w/v) for 30 min after which time residual staining solution was removed and plates were washed with water. Colonies consisting of 50 cells or more were counted.

Plating efficiencies (PE) were calculated using the formula:

$$PE = (\text{Number of colonies}) / (\text{Number of cells seeded}).$$

The surviving fraction (SF) was calculated using the formula:

$$SF = (\text{Number of colonies} / \text{Number of cells seeded}) \times (PE).$$

Survival curves were constructed for determination of survival ability of drug treated cells relative to parent cells in response to various concentrations of the nanoparticles.

2.13. Statistical analysis

All the data were expressed as means $\pm$ SEM. Results were statistically analysed by student's t-test for significant difference between group mean. The significant difference between the experimental and the control group was set as $p < 0.05$ (c); $p < 0.01$ (b); $p < 0.001$ (a).

3. Results and discussion

3.1. Optimisation of phase separation of xanthan gum

The optimum ratio of xanthan gum to CaCl₂ (10% w/v) at which phase separation obtained was 1:10. This was concluded by the clear separation of xanthan gum in fine particle form from its aqueous solution (Fig. 3) also the yield was found to be maximum (Table 1). This ratio was continued during preparation of nanoparticle in the subsequent formulations. All experiments were carried out in triplicate and results presented were the average values.

3.2. Effect of variation of surfactant and crosslinker concentration on the different properties of silibinin loaded xanthan gum nanoparticles

It was found that surfactant had no significant role to play in process yield. But the loading efficiency increased with the increase in surfactant concentration. Since on addition of surfactant the nanoparticles became smaller and there would be more number of newly formed surface areas which helped to incorporate more drugs in the newly formed surfaces of the nanoparticle (Saikia et al. 2016). On increasing surfactant, the *zeta potential* also increased. This implied the increase in stability of nanoparticles. The average diameter of the nanoparticles as estimated by DLS data are found to be in the range of 40–60 nm. On increasing the surfactant concentration, the average diameter decreased because surfactant reduced the agglomeration of nanoparticles and made them smaller.

From Table 2, it was found that the drug loading efficiency of the nanoparticles greatly enhanced due to crosslinking. The compact structure of the nanoparticle resulting from crosslinking of polymer chain enabled them to encapsulate properly and load more amount of drug molecule. Hence on increasing the crosslinker concentration the loading efficiency increased. The yield (%) is also found to increase with increase in crosslinker percentage. It may be due to lower loss during isolation. The stability of nanoparticles increased due to crosslinking of the polymeric chains which prevented the nanoparticles from agglomeration and thus the zeta potential value increased. The *zeta potential* values are found to be in the range of -37.8 to -29.2 mV. Also on increasing crosslinker concentration, the average diameter of the nanoparticle decreased due to their compact and rigid structure (da TrindadeNeto et al. 2005)

3.3. FTIR study

To confirm the successful preparation of nanoparticles, the presence of reactive groups on the silibinin loaded nanoparticles were studied in the FTIR. The FTIR spectra of commercially purchased xanthan gum, silibinin and nanoparticles are shown in Fig. 4.

Fig 4(a) shows the spectra of pure xanthan gum. Peak at 3434 cm^{-1} appeared due to -OH stretching. Peak at 2924 cm^{-1} was due to C-H stretching. The peaks at 1631 cm^{-1} and 1418 cm^{-1} might be due to the C=O and -COO⁻ stretching respectively (Su et al. 2003).

Fig 4(b) shows the spectra of pure silibinin where the characteristic absorption due to the -OH stretching vibration occurred at 3456 cm^{-1} . The absorption band at 1635 cm^{-1} was due to C=O stretching. Peaks appeared in the region $1518\text{--}1466\text{ cm}^{-1}$ was due to the skeletal vibration of the aromatic C=C ring stretching. The characteristic peak at 1269 cm^{-1} was probably due to the C-O-C stretching of the of the diaryl group (Pooja et al. 2014).

Fig 4(c) shows the spectra of silibinin loaded xanthan gum nanoparticles. With the interaction of xanthan gum and drug, the FTIR spectrum showed broadening of the functional groups and shifting of band compared to that of the polymer and pure drug. The peaks in the nanoparticles (3440 , 2088 & 1642 cm^{-1}) were seen to be shifted slightly from the pure polymer and drug.

3.4. XRD study

Fig. 5 gives a comparison of XRD patterns of pure silibinin, xanthan gum and silibinin loaded Xanthan gum nanoparticles. The characteristic peaks of silibinin appeared at $2\theta = 14.5^\circ$, 17.2° , 19.6° , 22.2° and 24.3° suggesting its highly crystalline nature (Fig. 5a). The diffraction pattern of xanthan gum on the other hand is typical of the amorphous nature (Fig. 5b). The XRD diffractogram of silibinin loaded nanoparticles showed a decrease in intensity of the characteristic crystalline peaks of silibinin, suggesting a formation of crystalline disorder or amorphous nature of silibinin within the polymer (Fig. 5c) (Shankar & Agrawal 2015).

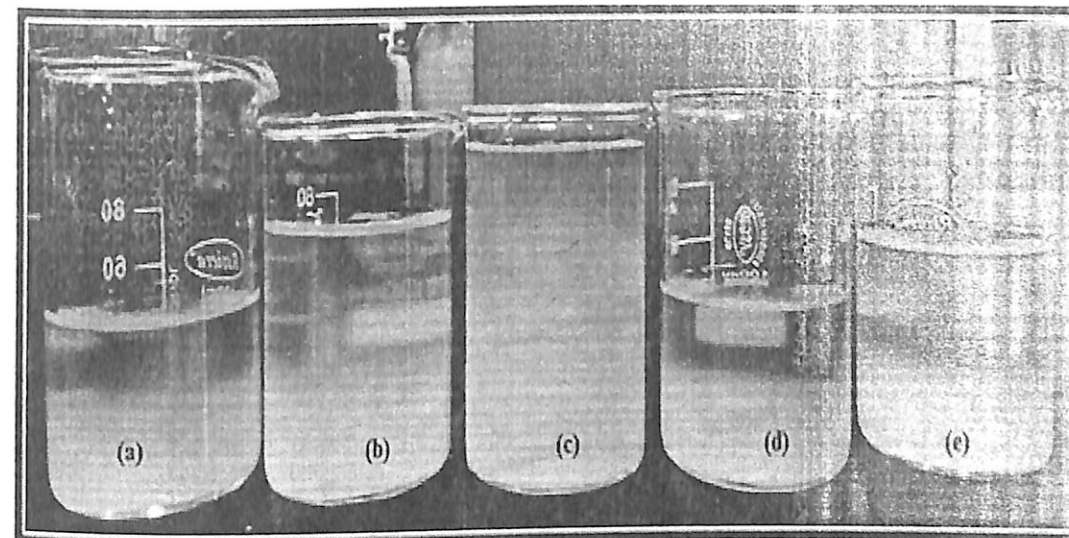


Fig. 3 Phase separation behaviour of xanthan gum. (a), (b) & (c) represents xanthan gum: CaCl_2 ratio of 1:5, 1:10 & 1:15 for 5% CaCl_2 and (d) & (e) represents xanthan gum: CaCl_2 ratio of 1:10 & 1:15 for 10% CaCl_2 respectively.

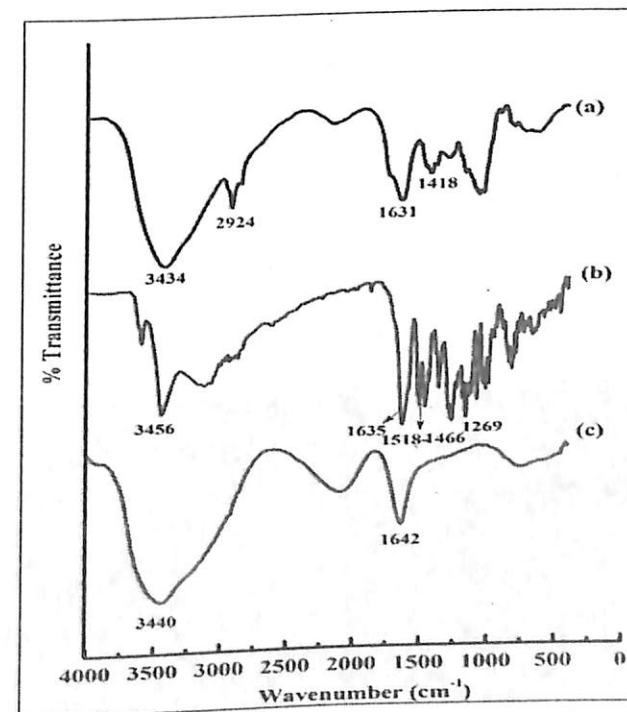


Fig. 4 FTIR spectra of (a) xanthan gum, (b) silibinin and (c) nanoparticle.

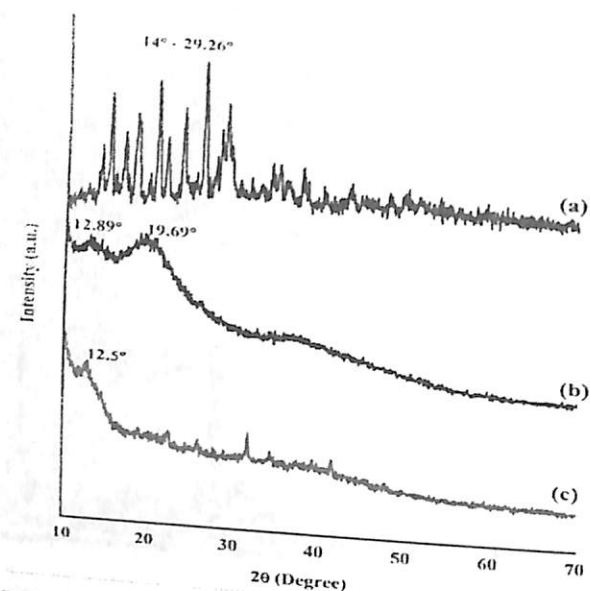


Fig. 5 XRD spectra of (a) silibinin, (b) xanthan gum and (c) nanoparticles.

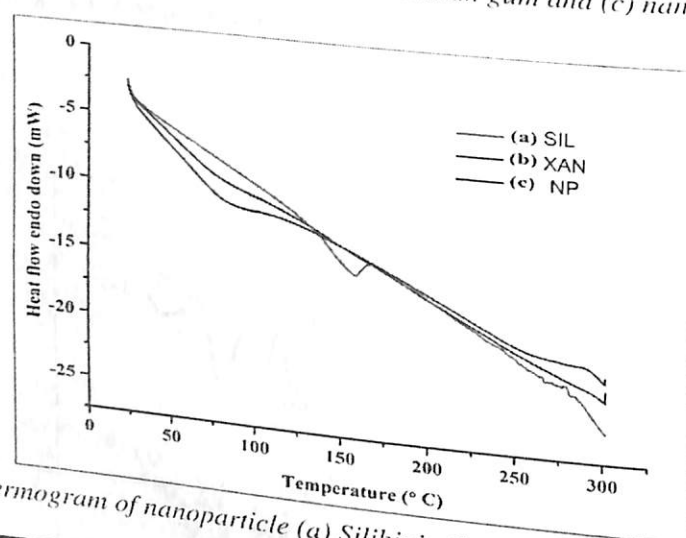


Fig. 6 DSC thermogram of nanoparticle (a) Silibinin (b) Xanthan gum and (c) nanoparticle

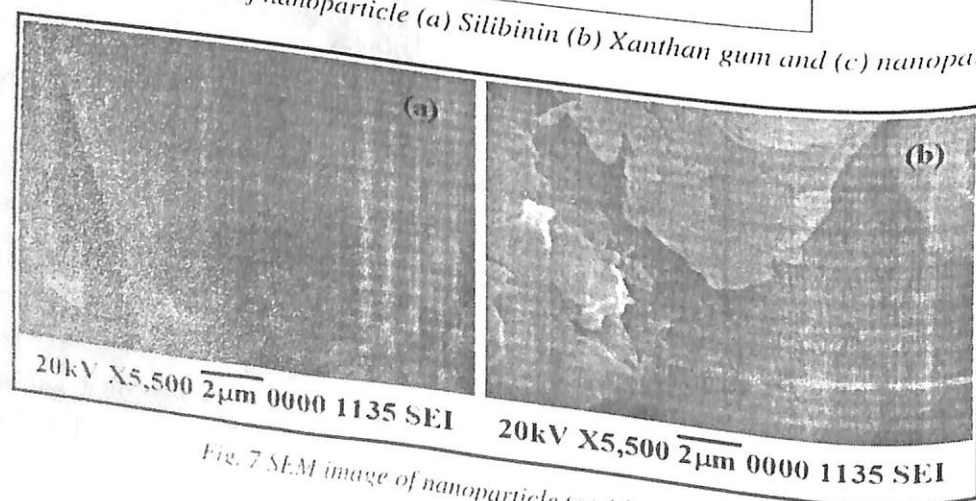


Fig. 7 SEM image of nanoparticle (a) A1 and (b) A5.

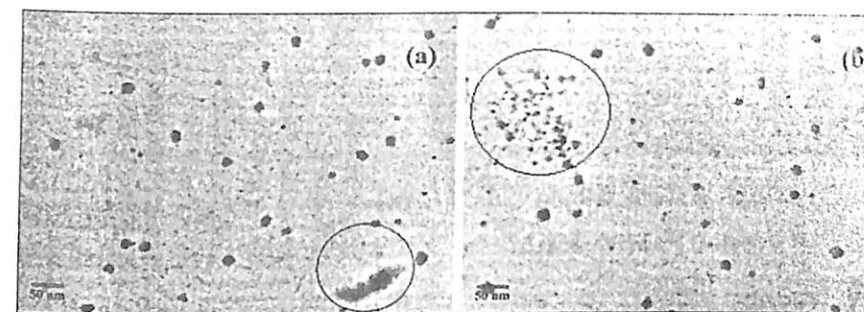


Fig. 8 TEM image of nanoparticle (a) A1 and (b) A5.

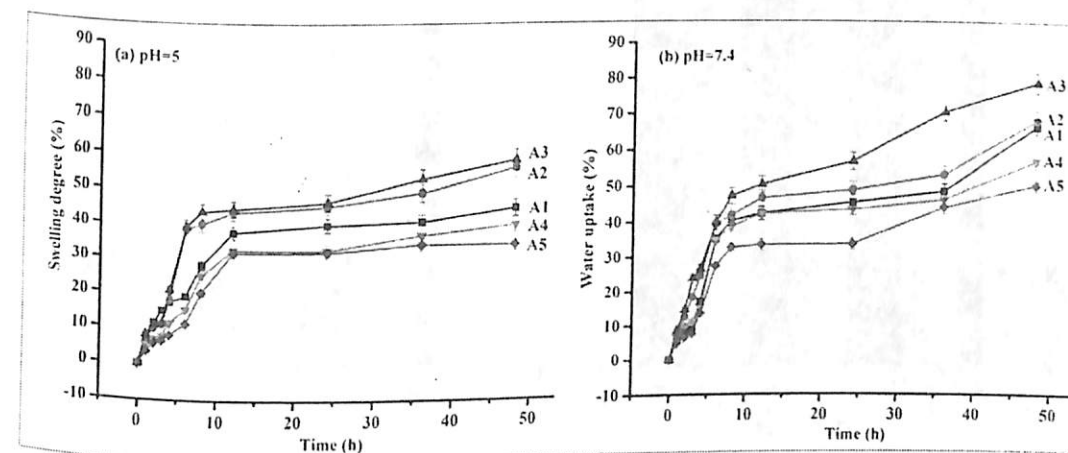


Fig. 9 Percentage swelling degree of nanoparticles A1, A2, A3, A4, A5 at (a) pH 5 and (b) pH 7.4.

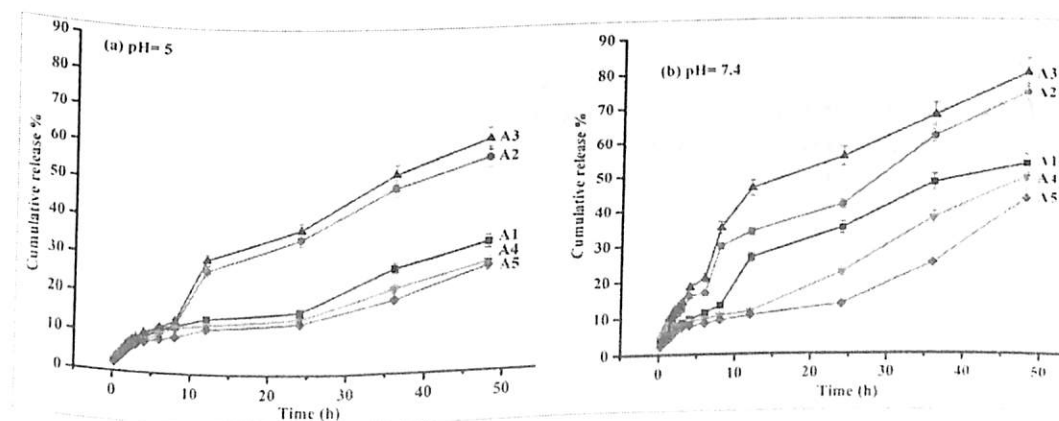
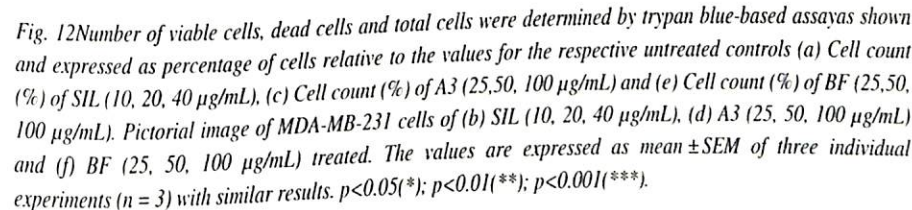


Fig. 10 Cumulative drug release of nanoparticles A1, A2, A3, A4, A5 at (a) pH 5 and (b) pH 7.4.



43



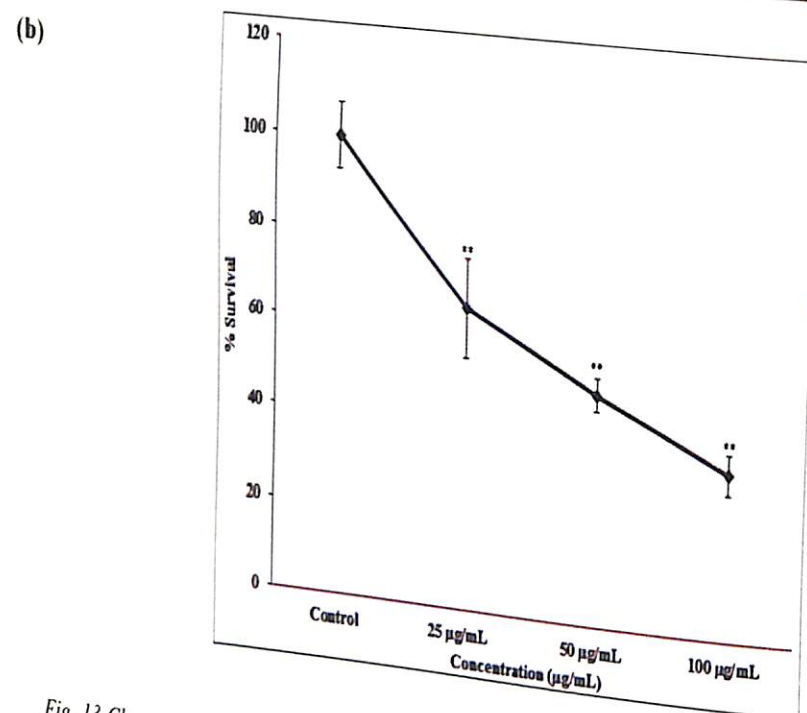
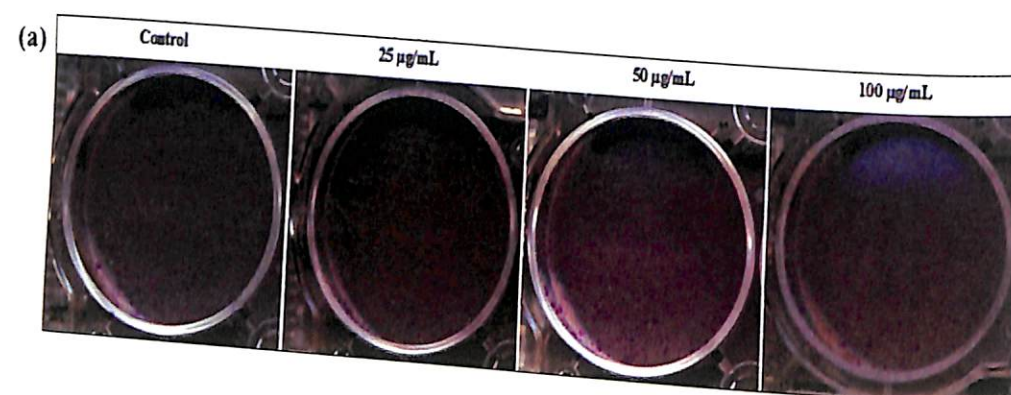


Fig. 13 Clonogenic cell survival assay of MDA-MB-231 cell lines performed in a 6 well plate using optimised seeding densities after exposure to A3 nanoparticle (25,50, 100 µg/mL) for 48 h. (a) Cells were fixed, stained with crystal violet, counted and imaged after 9 days. (b) Survival curve for determination of survival ability of treated MDA-MB-231 cells relative to untreated control cells. The values are expressed as mean $\pm$ SEM of three individual experiments ($n = 3$) with similar results. $p < 0.05$ (*); $p < 0.01$ (**); $p < 0.001$ (***)

3.5. DSC analysis

Fig. 6 represents the DSC thermograms of the silibinin, xanthan gum and the nanoparticle. DSC study showed a sharp endothermic peak corresponding to its m.p. at 150°C. Xanthan gum showed a broad peak at around 75°C showing its amorphous nature. The complete disappearance of melting point peak of silibinin in the nanoparticle indicated the amorphous nature of the drug as shown in Fig. 6.

3.6. Morphological studies

3.6.1. Scanning Electron Microscopy (SEM) study

SEM image revealed that the surface of the nanoparticles with minimum amount of surfactant and crosslinker (Fig. 7a) appeared to be homogeneously smooth than the nanoparticles with maximum amount of surfactant and crosslinker (Fig. 7b). In Fig. 7b, the surface morphology of the nanoparticles seemed to be irregular with a rough and a slight dense structure. Due to better crosslinking of the nanoparticles, the irregularity and roughness on the surface was more evident in the nanoparticles A5 with maximum amount of surfactant and crosslinker.

3.6.2. Transmission Electron Microscopy (TEM) study

Fig. 8 represents the TEM image of the silibinin loaded xanthan gum nanoparticles A1 (Fig. 8a) and A5 (Fig. 8b). Both the pictures resemble the shape of the nanoparticles to be almost spherical. But the image of A1 nanoparticles with least amount of surfactant and crosslinker were found to be agglomerated (shown within a circle). As the amount of surfactant and crosslinker were increased the process of agglomeration was decreased and the particles were thoroughly dispersed even with more fine sized particles (shown within a circle) (Fig. 8b). The size of the nanoparticles lies within a range of 10-20 nm. The average hydrodynamic diameter of the nanoparticles as obtained from DLS analysis was found in the range of 40-60 nm which was found to be slightly higher than the TEM analysis. The slight increase in the size of the nanoparticles as obtained in DLS data was due to the swelled state of the nanoparticles at which the DLS analysis was carried out.

3.7. Swelling study:

The effect of pH on the percentage swelling of xanthan gum nanoparticles at pH 5 and pH 7.4 are shown in Fig. 9. It was observed that the swelling of silibinin loaded xanthan gum nanoparticles was more in intestinal pH(7.4) than in tumor cell pH (5).

Under basic medium, the unreacted hydroxyl groups of xanthan gum became deprotonated and a repulsive force was generated between the adjacent negatively charged polymer chains causing the swelling of the polymer. Also in basic pH, the ester bonds became hydrolyzed. This was followed by rupturing and swelling of the nanoparticles (Bueno et al. 2013).

With the increase in surfactant concentration, particle size decreased. The lower the particle size, the higher was the surface area. This enhanced the contact area of nanoparticles

with the solvents and thus would increase the swelling. Similarly higher crosslinking concentration enhanced the crosslink density of the nanoparticles and hence decreased the swelling.

3.8. *In-vitro* release study

The drug release profile of the nanoparticles at two different pH namely 5 and 7.4 are shown in the Fig. 10. The cumulative release (%) of silibinin from silibinin loaded xanthan gum nanoparticles was found to be pH dependent. The major factor controlling the release profile of silibinin from nanoparticles was the swelling nature of the polymer. With increase in time, the drug release was found to be more in the intestinal pH than in acidic pH. At intestinal pH, the deprotonation of the unreacted hydroxyl groups occurred and a repulsive force was generated between the adjacent negatively charged polymer chains swelling increased. As a result the release also increased due to the formation of wobble structure of nanoparticles caused by swelling.

When the concentration of crosslinker increased, the drug release decreased due to the compact and rigid structure of the nanoparticle which became a palisade for the drug molecules to come out of the dense polymeric domain. Again, due to higher amount of surfactant the size of the nanoparticles decreased which in turn increase the surface area and accelerated the release of the drugs from the nanoparticles.

3.9. Cytotoxicity test

As described in the method section, MTT assay conducted on cell lines were spectrophotometrically identified when the colour of yellow tetrazolium salt turned to a purple intracellular formazan by active mitochondria. Absorbance recorded was converted to percent cell viability. The results represent the means of three independent runs with error bars denoting SEM.

The cytotoxicity of SIL, A3 and BF was studied using human lymphocyte and MDA-MB-231 cancer cell lines and are shown in Fig. 11(a) and Fig. 11 (b) respectively. It was found that SIL was not cytotoxic to normal human lymphocyte at any given concentration. Silibinin loaded A3 nanoparticle and the BF also did not reveal any cytotoxicity to the human lymphocytes.

It showed a dose dependent response against MDA-MB-231 cancer cell lines (Fig. 11 b). SIL when exposed to cancer cell line showed an increased cell growth inhibition from 6 % to 22% as the concentration of the doses increased from 10 to 40 $\mu\text{g/mL}$. The nanoparticle A3 at 50 $\mu\text{g/mL}$ showed a decrease in cell viability of 84% and a minimum of 80% cell viability at 100 $\mu\text{g/mL}$. The BF showed a negligible effect when exposed to the cell line.

There was no significant gain of less cell viability at higher concentration of blank formulation. The percentage of cell viability at 25 $\mu\text{g/mL}$ was 99.14% against 92.85% and 92.27 % at a concentration of 50 and 100 $\mu\text{g/mL}$ respectively. This could be attributed to the higher concentration of the excipient (Guo *et al.* 2011). But it had no significant effect on the viability of cells as compared to SIL or A3 that showed significant dose dependent cytotoxicity

. This was further validated by Trypan Blue Dye Exclusion assay which showed that BF had negligible effect on cell viability as compared to SIL or A3.

3.10. Cell count

As shown in Fig. 12 (e and f), blank formulation did not affect cell growth, even at high concentrations. However, as expected, both SIL and A3 nanoparticles Fig. 12 (a and c) exhibited a dose-dependent effect in cell proliferation of MDA-MB-231 cells. In comparison with the untreated cells, cell proliferation was significantly inhibited after 48 h of treatment. There was a significant dose dependent decrease in percentage of cell viability in cells treated with A3 nanoparticles as compared to SIL. Similarly both 25 and 50 $\mu\text{g/mL}$ doses of A3 nanoparticles showed a significant dose dependent increase in percentage of dead cells of 7.66 % and 17.01 % as compared to SIL of 7.40 % and 12.84 % respectively.

3.11. Clonogenic

Clonogenic assay (Fig. 13) was performed for evaluating the effect of silibinin loaded xanthan gum nanoparticles (A3) on long term viability and reproductive ability of MDA-MB-231 cells. At all tested doses of nanoparticles, the colonies were found to decrease significantly with the increase in the concentration of nanoparticles. The surviving fraction of colonies were 64.97 %, 49.25 % and 35.33% for doses of 25, 50 and 100 $\mu\text{g/mL}$ respectively compared to the control. It showed that silibinin loaded xanthan gum nanoparticles reduced the surviving fraction of cells in a concentration dependent manner and had the efficacy to inhibit regenerative ability of the treated cells.

4. Conclusion

Silibinin was successfully incorporated in xanthan gum nanoparticles by ionic gelation method using genipin as a chemical crosslinking agent. FTIR, XRD and DSC studies showed the incorporation of silibinin in the xanthan gum nanoparticles. Increasing the surfactant and crosslinker concentration resulted in higher drug loading efficiency. SEM study implied that the surface of the nanoparticles with minimum amount of surfactant and crosslinker appeared to be smooth compared to those of maximum amount of surfactant and crosslinker. DLS and TEM study revealed the size of the nanoparticle in nanometer range. Swelling and release properties of the drug from the nanoparticles were found to be both time and pH dependent and higher in alkaline pH. With the increase in surfactant content in the polymeric nanoparticles, swelling and release properties increased, but the behaviour of the nanoparticles reversed as the percentage of genipin crosslinker increased. Silibinin loaded xanthan gum nanoparticle inhibited the growth of MDA-MB-231 breast cancer cells. Further trypan blue dye assay and clonogenic cell survival assay showed the inhibition potentiality of MDA-MB-231 cell growth by silibinin loaded xanthan gum nanoparticles as compared to free silibinin. The results from the above studies showed that the anticancer efficacy of silibinin-loaded polymeric nanoparticles could be monitored by varying the parameters of surfactant and crosslinker in the nanoformulations.

Acknowledgement

Author would like to thanks Dept. of Chemical Sciences, Tezpur University for providing the necessary research infrastructure to carry out this work.

References

- [1] Banik, N., Ramteke, A., and Maji, T.K. Carboxymethyl chitosan montmorillonite nanoparticles for controlled delivery of isoniazid: evaluation of the effect of the glutaraldehyde and montmorillonite. *Polymer for Advanced Technology*, 25(12):1580-89, 2014.
- [2] Bhattacharya, S. Phytotherapeutic properties of milk thistle seeds: An overview. *Journal of Advanced Pharmacy Education and Research*, 1:69-79, 2011.
- [3] Broadhead, J., Edmond Rouan, S. K., and Rhodes, C. T. The spray drying of pharmaceuticals. *Drug Development and Industrial Pharmacy*, 18(11):1169-1206, 1992.
- [4] Bueno, V. B., Bentini, R., Catalani, L. H., and Petri, D. F. S. Synthesis and swelling behavior of xanthan-based hydrogels. *Carbohydrate Polymers*, 92(2), 1091-1099, 2013.
- [5] Cassano, R., Trombino, S., Ferrarelli, T., Mauro, M.V., Giraldi, C., Manconi, M., Fadda, A. M., and Picci, N., J. Respirable rifampicin based microspheres containing isoniazid for tuberculosis treatment. *Journal of Biomedical Materials Research Part A*, 100:536-542, 2012.
- [6] da Trindade Neto, C.G., Fernandes, A.L.P., Santos, A.I.B., Morais, W. A., Navarro, M., Dantas, T.N.C., Pereira, M.R., and Fonseca, J.L.C. Preparation and characterization of chitosan-based dispersions. *Polymer International*, 54:659-666, 2005.
- [7] Das, M., Mishra, D., Maiti, T. K., Basak, A., and Pramanik, P. Bio-functionalization of magnetite nanoparticles using an aminophosphonic acid coupling agent: new, ultradispersed, iron-oxide folate nanoconjugates for cancer-specific targeting. *Nanotechnology* 19:415101, 2008.
- [8] Deep, G., and Agarwal, R. Antimetastatic efficacy of silibinin: molecular mechanisms and therapeutic potential against cancer. *Cancer and Metastasis Reviews*, 29(3):447-463, 2010.
- [9] Devi, N. and Maji, T. K. Microencapsulation of isoniazid in genipin-crosslinked gelatin-A- β -carrageenan polyelectrolyte complex. *Drug Dev. Ind. Pharm.* 36:56-63, 2010.
- [10] Flickinger, F.C., and Draw, S.W. (1999). *Encyclopedia of Bioprocess Technology: Fermentation*, 5, 1999.
- [11] Gao, Z. H., Crowley, W. R., Shukla, A. J., Johnson, J. R., and Reger, J. F. Controlled release of contraceptive steroids from biodegradable and injectable gel formulations: in vivo evaluation. *Pharmaceutical Research*, 12(6):864-868, 1995.
- [12] García-Ochoa, F., Santos, V. E., Casas, J. A., and Gomez, E. Xanthan gum: production, recovery, and properties. *Biotechnology Advances*, 18(7):549-579, 2000.
- [13] Gogoi, P., Das, M.K., Ramteke, A., and Maji, T.K. Soy flour-ZnO nanoparticles for controlled release of silibinin: Effect of ZnO nanoparticle, surfactant, and cross-linker. *International Journal of Polymeric Materials and Polymeric Biomaterials*, 67(9):543-552, 2018.
- [14] Guo, G., Fu, S., Zhou, L., Liang, H., Fan, M., Luo, F., Qian, Z., and Wei, Y. Preparation of curcumin loaded poly (α -caprolactone)-poly (ethylene glycol)-poly (α -caprolactone) nanofibers and their in vitro antitumor activity against Glioma 9L cells. *Nanoscale*, 3(9):3825-3832, 2011.
- [15] Gupta, K. C., and Kumar, M. N. V. R. Drug release behavior of beads and microgranules of chitosan. *Biomaterials*, 21:1115-9, 2000.
- [16] Katzbauer, B. Properties and applications of xanthan gum. *Polymer Degradation and Stability*, 59:(1-3), 81-84, 1998.
- [17] Kizilay, E., Kayitmazer, A. B., & Dubin, P. L. Complexation and coacervation of polyelectrolytes with oppositely charged colloids. *Advances in Colloid and Interface Science*, 167(1-2), 24-37, 2011.
- [18] Kim, K. W., Choi, C. H., Kim, T. H., Kwon, C. H., Woo, J. S., and Kim, Y. K. Silibinin inhibits glioma cell proliferation via Ca²⁺/ROS/MAPK-dependent mechanism in vitro and glioma tumor growth in vivo. *Neurochemical research*, 34(8):1479-1490, 2009.
- [19] Mi, F. L. Sung, H. W. Shyu, S. S. Su, C. C. and Peng, C. K. Synthesis and characterization of biodegradable TPP/genipin co-crosslinked chitosan gel beads. *Polymer*, 44:6521-6530, 2003.
- [20] Mohd Zobir, B.H., Yarmo, M.A., Rahman, M.Z.H.A., Zainal, Z., and Liang, A.A.S. Texture and microstructure of chitosan-treated bentonite and its calcined products. *Malaysian Journal of Analytical Sciences*, 7(1), 35-40, 2001.
- [21] Muhamad, I. I., Fen, L.S., Hui, N.H., and Mustapha, N.A. Genipin cross-linked kappa-carrageena/ carboxymethyl cellulose beads and effect on beta-carotene release, *Carbohydrate Polymers*, 83(3):1207-1212, 2011.
- [22] Mudshinge, S. R., Deore, A. B., Patil, S., and Bhalgat, C. M. Nanoparticles: emerging carriers for drug delivery. *Saudi Pharmaceutical Journal*, 19(3), 129-141, 2011.
- [23] Pooja, D., Panyaram, S., Kulhari, H., Rachamalla, S. S., and Sistla, R. Xanthan gum stabilized gold nanoparticles: characterization, biocompatibility, stability and cytotoxicity. *Carbohydrate Polymers*, 110:1-9, 2014.
- [24] Pooja, D., Babu Bikkina, D. J., Kulhari, H., Nikhila, N., Chinde, S., Raghavendra, Y. M., Sreedhar, B., and Tiwari, A. K. Fabrication, characterization and bioevaluation of silibinin loaded chitosan nanoparticles. *International Journal of Biological Macromolecules*, 69:267-273, 2014.
- [25] Rosalam, S., and England, R. Review of xanthan gum production from unmodified starches by *Xanthomonas comprestis* sp. *Enzyme and Microbial Technology*, 39(2):197-207, 2006.
- [26] Saikia, C., Das, M. K., Ramteke, A. and Maji, T. K. Effect of cross linker on drug delivery properties of curcumin loaded starch coated iron oxide nanoparticles. *International Journal of Biological Macromolecules*, 93:1121-1132, 2016.
- [27] Selvaraj, S., Saravanakumar, N., Karthikeyan, J., Evangeline, D. D. Rajendran, N.N. and Lathamary, D. Acyclovir loaded chitosan nanoparticles for ocular delivery. *Asian Journal of Pharmaceutics*, 4(4), 2010.

- [28] Shankar, G. and Agrawal Y. K. Formulation and evaluation of polymeric micelles for a poorly absorbed drug. *Research Journal of Pharmaceutical, Biological and Chemical Sciences*, 6(3):1314-1321, 2015.
- [29] Sharma, B. R., Naresh, L., Dhuldhoya, N. C., Merchant, S. U., and Merchant, U. C. Xanthan gum-A boon to food industry. *Food Promotion Chronicle*, 1(5):27-30, 2006.
- [30] Singh, R. P., Deep, G., Chittezhath, M., Kaur, M., Dwyer-Nield, L. D., Malkinson, A. M., and Agarwal, R. Effect of silibinin on the growth and progression of primary lung tumors in mice. *Journal of the National Cancer Institute*, 98(12):846-855, 2006.
- [31] Singh, R. P., Gu, M., and Agarwal, R. Silibinin inhibits colorectal cancer growth by inhibiting tumor cell proliferation and angiogenesis. *Cancer research*, 68(6):2043-2050, 2008.
- [32] Singh, R. P., and Agarwal, R. Prostate cancer prevention by silibinin. *Current cancer drug targets*, 4(1):1-11, 2004.
- [33] Su, L., Ji, W.K., Lan, W.Z., Dong, X.Q. Chemical modification of xanthan gum to increase dissolution rate. *Carbohydrate Polymers* 53:497-499, 2003.
- [34] Tavakol, M., Farahani, E. V., and Najafabadi, S.H. The effect of polymer and CaCl_2 concentrations on the sulfasalazine release from alginate-N,O-carboxymethyl chitosan beads. *Progress in Biomaterials* 2:1-8, 2013.
- [35] Teng, Z., Luo, Y., and Wang, Q. Carboxymethyl chitosan-soy protein complex nanoparticles for the encapsulation and controlled release of vitamin D3. *Food chemistry*, 141(1):524-532, 2013.
- [36] da Trindade Neto, C.G., Fernandes, A.L.P., Santos, A.I.B., Morais, W. A., Navarro, M., Dantas, T.N.C., Pereira, M.R., and Fonseca, J.L.C. Preparation and characterization of chitosan-based dispersions. *Polymer International*, 54:659-666, 2005.
- [37] Xu, P., Yin, Q., Shen, J., Chen, L., Yu, H., Zhang, Z., and Li, Y. Synergistic inhibition of breast cancer metastasis by silibinin-loaded lipid nanoparticles containing TPGS. *International journal of pharmaceutics*, 454(1):21-30, 2013.
- [38] Yoo, H. J., Ahn, S. H., Kim, S. B., Kim, W. K., and Han, O. S. Efficacy of progressive muscle relaxation training and guided imagery in reducing chemotherapy side effects in patients with breast cancer and in improving their quality of life. *Supportive Care in Cancer*, 13(10):826-833, 2005.
- [39] Yuan, Y., Chesnutt, B.M., Utturkar, G., Haggard, W.O., Yang, Y., Ong, J.L., and Bumgardner, J.D. The effect of cross-linking of chitosan microspheres with genipin on protein release. *Carbohydrate Polymers*, 68(3):561-567, 2007.

Phase transition driven structural transformation in colloidal membranes

Lachit Saikia

*Department of Physics,
Dibru College, Dibrugarh
Email: lachit.iisc@gmail.com*

Abstract

The phenomenon of structural transformation in 2-D thin elastic sheets through differential strains has been extensively studied and applied in various fields like flexible electronics, origami, and tissue growth. In contrast, at the molecular scales of lipid bilayers and cell membranes, external agents such as proteins typically influence their curvature and shape. This article aims to discuss the novel approaches, at intermediate colloidal length scales, that have been utilized to carry out structural transformation of the fluid membranes using internal phase transitions based on intrinsic interactions and physical properties of constituent rod-like colloidal particles.

By employing an isotropic mixture of highly monodispersed rod-like viruses and adding non-adsorbing polymer through depletion attraction, membranes of aligned rods are spontaneously assembled. These membranes possess a thickness of one rod length and display long-wavelength thermal elastic fluctuations. The intrinsic chirality of the virus rods leads to the trapping of a uniform chiral rod twist at the membrane edge, which plays a crucial role in structural transformation of the membranes into globally buckled and locally wrinkled three-dimensional structures through crystallization. The presence of finite line tension in the membrane edge results in a circular geometry. However, introducing another rod with similar chirality but different aspect ratio shapes the membranes into an unusual cyclic polygon geometry. This anisotropic shape arises from the spreading phenomenon of a colloidal membrane composed of short-thick rods over another membrane of purely long rods in the presence of disorder. Eventually this process leads to formation of pinning junctions which form the cornerstones of the cyclic polygon structure of the membranes. These junctions consist of weakly cross-linked compressed defective rods, rather than topological defects.

L. Saikia, "Phase transition driven structural transformation in colloidal membrane", Horizon (Ed. Book), 51-73 (2023), copyright: Tinsukia College Publishing. ISBN: 978-81-19322-07-7

1. Introduction

a. Colloids

Colloids are composed of small solid particles that are suspended in a dispersion medium, which can be a liquid or gas. These particles typically have sizes ranging from tens of nanometers to micrometers, reaching a point where the influence of Brownian motion, caused by thermal energy fluctuations, becomes significant. Colloids are abundant in both natural and synthetic settings. Noteworthy examples include milk, fog, paints, cosmetics, and liquid crystalline displays, all of which contain suspended particle sordroplet satthemicrons scale within a medium. Furthermore, colloids play a crucial role in bio medical research, find in gextensive application in genetherapy, drug delivery, tumordetection, biosensors, and various other fields [1,2,3,4,5]. In essence, colloids are an indispensable component of our daily lives.

The study of colloidal suspensions has a rich and esteemed history, involving renowned scientists who made significant contributions to the field. It all began in 1827 when Robert Brown discovered the constant random motion observed in pollen grains, which is now known as Brownian motion. Nearly eighty years later, in 1905, Albert Einstein provided a theoretical explanation for this phenomenon, focusing on the random motion of pollen grains suspended in a liquid medium. In 1908, Jean Perrin conducted experimental work on gum colloids, which, when combined with Einstein's theory of Brownian motion, provided conclusive evidence for the discrete nature of matter [6, 7]. Viewing the liquid dispersion medium of a colloidal suspension as an inert continuum with macroscopic properties like density and dielectric constant allows one to consider colloidal particles as experiencing an effective pair potential $U(r)$ or a potential mean force. As a result, the thermodynamic equilibrium properties of a colloidal suspension can be understood using the laws of statistical physics developed for atoms and molecules. This means that theories formulated for simple liquids and solids using statistical mechanics for atoms and molecules can be applied to calculate and predict various phase behaviors of colloidal suspensions. Interestingly, under specific conditions, suspensions of uniform spherical colloids display phases analogous to atomic phases, such as colloidal gas, liquid, and crystals, exhibiting characteristics of long-range and short-range order for crystals and liquids, respectively. Consequently, colloids can be thought of as "giant atoms" from a thermodynamic perspective. However, colloidal suspensions also offer unique experimental opportunities to explore fundamental problems in condensed matter physics, including the kinetics of crystallization and the nature of the glassy state [8, 9].

Now, let's consider a suspension of uniformly shaped colloidal particles under going Brownian motion. The average displacement of these particles over time can be described by the equation

$$\langle \Delta r^2 \rangle = 2dDt,$$

Where 'd' represents the dimension, 'D' is the diffusion coefficient, and 't' denotes time. The diffusion coefficient 'D' can be determined using the Stokes-Einstein relation,

$$D = \frac{k_B T}{6\pi\eta R}$$

which relates it to the particle radius 'R' and the viscosity of the solvent ' η '.

By applying this equation, we can estimate the characteristic structural relaxation time (τ) for the system, which is the time it takes for a colloidal particle to diffuse a distance equal to its own size 'R' in three dimensions. The equation provides different values for 'D' depending on the system, such as ~1 pico second for atomic systems and ~1 second for colloidal systems. This distinction is significant because colloidal systems operate on a micron-sized length scale and a time scale on the order of seconds, allowing for direct experimental investigation using real-time optical microscopy, which is not possible in atomic systems. Furthermore, colloidal systems exhibit interaction energies on the scale of thermal energy ($k_B T$), while atomic crystals can have varying interaction energies ranging from values comparable to $k_B T$ to electron volts, depending on the nature of inter atomic bonding. Notably, the elastic constants of these systems, representing the lattice energy per unit volume, can be easily estimated. For a system with N particles in volume V , the elastic modulus E can be calculated. Comparing atomic size (R_{atom}) and colloidal size (R_{colloid}), where $R_{\text{colloid}} = R_{\text{atom}} \times 10^3$, the elastic constant of colloidal systems is approximately 10^9 times smaller than that of atomic crystals. This suggests that colloidal crystals are much more deformable compared to the softest atomic crystals held together by dispersion forces. As a result, colloidal systems are often referred to as "soft matter systems" that can easily undergo deformation under external influences such as temperature, optical, electrical, magnetic fields, shear, etc [10, 11, 12, 13]. These distinctive characteristics of colloidal suspensions make them valuable tools in soft matter experiments.

b. Phase behavior of colloidal suspension

The phase behavior of hard spherical colloids and rod-shaped colloids (such as cylinders or ellipsoids) is a fascinating area of research in soft matter physics and material science. These particles interact based on their shape, size, and mutual orientation, leading to a rich variety of phases and phase transitions [14,15,16]. Hard spherical colloids at low enough density exhibit a gas-like phase where the particles are dispersed randomly in space, with minimal interactions between them. The system resembles a dilute gas, with high fluidity and low particle density. As the particle density increases, the gaseous phase makes a transition into a liquid-like phase. In this phase, the particles are closer to each other, resulting in local structural order. However, there is no long-range order, and the system lacks a regular, repeating pattern over long distance. The liquid phase displays properties of both fluidity and rigidity. With further increase in particle density, the liquid phase undergoes a transition into a crystalline phase. The crystalline phase exhibits strong positional correlations and has a low degree of randomness. The particles arrange themselves in a well-defined, periodic pattern with long-range order. The crystalline phase exhibits strong positional correlations and has a low degree of randomness.

As stated earlier, the shape of the colloids plays an important role in their phase behavior. The rod-shaped colloids at low densities form an isotropic phase, where rods are randomly

oriented, and their positions are disordered, like the gaseous phase of hard spherical colloids. On increasing the number density of these colloids, the isotropic phase transforms into a nematic phase. In the nematic phase, the rod-shaped particles align along their long axes in a preferred direction while maintaining a disordered positional arrangement [15]. This alignment leads to a unidirectional orientational order, making the nematic phase anisotropic [17]. Further increasing the density can lead to more complex phases. For instance, rod-shaped colloids can form the cholesteric phase, where the particles form layers with a helical twist. In the smectic phase, the rods self-organize into distinct layers with positional ordering within each layer.

Phase transitions in both hard spherical colloids and rod-shaped colloids occur due to changes in thermodynamic parameters such as temperature, pressure, or particle concentration. The behavior during phase transitions provides insights into the collective behavior of the particles. In addition to fundamental research, understanding and controlling the phase behavior of hard spherical colloids and rod-shaped colloids are crucial for various scientific and technological applications. These colloidal systems serve as model systems for investigating fundamental principles of phase transitions and provide insights into complex phenomena observed in a wide range of materials, including liquid crystals, polymers, and biological systems. Moreover, they find applications in diverse fields such as materials science, liquid crystal displays, and advanced functional materials.

c. Depletion interaction

In presence of non-adsorbing polymers or particles in a suspension of colloidal particles which are larger in size compared to the polymer molecules, an effective attraction is induced between the colloidal particles even in the absence of any direct attractive forces between them. This induced attractive force between the larger particles due to the smaller ones is popularly known as depletion interaction [18]. When non-adsorbing polymers or particles are added to a colloidal system, they tend to minimize their contact with the colloidal particles. As a result, the polymers or particles create an excluded volume around the colloids, effectively depleting the available space. This depletion zone is devoid of the non-adsorbing species and creates a difference in osmotic pressure between the inside and outside of the depletion zone. On approach of one colloidal particle to the other, their depletion zones overlap leading to reduction in excluded volume. Overall, this creates a net osmotic pressure difference in the region between the colloids and outside of them. This generates an effective attraction leading to clustering of the particles (Figure 1). The depletion interaction arises because the particles are driven together to reduce the unfavorable excluded volume. The attraction is typically short ranged, decaying exponentially with distance, and can be modulated by the concentration and size of the non-adsorbing species. The depletion interaction is a reversible and tunable interaction that depends on the concentration and size of the non-adsorbing species. By adjusting these parameters, it is possible to control the strength and range of the interaction between the colloidal particles. This has important

implications in various fields, including colloid science, materials science, and biophysics. In colloidal suspensions, depletion interactions can lead to the aggregation or clustering of colloidal particles, altering the properties and behavior of the system. It can affect the phase behavior, rheology, and stability of colloidal dispersions. Additionally, depletion interactions play a significant role in biological systems, such as the behavior of proteins and the assembly of biological structures.

d. Self-assembly of colloids and colloidal membrane

As discussed earlier, suspension of rod-shaped colloids exhibits isotropic phase in lower concentrations which on increasing show orientationally prominent nematic phase and other complex phases like smectic phase, cholesteric phase, etc. Now, let us consider a dilute suspension of rod-shaped colloids. For the study discussed in this article, the considered rod-shaped colloids are filamentous viruses, e.g., fd, fdY21M, fdY21M-PEG, M13K07, etc. These are semiflexible, highly charged and monodispersed rod-shaped colloids. Addition of non-adsorbing polymers to such dilute virus suspensions results in phase separation into polymer-rich and virus-rich regions due to depletion forces. Depletion forces arise when excluded volumes overlap, leading to an increase in polymer coil entropy and a decrease in rod entropy. The strength and range of depletion interactions can be tuned experimentally. In suspensions of rod-shaped viruses, depletion interactions exhibit a directional nature, causing the formation of nematic droplets and colloidal membranes (discussed below) at low and high polymer concentrations, respectively [19]. The colloidal membrane phase, akin to a micron-sized lipid bilayer, presents distinct microscopic interactions and allows for direct visualization and manipulation. It offers a versatile system for self-assembly studies using simple experimental tools like light microscopy and optical laser tweezers. In the preceding section, it was discussed how the addition of non-adsorbing polymers like dextran or PEG to a dilute suspension of rod-shaped viruses induces the formation of rod clusters due to depletion interactions. The electrostatic interactions of charged biopolymers are typically screened by the addition of electrolyte, allowing the rods to behave as model hard rods with a slightly larger effective diameter [20]. These colloidal rods maximize their excluded volume overlap by aligning parallel to each other and maintaining the shortest distance between their centers. As a result, they arrange into small disc-like structures, with rods aligning along the surface normal (Figure 2). At intermediate polymer concentrations, these small discs merge laterally to form larger discs. These disc-like structures have been observed to be thermodynamically stable, even reaching diameters of hundreds of microns after equilibration for days. These self-assembled disc-like structures consisting of rod-shaped colloids are known as colloidal membranes. The stability of colloidal membranes and their lateral growth have been investigated, questioning why depletion does not act between the surfaces of two membranes. It was found that protrusion fluctuations of the individual rods, which displace surrounding polymer molecules, play a crucial role.

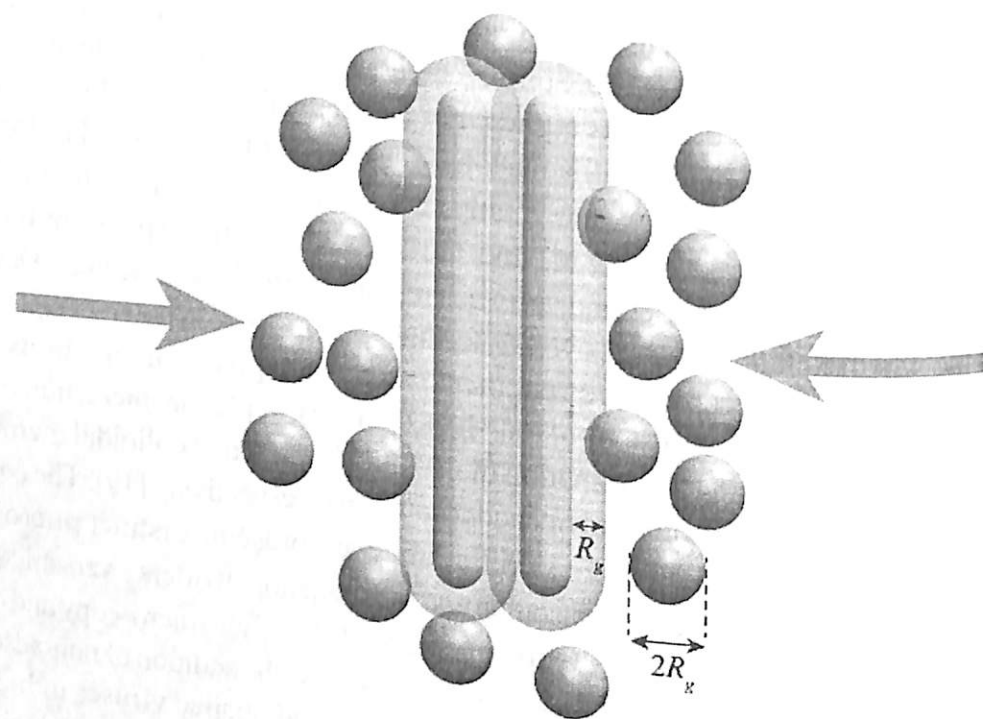


Figure 1: Schematic illustration of depletion forces acting between two colloidal rods (green). The transparent shell around each rod is the excluded volume zone which on overlapping create excess free volume for polymer coils (blue spheres). As a result, free volume entropy of the depleting polymers increases and thus leads to effective attraction between the rods.

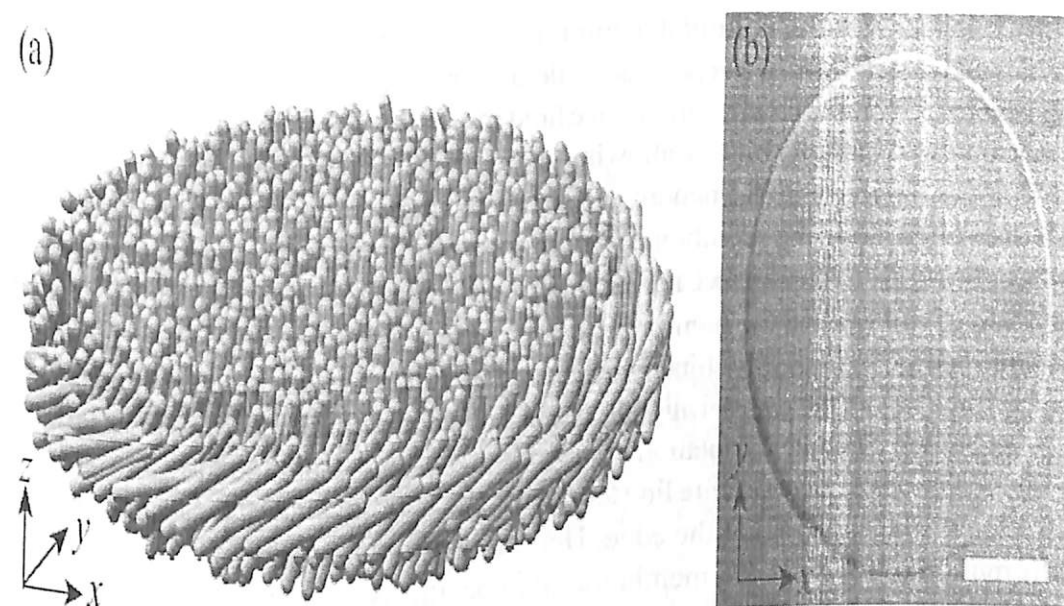


Figure 2: (a) Schematic of a colloidal membrane composed of rod-like particles. (b) DIC image of a colloidal membrane assembled from a dilute isotropic suspension of M13K07virus at 21 mg/ml of PEG 35k. Scale bar, 10 μm .

C. Microscopy tools

At low polymer concentrations, the rods require displacing fewer molecules during protrusion fluctuations, resulting in stable isolated 2D smectic monolayers [21]. However, at higher polymer concentrations, the displacement of surrounding polymer molecules becomes energetically costly, leading to smoother monolayer surfaces. Consequently, repulsive interactions between membranes approaching face-on are diminished, favoring stacking of membrane layers on top of each other. These membranes, formed through entropic forces in a mixture of filamentous viruses and non-adsorbing polymers, resemble conventional lipid bilayer membranes in certain properties. However, there are fundamental differences between colloidal membranes and lipid bilayers, making colloidal membranes a unique system for studying membrane biophysics and smectic monolayer physics. Colloidal virus rods are non-amphiphilic and interact through excluded volume interactions, unlike lipid molecules. Moreover, colloidal membranes have micron-sized building blocks, allowing direct visualization of structures and dynamics, which is not as feasible with the nanometer-sized lipid molecules. The rods in the bulk of the membranes align along the membrane normal, however, rods at the edge show continuous twist from being along the surface normal to lie on the membrane plane [22, 23]. This finite penetration of rod twist at the membrane's edge occurs due to two possible reasons. One, these virus rods are intrinsically chiral which favors a twist between neighboring rods, whereas depletion force favors the parallel alignment. As a result, the chiral rod twist is expelled to the membrane's edge. Second, orientation of the rods at the edge similar to the membrane's bulk is energetically costly due to finite line tension of the colloidal membranes and thus favors penetration of the rod twist at the edge. Here, in this article we discussed the structural transformation of the colloidal membranes that occurs as a result of crystallization in the membrane. The planar geometry of the membranes transformed into non-Euclidean three-dimensional structure. In another case, this article highlights the colloidal membranes assembled out of binary mixture of homochiral rods having similar handedness of chirality and different aspect ratio resulting in an implausible cyclic polygon shaped colloidal membranes. The inner polygons of these fluid membranes are composed of longer rods whereas the outer lobes extending from the sides of the polygon are composed of shorter thick rods.

1. Methodology

a. Preparation of virus: rod-shaped colloids

To conduct the studies outlined in this article, a range of filamentous viruses were utilized, namely fdY21M, fdY21M-PEG, M13K07, and M13K07Y21M. These viruses have been intricately engineered by nature to exhibit remarkable consistency in physical characteristics such as length, diameter, and charge density within a group of virus particles prepared at a time. This exceptional uniformity makes them exceptionally well-suited as a model system for exploring liquid crystalline phenomena in experimental research. These virus suspensions were prepared using standard molecular biology protocols where the host bacteria *Escherichia coli* is cultured in sufficient amounts and subsequently infected with filamentous virus [24]. Adequately

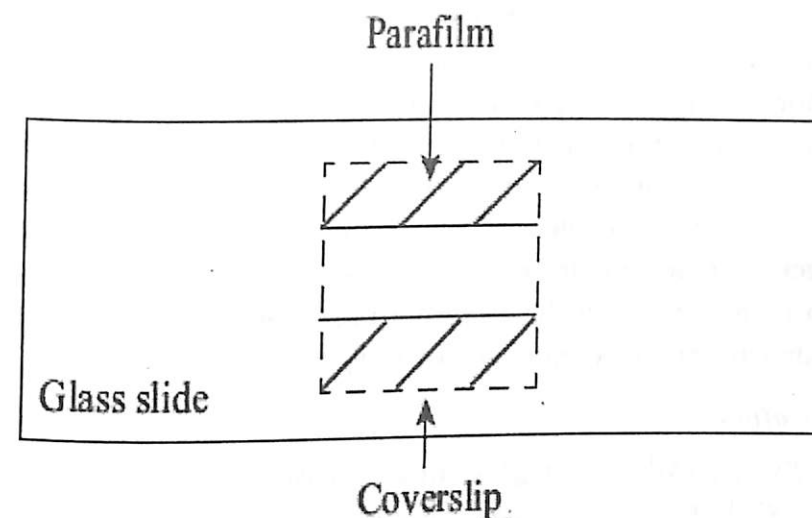


Figure 3: Schematic of sample chambers prepared using microscope glass slide and cover slip used for self-assembly of colloidal membranes.

grown infected culture is further purified and virus particles are separated through multiple filtration steps. The entire process takes 40 to 60 hours and about 5-7 gm of virus is obtained from 5 liters of culture. Fluorescent dyes containing N-hydroxysuccinimide (NHS) esters (DyLight 550 from Thermo Scientific) are used to label the coat protein of the virus rods. This allows visualization of the labeled proteins or the virus rods under fluorescence microscopy. The NHS esters in the dyes bind covalently to the primary amines of the virus coat protein, releasing the NHS group. Similarly, N-hydroxysuccinimide (NHS) ester terminated poly (N-isopropylacrylamide) (Mn 2000) and succinimidyl valeric acid functionalized methoxy polyethylene glycol (mPEG-SVA-20k, Nanocs) are attached to the virus surface through NHS ester bonding. In the following paragraphs, a detailed step-by-step description of the procedure for labeling the viruses with the fluorescent dye molecules are discussed. The same procedure is also used for labeling the polymers onto the virus surface. Typically, the virus particles are stored in a buffer containing 20 mM Tris HCl and 100 mM NaCl at pH 8.0. However, the Tris in this buffer is likely to interfere with the dye labeling reaction as it reacts with the NHS-ester moiety. To avoid this interference, the virus suspension buffer is changed to 50 mM sodium borate at pH 8.5, which does not contain any primary amines. For efficient bonding, the virus suspensions are prepared at a concentration of 5-10 mg/ml for fluorophore labeling reaction and 1 mg/ml for polymer labeling reaction. The DyLight NHS-ester reagents are highly sensitive to moisture and are stored at -20 degree Celsius with a desiccant. Therefore, the vial must be allowed to equilibrate to room temperature for 10-15 minutes before opening. The amine-reactive fluorescent dye and polymers come in powdered form and are dissolved in dimethyl sulfoxide (DMSO) at concentrations of 10 mg/ml for the fluorescent dye, 100 mg/

ml for mPEG-SVA, and 20 mg/ml for poly(N-isopropylacrylamide). Depending on the desired degree of labeling, the appropriate amount of dye labeling reagent is added to the virus suspension in the borate buffer and incubated at room temperature for 1-2 hours. We usually incubate the mixture for the first 1-2 hours at room temperature and then overnight at 4 degree Celsius. In the case of co-labelling the virus rods with both the polymer and the fluorophore, first the amine-reactive polymer solution is added to the virus suspension in the borate buffer and the same is incubated for about 1-2 hours at room temperature. Then, the dye labeling reagent is added and incubated for another 2-3 hours at room temperature. Excess dye/polymer is removed from the suspension through ultracentrifugation, and the virus pellet is resuspended in a buffer containing 20 mM Tris HCl and 100 mM NaCl at pH 8.0.

b. Sample Preparation

The samples are prepared by mixing the virus suspension with a non-adsorbing polymer (PEG 35k or Dextran 500k, Sigma Aldrich). The mixture is injected into a sample chamber shown in Figure 3. The chambers are made using cleaned microscope slides and coverslips with a parafilm spacer. The coverslips are coated with a polyacrylamide brush to minimize virus-glass surface interaction. Finally, the chambers are sealed with Araldite epoxy glue or Norland optical glue to prevent evaporation.

The colloidal membranes were imaged using an inverted microscope (Olympus IX73) with various imaging techniques including differential interference contrast (DIC), fluorescence, bright field, and phase contrast microscopy. Images were captured using a charged-coupled-device camera (Cool Snap HQ2, Photometrics) and a TRITC filter cube with specific excitation and emission wavelengths. Three-dimensional images of solid membranes were obtained using a laser scanning microscopy setup (Zeiss LSM 880 Airyscan). The membranes were assembled from fluorescently labeled M13K07 rods, and fluorescence images were acquired using a 543 nm He-Ne laser. Cross-section images of the membranes were stacked and rendered using Matlab. The fast kinetics of virus rods within colloidal membranes were recorded using a high-speed CMOS camera (PCO1200HS) at a frame rate of 500 frames per second. The orientation of virus rods within the colloidal membrane was determined using an advanced quantitative polarization microscopy technique called LC-PolScope (Marine Biological Labs). LC-PolScope replaces the traditional compensator with an electronically controlled liquid crystalline plate, enabling rapid switching of the polarization state of trans-illuminating light. The raw images were processed using micromanager to obtain high-resolution images of retardance and slow axis orientation. The pixel intensity of the retardance image directly corresponds to the spatial dependence of sample birefringence resulting from the orientation of the rods.

d. Analysis

The images of colloidal membranes captured using various microscopy techniques were analyzed using *ImageJ* and *MATLAB*. Specific codes were written for every analysis of the data.

1. Results

a. Crystallization of chiral colloidal membranes

The M13K07 rod-shaped viruses, 1.2 μm in length and 6.6 nm in diameter, naturally come together to form layers of aligned rods when a non-adsorbing polymer called Polyethylene Glycol (PEG 35k, Sigma Aldrich) is added to its suspension which induces depletion attraction [25]. As explained in the earlier sections, the depletion interaction in a suspension of these viruses promotes parallel alignment of the rods. Consequently, the virus rods align along their long axes within monolayer membranes that are as thick as one rod length. The virus rods within these membranes exhibit fluid-like dynamics and order at low PEG concentration (20 mg/ml), allowing for free diffusion of the rods within the membrane. Furthermore, due to the finite line tension at the fluid membrane's edge, the membranes adopt a circular shape (Figure 1) and exhibit thermal fluctuations at macroscopic scales [21].

Theoretical studies on colloidal membranes' phase behavior suggest that rod-like viruses having a sufficiently high aspect ratio in the presence of strong depletion interaction will undergo crystallization [26]. The research work presented in this article not only confirms the predicted solid phase of the colloidal membrane at a high PEG concentration (24 mg/ml) but also discovers that the solid membranes are highly roughened and buckled (Figure 4). Interestingly, along with crystallization, the colloidal membranes exhibit an unprecedented curvature instability.

The microscopic structure of fluid and solid colloidal membranes is examined using small angle x-ray scattering. For the fluid membranes, a broad scattering peak is observed, indicating an inter-rod spacing of 11.74 nm (Figure 5). In contrast, the peak shifts to 10.03 nm for solid membranes, which aligns with the 2D crystallization process involving increased density [27]. By modelling the virus rods as cylinders with a uniform electron density of 6.6 nm, it is found that the first zero of the virus rods' form factor occurs at $q = 1.1 \text{ nm}^{-1}$, approximately halfway between the positions of the (11) and (20) peaks expected from a two-dimensional hexagonal lattice. Consequently, the higher-order scattering peaks of the crystalline phase beyond (10) become too small to measure, limiting the experimental observation to only one peak in the scattering profile.

To observe the curvature instabilities related to the transition from a fluid to solid state in colloidal membranes, the membranes are assembled using thermo-responsive polymer, PEG whose osmotic pressure changes with temperature [28]. Therefore, the osmotic pressure of PEG in a buffer solution is adjusted by temperature manipulation. The colloidal membranes are assembled at an intermediate polymer concentration (22 mg/ml) and transformed into a solid state by cooling the sample chamber to approximately 15 degree Celsius. The time-lapse images in Figure 6a-c depict the transformation process of a colloidal membrane, which involves nucleation and growth. In a typical equilibrated sample, the nucleation time is around 30 minutes, however it may vary depending on the level of supercooling and the overall polymer concentration used during membrane assembly. After nucleation, as shown in Figure 6b,c, growing domains with faceted boundaries become apparent, suggesting possible crystalline positional order of rods within the solid membranes. Within these growing solid

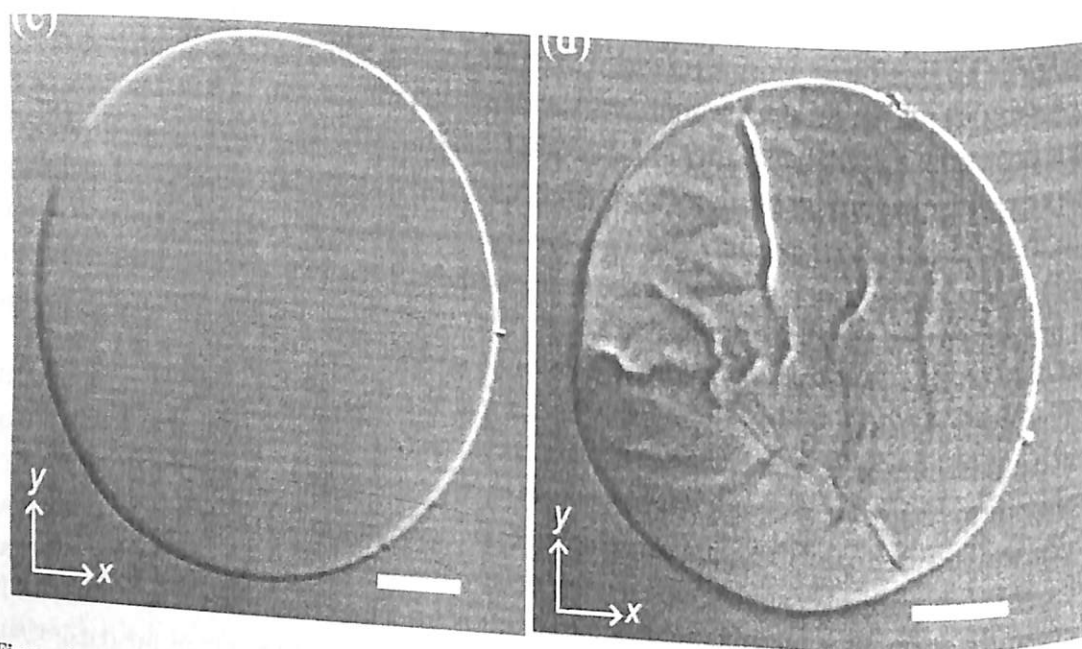


Figure 4: (a) Differential interference contrast (DIC) image of a fluid colloidal membrane at 20 mg/ml PEG concentration. Scale bar, 10 μm . (b) DIC image of a solid colloidal membrane at 24 mg/ml PEG concentration. Scale bar, 10 μm .

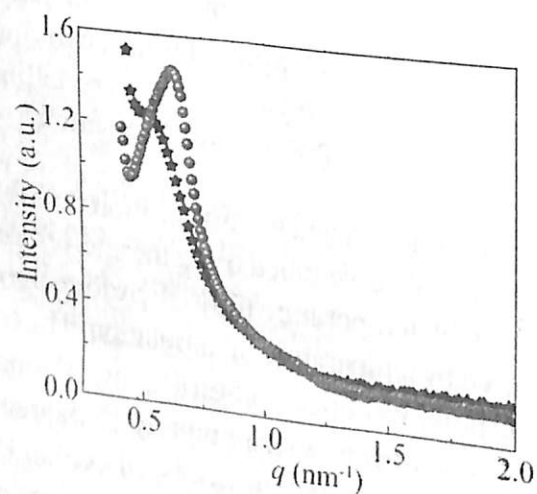


Figure 5: SAXS intensity as a function of scattering wave vector (q) for fluid (black stars) and crystalline (red circles) colloidal membranes at 20 and 25 mg/ml PEG concentration, respectively.

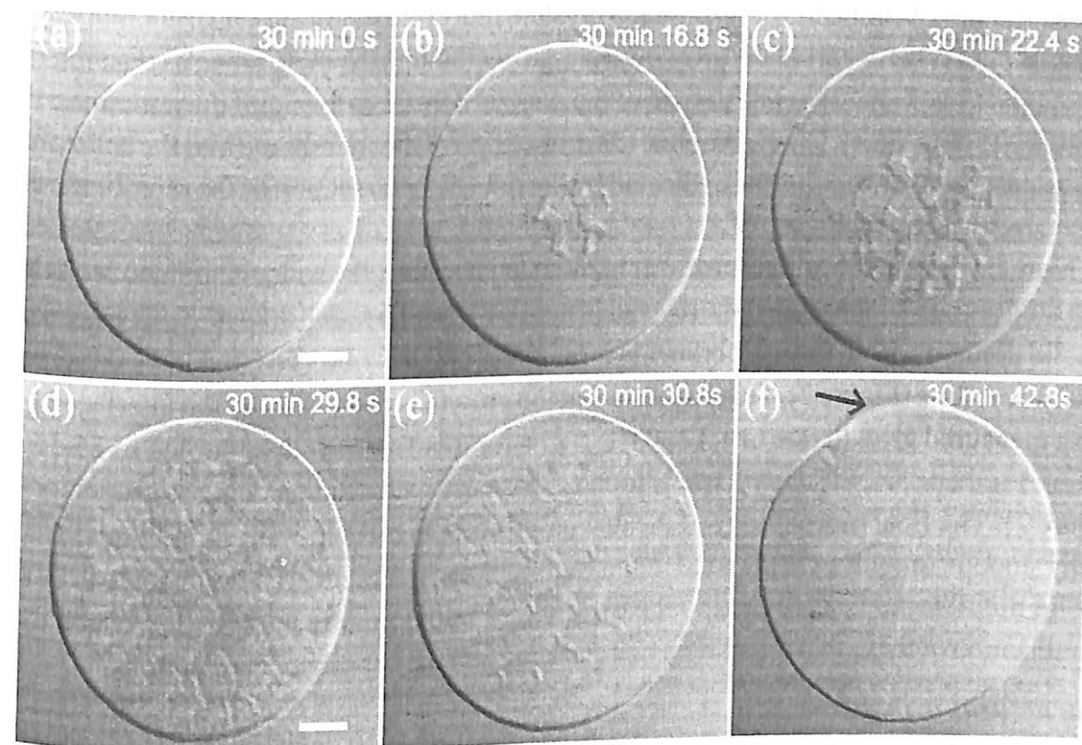


Figure 6: Curvature instability during fluid-crystal transition in colloidal membranes. (a)-(c) Time lapse sequence of nucleation and growth of a crystalline domain within a fluid membrane. The growing domain spontaneously develops a network of local ridges (protrusions). Scale bar, 10 μm . (d)-(f) Time lapse sequence showing annealing of the ridges, shown in (a)-(c), as the growing domain encounters the edge. The black arrow in the last panel indicates the out of focus edge of the transformed membrane. Scale bar, 10 μm .

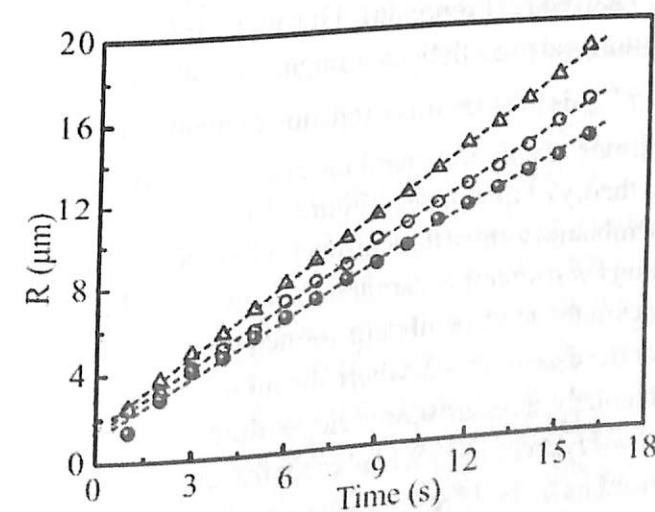


Figure 7: Growth kinetics of crystalline domains. Experimental data for the size of the crystalline domain as a function of time at 16 degree Celsius (triangles), 18 degree Celsius (open circles), and 19 degree Celsius (closed circles).

domains, a network of local ridges (protrusions) emerges to get rid of the excess strain energy that arises during membrane crystallization through out-of-plane buckling (Figure 6b-d). The primary sources of strain during membrane crystallization are dislocations, disclinations, and tilt distortions. When a part of the growing domain encounters the membrane's edge, the local ridges begin to self-anneal, while the rest of the domain continues to engulf the remaining fluid membrane (Figure 6d-f). Upon complete transformation of the membrane into the solid phase, it experiences global buckling, as indicated by the out-of-focus section of the membrane edge marked by an arrow in Figure 6f.

Now, on heating the sample chamber to room temperature, the solid membrane melts back into a fluid phase, demonstrating the reversible nature of the phase transition [25]. The growth of the solid domain within the membranes is monitored, and it is observed that its size increases linearly over time (Figure 7). This suggests that the growth is controlled by the interface [114], with a measured growth rate of $1.12 \pm 0.01 \mu\text{m/s}$ for membranes super cooled to 15 degree Celsius. Furthermore, the growth rate is found to increase with a higher degree of supercooling (Figure 7). The final morphology of the discussed solid membranes is globally buckled but locally smooth, in contrast to the solid membrane depicted in Figure 4b. This distinct difference in solid membranes is attributed to the differences in the kinetic pathway involved in the crystallization of these two membranes, as elaborated in the subsequent paragraphs.

b. Structural transformation

As mentioned in the previous section, the assembly of the colloidal membrane using M13K07 viruses' results in local wrinkling during the crystallization process and global buckling after crystallization. This leads to the formation of 3D structures in solid membranes. The degree of out-of-plane buckling is determined by capturing cross-sectional images of the solid membranes using confocal microscopy, and these 2D images are processed and stacked using *Matlab* and *ImageJ* software (Figure 8a). The analysis focuses on membranes that solidified through nucleation and growth from a single nucleation center. The deformation amplitude, $\Delta H = H_{\text{max}} - H_{\text{min}}$ is measured as a function of membrane size, and it is observed that the deformation amplitude of the solid membranes increases monotonically with size, as expected from the elastic theory of thin plates (Figure 8b).

Additionally, solid membranes formed through nucleation and growth at multiple nucleation sites exhibit local wrinkling even after the completion of the crystallization process. Figure 9 displays 3D confocal images of the solid membranes formed from varying numbers of nucleation sites (N). Specifically, for the case of $N=2$, where the membrane nucleates at two different sites and grows simultaneously, a macroscopic ridge (domain wall) gets trapped near the center of the membrane upon crystallization. Membranes that crystallize with $N>2$ trap multiple domain walls within the membrane, resulting in local height variations across the entire membrane area. The surface morphology of the membranes is further characterized by plotting the area fraction (ϕ) of the membrane covered by local peaks and troughs as a function of N (Figure

10a). It is observed that ϕ is lowest for membranes with $N=1$, indicating smooth surface topography in the absence of ridges and buckling at larger scales (Figure 8a). ϕ sharply increases for $N=2$ and remains constant for $N>2$, indicating that solid membranes formed through nucleation and growth with $N>2$ are locally rough. The number of local peaks and troughs in the membrane also increases with an increase in N (Figure 10b).

In order to examine the effect of substrate in the crystallization of the membranes, the membranes were assembled from a high virus concentration mixture (5 mg/ml) where the membranes have a tendency to grow perpendicular to the plane of the cover slip. It is observed that these membranes also undergo wrinkling during the fluid-solid transition in the face-on configuration. Thus, it shows that substrate doesn't have any role in the structural transformation of the solid membranes.

c. Colloidal membranes of binary rod mixture

In the preceding sections, we have discussed how on-set of crystallization plays a crucial role in shaping the chiral colloidal membranes into 3D structures. In this section, we delve into the impact of the geometrical differences among the constituent virus rods and their inherent chiral interaction on the self-assembly of colloidal membranes. By adding a non-adsorbing polymer dextran of sufficient quantity to an isotropic mixture of long and short-thick rods, the polymer rods spontaneously arrange themselves into fluid colloidal membranes with cyclic polygon shapes (Figure 11b,d). Interestingly, the short-thick rods segregate into lobes, while the long rods form the faceted core of the cyclic polygon [29]. We attribute this anisotropic shape to the formation of pinning centers during the spreading of one liquid over another, particularly in the presence of disorder.

At low polymer concentrations, the colloidal membranes assembled from the mixture of long and short-thick rods exhibit a circular shape with uniform mixing of the two types of rods (Figure 11a,c). However, at higher polymer concentrations, the dominance of excluded volume effects over the entropy of mixing leads to phase separation. This results in the formation of an intriguing cyclic polygon shaped colloidal membranes where the long and short-thick rods within the membranes segregate into distinct regions. The self-assembly of this unexpected geometry in a colloidal rod assemblage governed by line tension is attributed to liquid-liquid wetting in the presence of disorder. Unlike the spreading of liquids on solid substrates, disorder in liquid-liquid spreading acts as mobile pinning centers, creating pinning junctions that prevent rods of the membrane act as mobile pinning centers. Notably, in the assemblage of cyclic polygon shaped complete spreading of the membrane. Unlike conventional self-assembly approaches that aim to avoid disorder and defects, we present here a mechanism that utilizes defects to self-assemble higher-order complex structures that would otherwise be unattainable.

As stated earlier, upon addition of a non-adsorbing polymer, dextran (500 kDa, *Sigma Aldrich*), to a mixture of rod-shaped viruses with a high aspect ratio (~ 140), the virus rods

undergo spontaneous alignment, forming monolayer membranes that are one rod-length thick. Within the membrane, the rods move freely, exhibiting fluid-like order and dynamics. The behavior of colloidal membranes formed from a mixture of two types of viruses depends on the strength of the depletion interaction and the chirality of the virus rods. For this particular study, a dilute isotropic mixture of rod-shaped viruses, namely M13K07Y21M (1200 nm long, 6.6 nm in diameter) and fdY21M-PEG (880 nm long, 34 nm in diameter) were used. These filamentous viruses are excellent model systems for studying the physics of liquid crystals. Both types of viruses form right-handed cholesteric structures in their bulk samples [30]. The fdY21M-PEG viruses are prepared by grafting neutral polymer PEG 20k ($R_g \sim 7$ nm) uniformly onto the surface of fdY21M viruses, resulting in an approximately five-fold increase in the diameter of the virus rods (~ 34 nm) without affecting their chiral behavior. At low depletant concentration, the entropy of mixing dominates, leading to the formation of membranes with uniformly mixed virus rods that display fluid-like order and dynamics (Figure 11a,c). Conversely, at high depletant concentration, excluded volume interactions become dominant, resulting in a near bulk phase separation of the two types of virus rods within the membrane (Figure 11b,d). Membranes with uniformly mixed rods have circular shapes, while those assembled at high depletant concentrations exhibit cyclic polygon-like shapes. The membranes consist of short-thick rods forming the lobes protruding in-plane from the faceted core consisting of long rods. Both lobes and core exhibit fluid-like order and dynamics. Adjacent lobes and the faceted liquid interfaces are connected via a single junction.

The degree of phase separation is quantified by using fluorescently labeled fdY21M-PEG (short-thick) rods in the membranes and imaging them with fluorescence microscopy. The fluorescence intensity of the lobes and the core is measured, and the intensity ratio is computed. Figure 12 shows that the intensity ratio increases monotonically with dextran concentration and eventually saturates to a finite value. The size of the lobes, measured as the maximum protrusion length from the faceted liquid core, increases with an increase in the composition of short-thick rods (inset of Figure 12). As the number of lobes decreases, the cyclic polygon-shaped membrane transitions to a nearly concentric circle (Figure 13). It is to be noted that the pinning junctions in the cyclic polygon membrane can be removed using an optical trap, but they cannot be created or engineered as other defects in colloidal membrane systems, however it can be manipulated with an optical trap.

1. Discussion and Conclusion

Condensed matter physicists have long been intrigued by the challenge of connecting macroscopic properties of a system to the microscopic behavior of its constituent particles. One such phenomenon that captivated researchers in the mid-twentieth century was the crystalline order observed in a suspension of hard spheres, which lacked any inter-particle interaction except for a repulsive hard wall potential at close distances. Another interesting phenomenon involves rod-like anisotropic particles that experience an entropy-driven ordering transition due to excluded volume interactions. This has led to the study of liquid

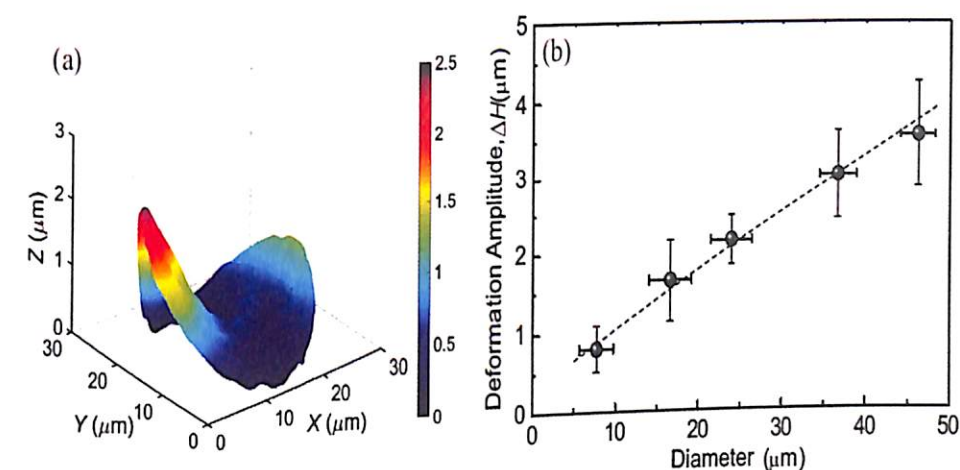


Figure 8: Macroscopic structure of crystalline colloidal membranes. (a) 3D rendered confocal image of a single nucleation center. Sidebar shows the color code for z-axis crystalline colloidal membrane formed out of a single nucleation center. (b) Deformation amplitude, $\Delta H = H_{max} - H_{min}$ as a function of diameter of the colloidal membrane. The raw data corresponding to different membrane diameters were binned to compute a mean deformation. The standard deviation of a particular bin has been used as the error bar for that bin.

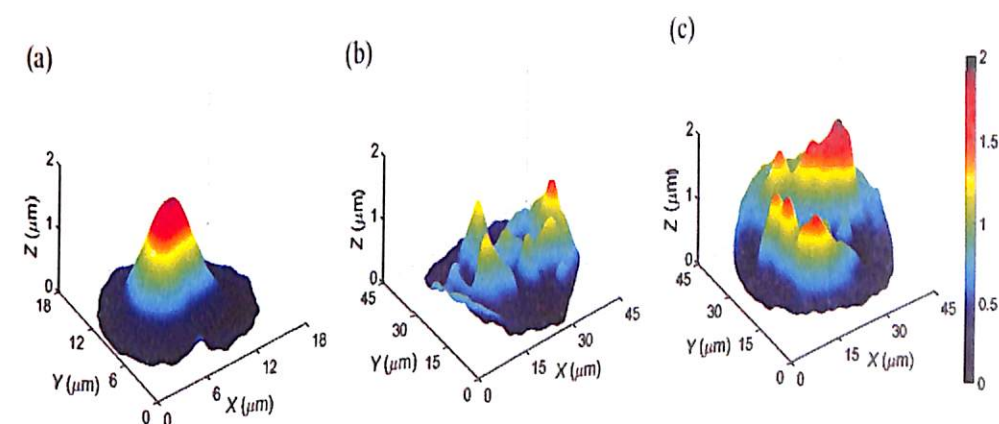


Figure 9: Wrinkling in crystalline membrane. 3D rendered confocal images of crystalline colloidal membranes formed out of chiral uid membranes containing (a) two, (b) three, (c) four nucleation centers. All panels have the same color code for z-axis values as shown in the sidebar of (c).

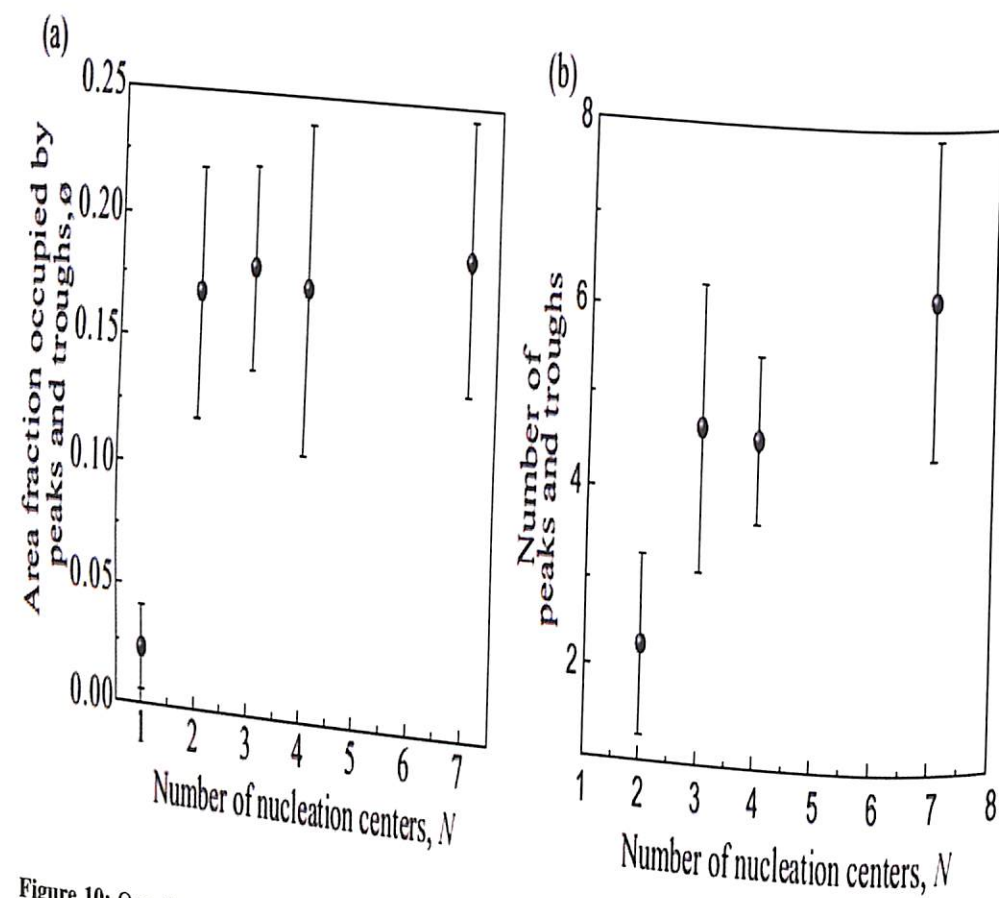


Figure 10: Quantitative measure of surface morphology. (a) ϕ , area fraction of crystalline membranes covered by peaks and valleys, as a function of number of nucleation centers, N . The error bars correspond to the standard deviation of ϕ . (b) The number of peaks and troughs within crystalline membranes as a function of number of nucleation centers (N) for membranes formed out of more than one nucleation center. The error bars correspond to the standard deviation of the number of peaks and troughs.

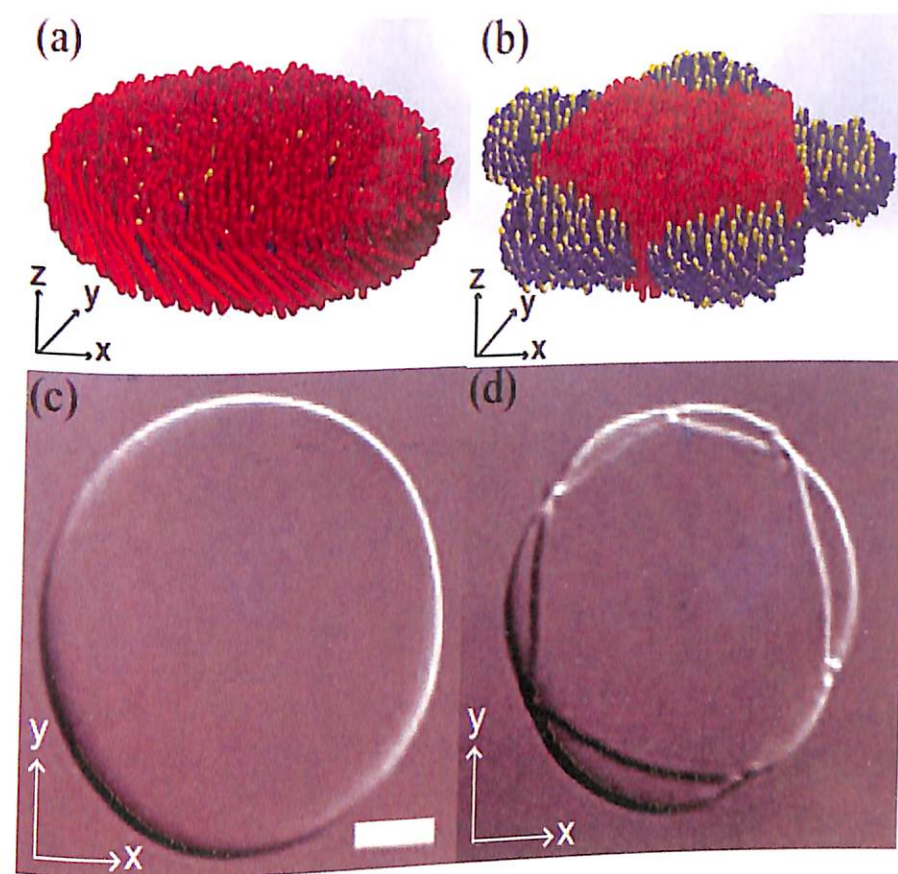


Figure 11: Cyclic polygon shaped colloidal membranes assembled from fdY21M-PEG and M13K07Y21M mixtures. (a) Schematic of uniformly mixed fdY21M-PEG and M13K07Y21M rods within a cyclic membrane. (b) Schematic of completely demixed fdY21M-PEG and M13K07Y21M rods within a cyclic membrane. (c) DIC image of a colloidal membrane with uniformly mixed fdY21M-polygon shaped colloidal membrane. (d) DIC image of a cyclic polygon shaped colloidal membrane with outer lobes composed of predominantly fdY21M-PEG rods and the inner faceted core composed of predominantly M13K07Y21M rods. The cyclic polygon membrane was assembled at a polymer concentration of 58 mg/ml.

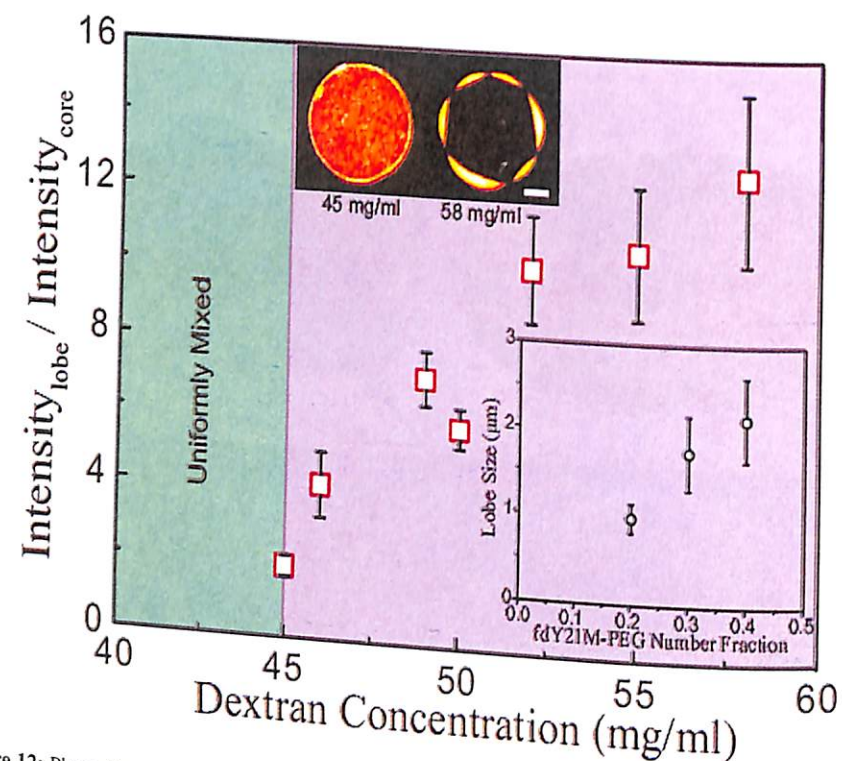


Figure 12: Phase diagram of cyclic polygon shaped membranes. Ratio of lobes and the inner core fluorescence signal intensity as a function of dextran concentration. Data points are an average of the intensity ratio measured for 639 lobes across multiple membranes and sample chambers, with the standard deviation plotted as error bars. The inset shows the size of the lobe (maximum protrusion length of the lobe) as a function of the number fraction of fdY21M-PEG rods for membranes assembled at 52 mg/ml. Data points are an average of the lobe sizes measured for 227 lobes across multiple membranes and sample chambers, with the standard deviation plotted as error bars. Inset images show colloidal membranes of fluorescently labelled fdY21M-PEG and unlabelled M13KO7Y21M at 45mg/ml and 58 mg/ml dextran concentration. Scale bar, 5 μ m.

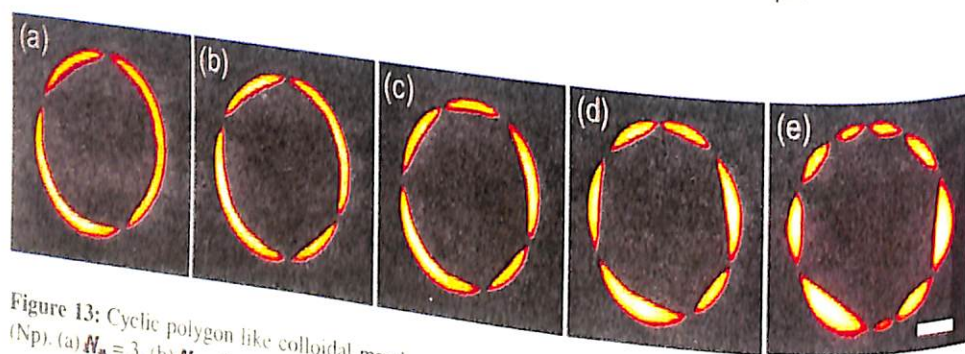


Figure 13: Cyclic polygon like colloidal membranes with sequentially increasing number of pinning junctions (N_p). (a) $N_p=3$, (b) $N_p=4$, (c) $N_p=5$, (d) $N_p=6$ and (e) $N_p=9$. Scale bar, 5 μ m.

crystals and their various phases, such as nematic, cholesteric, smectic, columnar, and blue phases, each displaying distinct arrangements of particles. However, relating the microscopic properties of the constituent particles to the macroscopic arrangement in the system is a formidable challenge. For instance, various filamentous viruses, despite having similar structures and microscopic chirality, form different phases like cholesteric and nematic. The complexity of chiral interactions in the formation of these phases has been explored, particularly when polymer grafted virus rods form cholesteric phase, where a layer of polymer brush coating could possibly screen the chiral interactions. Therefore, the intrinsic chiral interaction of the filamentous viruses remains not fully understood, however it plays a crucial role in self-assembly of colloidal membrane phases. In this article, we have discussed the structural transformation of colloidal membranes composed of rod-like viruses due to internal phase transitions based on intrinsic interactions and physical properties of the constituent particles. Dilute isotropic suspensions of rod-like viruses self-assemble into colloidal monolayer membranes of aligned rods through depletion attraction. These 2D fluid colloidal membranes are circular, macroscopically flat, and exhibit long wavelength thermal elastic fluctuations. Upon crystallization, a curvature instability sets-in in the colloidal membranes due to chirality of the virus rods leading to local wrinkling and global buckling of the membrane and eventually transforming it into three-dimensional structures. The emergence of curvature instability in crystalline membranes is associated with the increase in intrinsic chiral interaction energy, which is released through the formation of local ridges and out-of-plane global buckling. The surface roughness of the crystalline membrane can be controlled by adjusting the number of nucleation centers. Interestingly, weakly chiral colloidal membranes do not display curvature instability during crystallization, indicating the essential role of virus chirality in this phenomenon. Furthermore, a binary mixture of homochiral rods with different aspect ratios self-assemble into an unexpected cyclic polygon-shaped fluid colloidal membrane with the addition of non-adsorbing polymer. Defective rods at the membrane edge act as mobile pinning centers, compressing into one location and pinning of different rod interfaces. These pinning centers form the cornerstones of the cyclic polygon geometry and are weakly cross-linked, however these are not any topological defects. Unlike conventional self-assembly approaches that aim to avoid disorders, this work shows how disorder can be harnessed to assemble higher-order self-assembled structures with rod-shaped colloidal particles.

References

- [1] B. Thompson, N. Mignet, H. Hofland, D. Lamons, J. Seguin, C. Nicolazzi, N. de la Figuera, R. L. Kuen, X. Y. Meng, D. Scherman, and M. Bessodes. Neutral postgrafted colloidal particles for gene delivery. *Bioconjugate Chemistry*, 16(3):608–614, May 2005.
- [2] A. Fahr, K. Müller, Th. Nahde, Rolf Müller, and Sabine Brüsselbach. A new colloidal lipidic system for gene therapy. *Journal of Liposome Research*, 12(1-2):37–44, 2002. PMID: 12604036.

- [3] Monzer Fanun. *Colloids in Drug Delivery*. CRC Press, 2010.
- [4] Ximei Qian, Xiang-Hong Peng, Dominic O. Ansari, Qiqin Yin-Goen, Georgia Z. Chen, Dong M. Shin, Lily Yang, Andrew N. Young, May D. Wang, and Shuming Nie. In vivo tumor targeting and spectroscopic detection with surface-enhanced raman nanoparticle tags. *Nature Biotechnology*, 26:83 EP–, Dec 2007.
- [5] O. D. Velev and E. W. Kaler. In situ assembly of colloidal particles into miniaturized biosensors. *Langmuir*, 15(11):3693–3698, May 1999.
- [6] Albert Einstein and R. Fu"rth. *Investigations on the theory of Brownian movement*. Dover Publications, New York, N.Y., 1956.
- [7] E. R. Brownian movement and molecular reality. *Nature*, 86(2160):105–105, 1911.
- [8] S. Auer and D. Frenkel. Numerical prediction of absolute crystallization rates in hard-sphere colloids. *The Journal of Chemical Physics*, 120(6):3015–3029, 2004.
- [9] Eric R. Weeks, J. C. Crocker, Andrew C. Levitt, Andrew Schofield, and D. A. Weitz. Three-dimensional direct imaging of structural relaxation near the colloidal glass transition. *Science*, 287(5453):627–631, 2000.
- [10] Hong-Ren Jiang, Hirofumi Wada, Natsuhiko Yoshinaga, and Masaki Sano. Manipulation of colloids by a nonequilibrium depletion force in a temperature gradient. *Phys. Rev. Lett.*, 102:208301, May 2009.
- [11] Linhan Lin, Eric H. Hill, Xiaolei Peng, and Yuebing Zheng. Optothermal manipulations of colloidal particles and living cells. *Accounts of Chemical Research*, 51(6):1465–1474, Jun 2018.
- [12] Tara D. Edwards and Michael A. Bevan. Controlling colloidal particles with electric fields. *Langmuir*, 30(36):10793–10803, Sep 2014.
- [13] Angel Martinez, Hector C. Mireles, and Ivan I. Smalyukh. Large-area optoelastic manipulation of colloidal particles in liquid crystals using photoresponsive molecular surface monolayers. *Proceedings of the National Academy of Sciences*, 108(52):20891–20896, 2011.
- [14] B. J. Alder and T. E. Wainwright. Phase transition for a hard sphere system. *The Journal of Chemical Physics*, 27(5):1208–1209, 1957.
- [15] Lars Onsager. The effects of shape on the interaction of colloidal particles. *Annals of the New York Academy of Sciences*, 51(4):627–659, 1949.
- [16] Seth Fraden. *Phase Transitions in Colloidal Suspensions of Virus Particles*, pages 113–164. Springer Netherlands, Dordrecht, 1995.
- [17] Jianxin Tang and Seth Fraden. Isotropic-cholesteric phase transition in colloidal suspensions of filamentous bacteriophage fd. *Liquid Crystals*, 19(4):459–467, 1995.
- [18] Sho Asakura and Fumio Oosawa. On interaction between two bodies immersed in a solution of macromolecules. *The Journal of Chemical Physics*, 22(7):1255–1256, 1954.
- [19] Zvonimir Dogic and Seth Fraden. *Phase Behavior of Rod-Like Viruses and Virus-Sphere Mixtures*, chapter 1, pages 1–86. John Wiley & Sons, Ltd, 2007.
- [20] Eric Grelet. Hard-rod behavior in dense mesophases of semiflexible and rigid charged viruses. *Phys. Rev. X*, 4:021053, Jun 2014.
- [21] Edward Barry and Zvonimir Dogic. Entropy driven self-assembly of non-amphiphilic colloidal membranes. *Proceedings of the National Academy of Sciences*, 107(23):10348–10353, 2010.
- [22] Thomas Gibaud, Edward Barry, Mark J. Zakhary, Mir Henglin, Andrew Ward, Yasheng Yang, Cristina Berciu, Rudolf Oldenbourg, Michael F. Hagan, Daniela Nicastro, Robert B. Meyer, and Zvonimir Dogic. Re-configurable self-assembly through chiral control of interfacial tension. *Nature*, 481:348 EP–, Jan 2012.
- [23] Thomas Gibaud, C. Nadir Kaplan, Perna Sharma, Mark J. Zakhary, Andrew Ward, Rudolf Oldenbourg, Robert B. Meyer, Randall D. Kamien, Thomas R. Powers, and Zvonimir Dogic. Achiral symmetry breaking and positive gaussian modulus lead to scalloped colloidal membranes. *Proceedings of the National Academy of Sciences*, 114(17):E3376–E3384, 2017.
- [24] M R Green and J Sambrook. *Molecular Cloning*. Cold Spring Harbor Laboratory Press, 2012.
- [25] Lachit Saikia, Tanmoy Sarkar, Meera Thomas, V. A. Raghunathan, Anirban Sain, and Perna Sharma. Curvature instability of chiral colloidal membranes on crystallization. *Nature Communications*, 8(1):1160, Oct 2017.
- [26] Yasheng Yang, Edward Barry, Zvonimir Dogic, and Michael F. Hagan. Self-assembly of 2d membranes from mixtures of hard rods and depleting polymers. *Soft Matter*, 8:707–714, 2012.
- [27] Stefan Luding. Global equation of state of two-dimensional hard sphere systems. *Phys. Rev. E*, 63:042201, Mar 2001.
- [28] Christopher B. Stanley and Helmut H. Strey. Measuring osmotic pressure of poly(ethylene glycol) solutions by sedimentation equilibrium ultracentrifugation. *Macromolecules*, 36(18):6888–6893, Sep 2003.
- [29] Lachit Saikia and Perna Sharma. Self assembly of cyclic polygon shaped fluid colloidal membranes through pinning. *Soft Matter*, 14:9959–9966, 2018.
- [30] Edward Barry, Daniel Beller, and Zvonimir Dogic. A model liquid crystalline system based on rodlike viruses with variable chirality and persistence length. *Soft Matter*, 5:2563–2570, 2009.

Obesity and Its Treatment with Herbal Remedies

Kalyani Pathak⁽¹⁾
Parimita Kalita⁽²⁾
Aparoop Das⁽³⁾
Riya Saikia⁽⁴⁾

^{(1),(2),(4)}Dept of Pharmaceutical sciences,
Dibrugarh University, Dibrugarh, Assam

⁽³⁾School of Pharmacy,
The Assam Kaziranga University, Jorhat, Assam
Email: kalyakster@gmail.com

Abstract

Obesity has emerged as a major global health issue, associated with numerous diseases such as diabetes, heart disease, and certain cancers. Conventional methods of treating obesity, including diet, exercise, and pharmaceutical interventions, often prove insufficient for sustainable weight loss or are associated with adverse side effects. Consequently, there has been a growing interest in the potential role of herbal remedies in the treatment of obesity. This paper reviews the literature on herbal remedies used in the treatment of obesity, examining both their efficacy and safety profiles. A wide range of plant species have been studied for their potential anti-obesity effects, including green tea, *Garcinia cambogia*, *Hibiscus sabdariffa*, and *Capsicum annuum*. These herbs are proposed to exert their effects through various mechanisms, such as enhancing fat metabolism, suppressing appetite, and inhibiting fat absorption. Overall, preliminary evidence suggests that certain herbal remedies may offer promising adjunctive treatment for obesity. However, more extensive clinical trials are needed to confirm these effects and evaluate the long-term safety of these herbal treatments. The review also underscores the importance of standardizing herbal products to ensure consistency in their composition and effects. This book chapter provides insights into alternative and complementary approaches to tackle the obesity epidemic, potentially offering a novel direction for future research and treatment strategies.

K. Pathak, P. Kalita, A. Das, R. Saikia, "Obesity and Its Treatment with Herbal Remedies", Horizon (Ed. Book), 74-91 (2023), copyright: Tinsukia College Publishing, ISBN: 978-81-19322-07-7

1. Introduction

Obesity has become a major global health concern, with its prevalence reaching epidemic proportions in recent years. It is characterized by the excessive accumulation of body fat, leading to numerous health complications and an increased risk of chronic diseases such as cardiovascular disorders, type 2 diabetes, and certain cancers. The World Health Organization (WHO) has recognized obesity as a significant public health challenge, highlighting the need for effective prevention and treatment strategies. Conventional approaches to obesity management, including dietary interventions, physical activity, pharmacotherapy, and bariatric surgery, form the foundation of treatment. However, the increasing interest in natural remedies has sparked a growing exploration of herbal treatments as adjunctive therapies for obesity. Herbs, derived from plants and used in various traditional medicine systems across cultures, offer a unique perspective in addressing the multifaceted nature of obesity. Historically, herbs have been valued for their medicinal properties and have been used for centuries in traditional healing practices. They contain a diverse array of bioactive compounds that may exert physiological effects on the body. While herbs are not a standalone solution for weight loss, they may provide valuable support to conventional approaches and contribute to improved outcomes.^[1]

This chapter aims to provide a comprehensive overview of obesity and delve into the role of herbal remedies in its treatment. We will begin by examining the underlying causes of obesity, which involve a complex interplay of genetic, environmental, and lifestyle factors. The global rise in obesity can be attributed to changes in dietary patterns, sedentary lifestyles, and socioeconomic factors, among others. Understanding these causes is crucial for developing effective treatment strategies. Obesity is associated with a myriad of health consequences that significantly impact an individual's quality of life and overall well-being. These consequences include an increased risk of cardiovascular diseases, insulin resistance, metabolic syndrome, and psychological disorders such as depression and anxiety. By addressing obesity, we can not only improve weight management but also reduce the risk of developing these associated health conditions. Conventional approaches to obesity management have shown varying degrees of success. Dietary interventions, such as calorie restriction and balanced nutrition, are key components of weight loss programs. Physical activity, including both aerobic exercise and strength training, plays a vital role in enhancing energy expenditure and promoting overall health. Pharmacotherapy, including prescription medications, may be recommended for individuals with obesity-related comorbidities. In severe cases, bariatric surgery can be an effective option for sustained weight loss. However, these interventions may have limitations, and there is a need for additional strategies to enhance treatment outcomes.^[2]

Herbal remedies have garnered increasing attention as potential complementary therapies for

obesity. Many herbs possess bioactive compounds that have been found to influence various aspects of metabolism, appetite regulation, and fat storage. While the scientific evidence supporting their efficacy is still evolving, numerous studies have explored the potential of specific herbs in aiding weight loss and improving metabolic health. One herb that has gained significant attention in the context of obesity treatment is green tea (*Camellia sinensis*). Green tea contains catechins, particularly epigallocatechin gallate (EGCG), which have been shown to have thermogenic and fat oxidation properties. Clinical studies have demonstrated that green tea consumption may lead to modest weight loss and improvements in body composition. Garcinia Cambogia (*Garcinia gummi-gutta*) is another herb that has attracted interest for its potential anti-obesity effects. Its active ingredient, hydroxycitric acid (HCA), has been studied for its ability to inhibit an enzyme involved in fat synthesis and suppress appetite. While some studies have shown promising results, the overall evidence is still limited and conflicting. Cinnamon (*Cinnamomum verum*), ginger (*Zingiber officinale*), turmeric (*Curcuma longa*), and fenugreek (*Trigonella foenum-graecum*) are among other herbs that have been investigated for their potential benefits in obesity management. These herbs have been found to influence various metabolic pathways, including glucose and lipid metabolism, inflammation, and appetite regulation. Despite the potential benefits of herbal remedies, it is important to approach their use with caution. Lack of standardization, potential herb-drug interactions, and the need for proper regulation and quality control are important considerations. Additionally, the effectiveness of herbal treatments can vary among individuals, and more research is needed to identify specific subgroups that may benefit the most. In conclusion, obesity is a complex health issue with far-reaching consequences. While conventional approaches to obesity management remain fundamental, herbal remedies offer a complementary and holistic perspective on weight loss and improved metabolic health. By exploring the potential of herbs, we can expand our understanding of natural therapies and their role in combating obesity. However, further research is needed to establish the efficacy, safety, and optimal use of herbal treatments in the context of obesity management. Obesity has become a global epidemic, affecting individuals of all age groups and socioeconomic backgrounds. It is characterized by excessive body fat accumulation, leading to an increased risk of various chronic diseases such as cardiovascular disorders, diabetes, and certain cancers. While conventional approaches to weight management, such as diet and exercise, remain fundamental, there is growing interest in exploring the potential of herbal remedies as adjuvant therapies. This chapter aims to provide an overview of obesity and discuss the use of herbs in its treatment.^[3]

2. Understanding Obesity

2.1 Definition and Prevalence

Obesity is defined as the excessive accumulation of body fat, resulting in an increased body weight beyond what is considered healthy for a given height and age. Body Mass Index

(BMI), calculated by dividing an individual's weight in kilograms by the square of their height in meters, is commonly used as a measure to classify obesity. According to the World Health Organization (WHO), a BMI of 30 or higher is classified as obesity.^[4]

The prevalence of obesity has reached alarming levels globally. In both developed and developing countries, the rates of obesity have been steadily rising over the past few decades. It is estimated that more than 650 million adults worldwide were obese in 2016, representing approximately 13% of the global population. The prevalence of childhood obesity has also increased significantly, affecting over 340 million children and adolescents aged 5-19 years in 2016.^[5]

2.2 Causes and Risk Factors

Obesity is a complex condition with multiple underlying causes and risk factors. It is influenced by a combination of genetic, environmental, behavioral, and socioeconomic factors.^[6]

Genetic factors play a role in determining an individual's susceptibility to obesity. Certain genes can affect appetite regulation, metabolism, and fat storage, contributing to weight gain and obesity. However, genetic factors alone cannot account for the rapid increase in obesity rates, suggesting that environmental factors have a significant impact.^[6]

Environmental factors, such as the modern sedentary lifestyle and easy access to high-calorie foods, contribute to the development of obesity. Changes in dietary patterns, characterized by the consumption of energy-dense, nutrient-poor foods, coupled with a decrease in physical activity, have led to an energy imbalance, with more calories consumed than expended. Additionally, the obesogenic environment, characterized by an abundance of fast food outlets, sedentary jobs, and limited access to healthy foods, promotes unhealthy behaviors and further exacerbates the obesity epidemic.^[7]

Behavioral factors, including eating habits, physical activity levels, and lifestyle choices, also play a significant role in the development and progression of obesity. Poor dietary choices, such as high intake of sugary beverages, processed foods, and large portion sizes, contribute to excessive calorie consumption. Sedentary behaviors, such as prolonged sitting and low levels of physical activity, reduce energy expenditure and promote weight gain.^[8]

Socioeconomic factors, such as income, education, and access to healthcare, can influence the prevalence of obesity. Individuals from lower socioeconomic backgrounds may face barriers to accessing healthy foods, safe recreational spaces, and quality healthcare, making it more challenging to adopt healthy lifestyles and manage obesity effectively.^[9]

2.3 Health Consequences

Obesity is associated with a wide range of health consequences, affecting nearly every organ system in the body. Excess body fat not only alters physical appearance but also contributes to numerous chronic diseases and health conditions.^[10]

Cardiovascular disorders, including hypertension (high blood pressure), coronary artery disease, and stroke, are more prevalent in individuals with obesity. The excess body fat increases the workload on the heart, leading to elevated blood pressure and an increased risk of atherosclerosis (hardening of the arteries).^[11]

Type 2 diabetes is strongly associated with obesity, particularly abdominal obesity. Adipose tissue, or fat cells, release hormones and inflammatory substances that interfere with insulin signaling and glucose metabolism, leading to insulin resistance and the development of diabetes.^[12]

Obesity is also a significant risk factor for certain types of cancer, including breast, colorectal, and endometrial cancer. Adipose tissue produces hormones and growth factors that can promote tumor growth and metastasis.^[6]

Other health consequences of obesity include respiratory disorders (such as sleep apnea and asthma), musculoskeletal disorders (such as osteoarthritis), liver disease (including non-alcoholic fatty liver disease)^[13]

3. Conventional Approaches to Obesity Management

The management of obesity involves a comprehensive and multidimensional approach that targets various aspects of weight loss, improved metabolic health, and long-term weight maintenance. Conventional approaches to obesity management encompass lifestyle modifications, dietary interventions, physical activity, pharmacotherapy, and, in severe cases, bariatric surgery.^[13]

3.1 Dietary Interventions

Dietary interventions are a fundamental component of obesity management. They focus on achieving energy balance by reducing caloric intake, improving the quality of the diet, and promoting sustainable changes in eating habits. The goals of dietary interventions include weight loss, improvement in metabolic parameters, and long-term weight maintenance. Calorie restriction is often a key strategy in dietary interventions. It involves reducing overall energy intake through portion control, moderation of high-calorie foods, and the incorporation of nutrient-dense foods. A balanced approach that emphasizes a variety of whole foods, including fruits, vegetables, whole grains, lean proteins, and healthy fats, is recommended. Specific dietary patterns have shown effectiveness in promoting weight loss and improving metabolic health. The Mediterranean diet, characterized by high consumption of fruits, vegetables, whole grains, legumes, and healthy fats (such as olive oil), has been associated with favorable outcomes in weight management and cardiovascular health. Behavioral strategies, such as mindful eating, portion control, and self-monitoring, are often integrated into dietary interventions to promote long-term adherence and sustainable changes in eating behaviors.^[14]

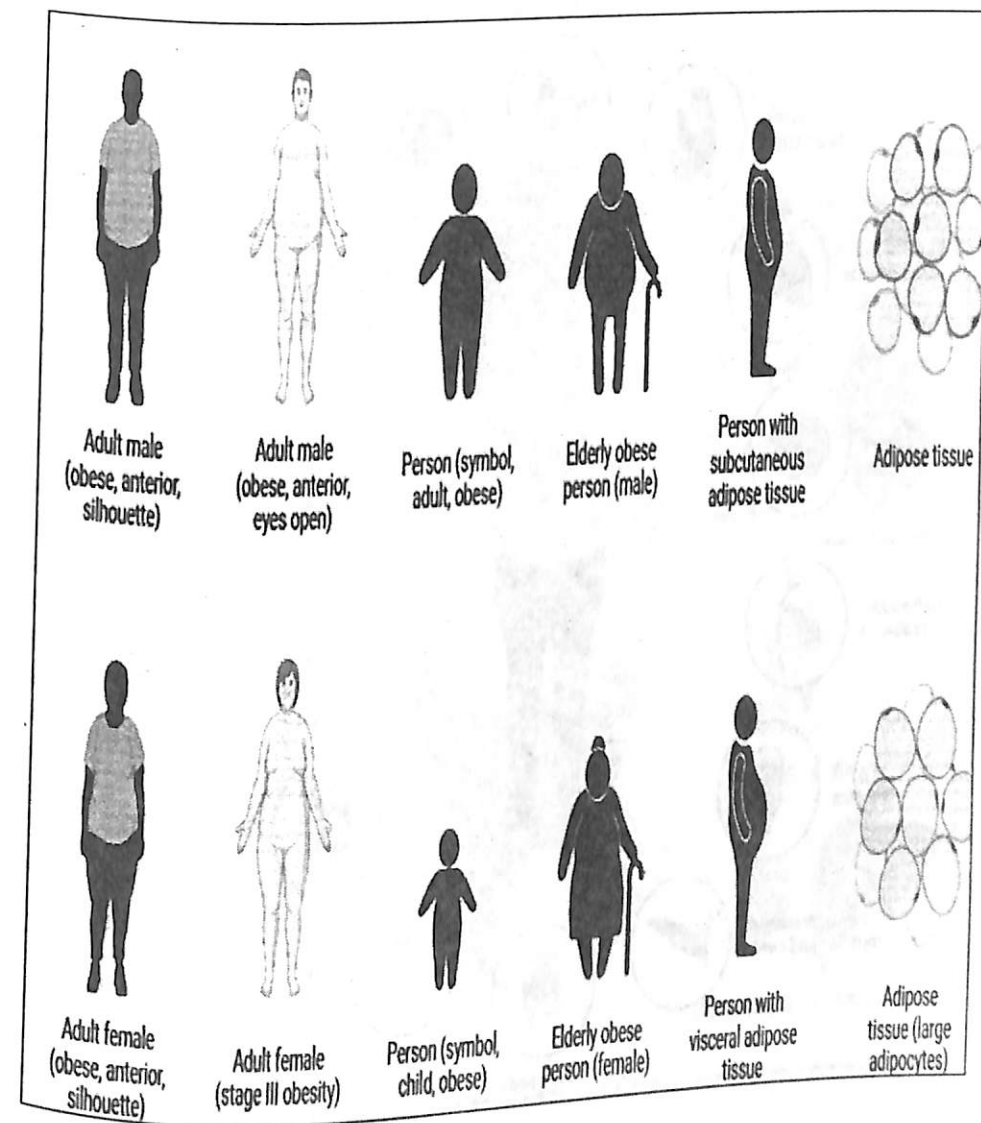


Figure 1: Obesity in Human

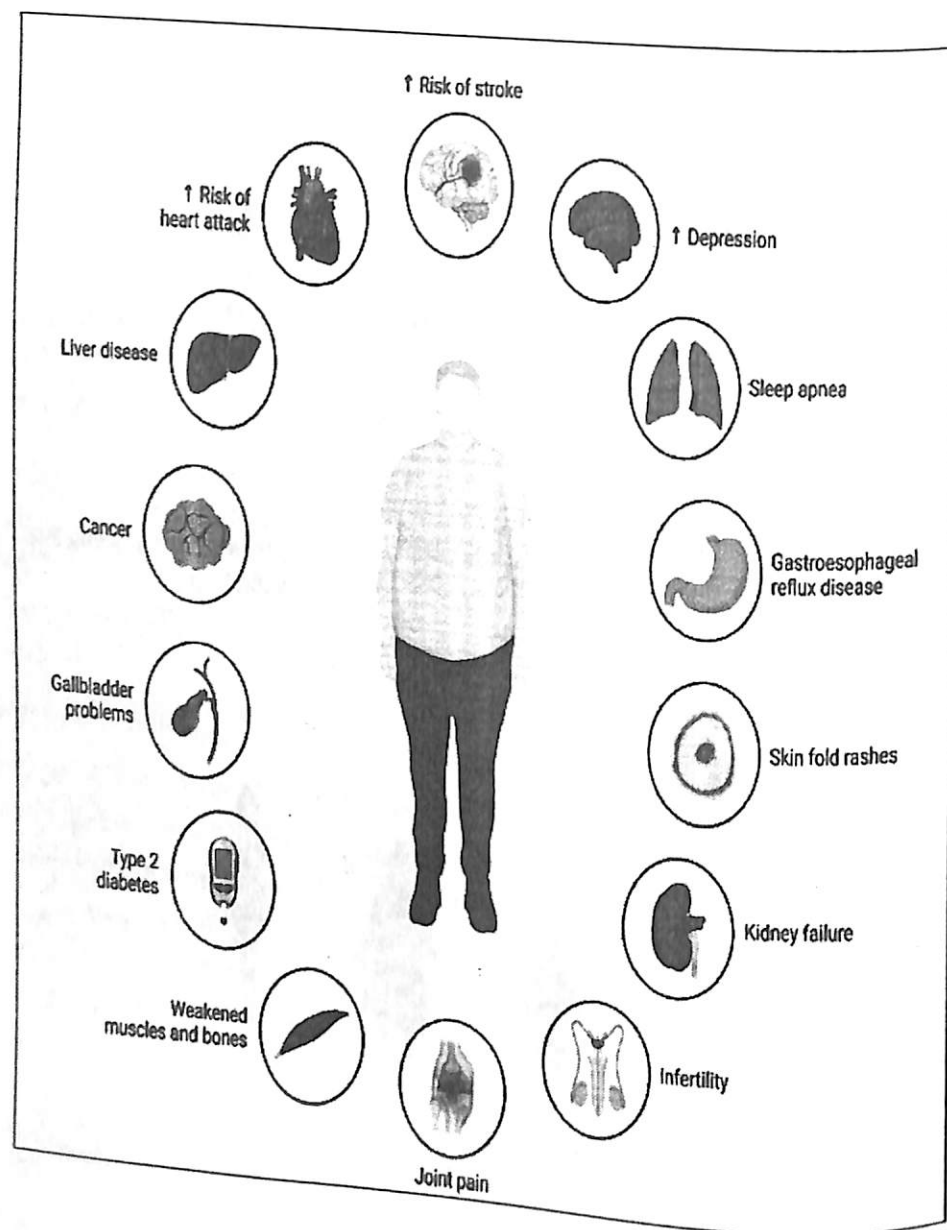


Figure 2: The Effect of Obesity on the Body

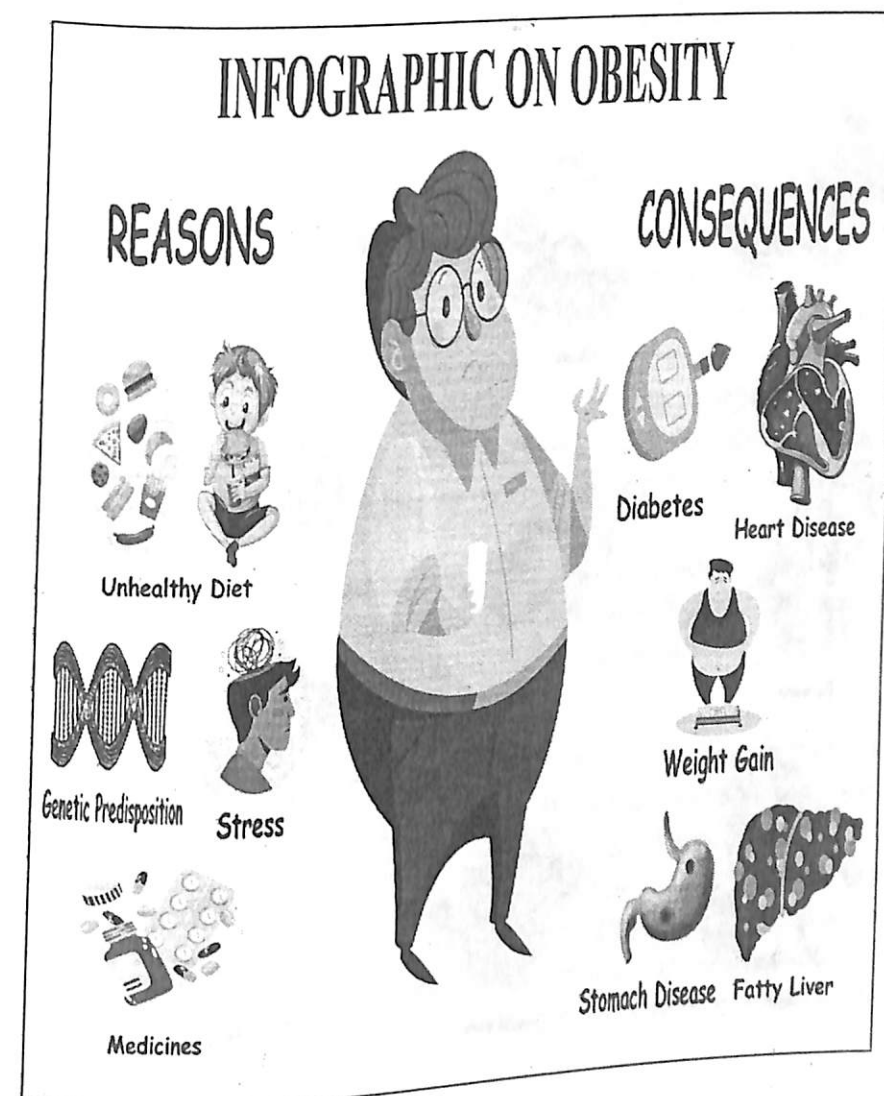


Figure 3: Infographic on obesity

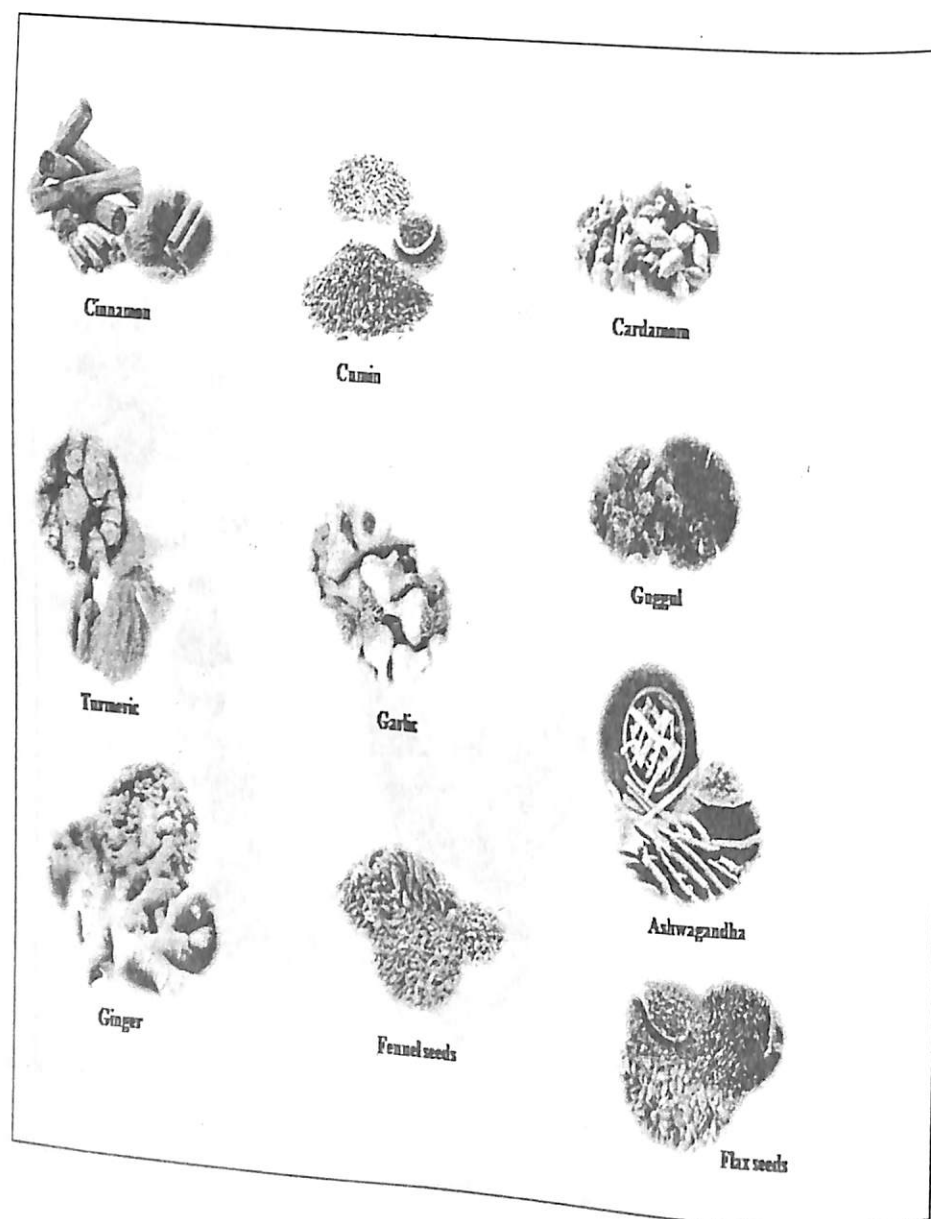


Figure 4: Treatment of obesity using common herbs

3.2 Physical Activity

Regular physical activity is crucial in the management of obesity. Exercise and physical activity increase energy expenditure, promote fat loss, improve cardiovascular fitness, and enhance overall well-being. The American Heart Association and other organizations recommend a minimum of 150 minutes of moderate-intensity aerobic activity or 75 minutes of vigorous-intensity aerobic activity per week, along with muscle-strengthening activities on two or more days per week. A combination of aerobic exercise (such as brisk walking, jogging, cycling) and resistance training (such as weightlifting, bodyweight exercises) is recommended to optimize weight loss, preserve muscle mass, and improve metabolic function. Physical activity interventions may also include behavioral strategies, such as goal setting, self-monitoring, and social support, to enhance adherence and long-term maintenance of exercise habits.^[15]

3.3 Pharmacotherapy

Pharmacotherapy may be considered as an adjunctive treatment for individuals with obesity who have not achieved adequate weight loss with lifestyle interventions alone. Medications approved for obesity treatment aim to suppress appetite, reduce food intake, or interfere with fat absorption. Prescription medications, such as orlistat, liraglutide, and phentermine/topiramate, are available for short-term use under medical supervision. These medications are typically prescribed in combination with lifestyle modifications and are intended for individuals with a BMI of 30 or higher, or those with a BMI of 27 or higher with obesity-related co morbidities.^[15]

It is important to note that pharmacotherapy is not a standalone solution for obesity management. The use of medications should be carefully monitored by healthcare professionals, and potential side effects and contraindications should be considered.^[16]

3.4 Bariatric Surgery

Bariatric surgery is a consideration for individuals with severe obesity (BMI of 40 or higher) or those with a BMI of 35 or higher with significant obesity-related comorbidities. It involves various surgical procedures, such as gastric bypass, sleeve gastrectomy, and gastric banding, which modify the anatomy of the digestive system to restrict food intake, reduce absorption, or both. Bariatric surgery can lead to significant weight loss and improvement in metabolic parameters. It is often recommended for individuals who have not achieved successful weight loss with non-surgical interventions or those with severe obesity-related health complications.

4. Herbal Remedies for Obesity Treatment

In recent years, there has been growing interest in exploring the potential of herbal remedies as adjunctive therapies for obesity treatment. Herbs, derived from plants and used in traditional medicine systems across cultures, contain bioactive compounds that may exert physiological effects relevant to weight management. While herbal remedies are not a standalone solution for obesity, they may complement conventional approaches and contribute to improved

outcomes. This section will discuss several herbs that have shown promise in scientific studies for their potential anti-obesity effects. [16]

4.1 Green Tea (*Camellia sinensis*)

Green tea, derived from the leaves of *Camellia sinensis*, is one of the most widely consumed beverages worldwide. It contains bioactive compounds, including catechins, particularly epigallocatechin gallate (EGCG), which have been studied for their potential role in weight management. [16]

Mechanisms of Action: Green tea catechins have been found to enhance energy expenditure and fat oxidation. They may also inhibit enzymes involved in fat synthesis and stimulate thermogenesis, thereby promoting weight loss. Furthermore, green tea has antioxidant and anti-inflammatory properties, which may be beneficial for overall metabolic health.

Clinical Evidence: Numerous studies have examined the effects of green tea catechins on body weight and body composition. Some studies have reported modest reductions in body weight, body mass index (BMI), and waist circumference with green tea consumption. However, the overall evidence is still limited and inconsistent, with variations in study designs, dosages, and participant characteristics. [16]

Safety Considerations: Green tea is generally considered safe for consumption as a beverage. However, high doses of green tea extract or supplements containing concentrated catechins should be used with caution due to potential hepatotoxicity and interactions with certain medications. [16]

4.2 Garcinia gummi-gutta (*Garcinia Cambogia*)

Garcinia Cambogia, also known as Malabar tamarind, is a tropical fruit commonly used in traditional culinary practices. The active ingredient in *Garcinia Cambogia* is hydroxycitric acid (HCA), which has been studied for its potential effects on appetite suppression and fat metabolism. [17]

Mechanisms of Action: HCA is believed to inhibit an enzyme called citrate lyase, which plays a role in converting excess carbohydrates into fat. By inhibiting this enzyme, HCA may reduce fat synthesis and promote fat oxidation. HCA may also increase serotonin levels in the brain, potentially reducing appetite and food cravings. [17]

Clinical Evidence: Studies investigating the effects of *Garcinia Cambogia* on weight loss have shown mixed results. Some studies have reported modest reductions in body weight and body fat percentage, while others have found no significant differences compared to placebo. The overall evidence is limited, and larger, well-designed studies are needed to determine the effectiveness of *Garcinia Cambogia* as an anti-obesity agent. [17]

Safety Considerations: *Garcinia Cambogia* is generally considered safe when consumed as part of a balanced diet. However, high doses of HCA or the use of *Garcinia Cambogia* supplements may cause gastrointestinal discomfort, headache, and in rare cases, hepatotoxicity. Individuals with liver disease or taking medications should exercise caution and consult with a healthcare professional before using *Garcinia Cambogia* products. [17]

4.3 Cinnamon (*Cinnamomum verum*)

Cinnamon is a spice derived from the bark of *Cinnamomum verum* or *Cinnamomum cassia* trees. It has been used for its culinary and medicinal properties for centuries. Cinnamon contains various bioactive compounds, including cinnamaldehyde, which have been studied for their potential effects on blood sugar regulation and weight management.

Mechanisms of Action: Cinnamon has been found to improve insulin sensitivity and enhance glucose metabolism. It may also delay gastric emptying, leading to increased feelings of fullness and reduced appetite. Additionally, cinnamon may have anti-inflammatory effects that could be beneficial for metabolic health.

Clinical Evidence: Several small-scale studies have explored the effects of cinnamon supplementation on weight loss and metabolic parameters. While some studies have reported improvements in body weight, body fat percentage, and metabolic markers such as blood glucose and lipid levels, the evidence is limited and inconsistent. Further well-designed studies are needed to establish the efficacy of cinnamon as an anti-obesity remedy. [17]

Safety Considerations: Cinnamon is generally safe for consumption as a spice in food amounts. However, high doses of cinnamon supplements or prolonged use of cinnamon oil may be associated with potential side effects, including liver toxicity and allergic reactions. Individuals with liver disease or taking medications should exercise caution and consult with a health care professional before using cinnamon supplements. [17]

4.4 Ginger (*Zingiber officinale*)

Ginger is a widely used spice and herbal remedy derived from the rhizome of *Zingiber officinale*. It has a long history of traditional use for its various health benefits, including potential effects on weight management.

Mechanisms of Action: Ginger contains bioactive compounds, such as gingerol and shogaol, which have been studied for their potential thermogenic and appetite-suppressing effects. Ginger may increase thermogenesis, which can lead to increased calorie expenditure and fat burning. It may also influence satiety and reduce food intake. [17]

Clinical Evidence: Several small-scale studies have explored the effects of ginger supplementation on body weight and metabolic parameters, and appetite, the overall evidence is limited and inconsistent. Further research is needed to determine the effectiveness of ginger as an anti-obesity remedy.

Safety Considerations: Ginger is generally considered safe when consumed as part of the diet or used as a spice. However, high doses of ginger supplements may cause gastrointestinal discomfort, heartburn, and interactions with certain medications. Individuals with gallstone disease, bleeding disorders, or taking anticoagulant medications should exercise caution and consult with a healthcare professional before using ginger supplements. [17]

4.5 Turmeric (*Curcuma longa*)

Turmeric, derived from the root of *Curcuma longa*, is a spice commonly used in Asian cuisine and traditional medicine. It contains the bioactive compound curcumin, which has been studied for its potential anti-inflammatory and metabolic effects. [18]

Mechanisms of Action: Curcumin has been found to modulate various molecular targets involved in obesity and metabolic regulation. It may help reduce inflammation, enhance insulin sensitivity, and inhibit fat cell differentiation. Curcumin may also increase thermogenesis and fat oxidation, leading to potential weight loss effects.

Clinical Evidence: Studies investigating the effects of curcumin on weight loss and metabolic health have shown promising results in preclinical and small-scale human studies. Some studies have reported improvements in body weight, body fat percentage, and metabolic markers such as blood glucose and lipid levels. However, larger and well-controlled studies are needed to confirm these findings and establish the efficacy of curcumin as an anti-obesity agent. [18]

Safety Considerations: Turmeric and curcumin are generally considered safe when consumed as part of the diet. However, high doses of curcumin supplements may cause gastrointestinal discomfort and interactions with certain medications. Individuals with gallstone disease, bleeding disorders, or taking anticoagulant medications should exercise caution and consult with a healthcare professional before using curcumin supplements.

4.6 Fenugreek (*Trigonella foenum-graecum*)

Fenugreek is an herb commonly used in traditional medicine and culinary practices. It contains various bioactive compounds, including soluble fiber, galactomannan, and saponins, which have been studied for their potential effects on weight management.

Mechanisms of Action: Fenugreek may help regulate blood sugar levels and improve insulin sensitivity, which can have implications for weight management. It may also increase feelings of fullness and satiety, reducing overall food intake. Additionally, fenugreek may inhibit fat accumulation and promote fat oxidation.

Clinical Evidence: Several small-scale studies have explored the effects of fenugreek supplementation on body weight, body composition, and metabolic parameters. Some studies have reported modest reductions in body weight, body fat percentage, and waist circumference with fenugreek consumption. However, the overall evidence is limited and further research is needed to establish the efficacy of fenugreek as an anti-obesity remedy.

Safety Considerations: Fenugreek is generally considered safe for consumption in food amounts or as a spice. However, high doses or prolonged use of fenugreek supplements may cause gastrointestinal discomfort, maple syrup-like odor in urine, and interactions with certain medications. Pregnant women should avoid fenugreek supplementation due to potential effects on uterine contractions. [18]

4.7 Other Herbs with Potential Anti-obesity Effects

In addition to the herbs mentioned above, there are several other herbs that have been studied for their potential anti-obesity effects. These include bitter orange (*Citrus aurantium*),

dandelion (*Taraxacum officinale*), and Hoodia gordonii. However, the evidence for their efficacy and safety in obesity management is limited and further research is needed. [19]

It is important to note that while herbal remedies may offer potential benefits in obesity management, they are not a standalone solution. The effectiveness of herbal treatments can vary among individuals, and more research is needed to identify specific subgroups that may benefit the most. Furthermore, it is crucial to approach the use of herbal remedies with caution, considering factors such as lack of standardization, potential herb-drug interactions, and the need for proper regulation and quality control. [18]

In conclusion, herbal remedies have shown promise in supporting weight loss and improving metabolic health. While scientific evidence supporting their use is still evolving, they may serve as valuable adjunctive therapies in the comprehensive management of obesity. It is recommended to consult with a healthcare professional before incorporating herbal remedies into an obesity treatment plan, especially for individuals with underlying health conditions or those taking medications.

5. Challenges and Future Directions

The use of herbal remedies as adjunctive therapies for obesity treatment presents various challenges and considerations. These challenges encompass aspects such as lack of standardization, potential herb-drug interactions, regulation and quality control, combining herbal therapies with conventional approaches, and the need for further research. Addressing these challenges and exploring future directions will contribute to a more comprehensive understanding and utilization of herbal treatments in obesity management.

5.1 Lack of Standardization

One of the challenges in herbal treatments for obesity is the lack of standardization. Herbal remedies can vary in their composition and potency due to differences in plant species, growing conditions, harvesting methods, and processing techniques. This variability can affect the bioavailability and effectiveness of the active compounds. Establishing standardized protocols for the production and quality control of herbal products is essential to ensure consistency and reproducibility in their therapeutic effects. [19]

5.2 Potential Herb-Drug Interactions

Herbal remedies may interact with prescription medications, affecting their efficacy or safety. Some herbs can interfere with drug metabolism, leading to altered drug concentrations in the body. Additionally, herbs may have additive or synergistic effects with certain medications, potentially increasing the risk of adverse reactions. Healthcare professionals should be aware of potential herb-drug interactions and advise patients accordingly, especially those with underlying health conditions or taking multiple medications.

4.5 Turmeric (*Curcuma longa*)

Turmeric, derived from the root of *Curcuma longa*, is a spice commonly used in Asian cuisine and traditional medicine. It contains the bioactive compound curcumin, which has been studied for its potential anti-inflammatory and metabolic effects.^[18]

Mechanisms of Action: Curcumin has been found to modulate various molecular targets involved in obesity and metabolic regulation. It may help reduce inflammation, enhance insulin sensitivity, and inhibit fat cell differentiation. Curcumin may also increase thermogenesis and fat oxidation, leading to potential weight loss effects.

Clinical Evidence: Studies investigating the effects of curcumin on weight loss and metabolic health have shown promising results in preclinical and small-scale human studies. Some studies have reported improvements in body weight, body fat percentage, and metabolic markers such as blood glucose and lipid levels. However, larger and well-controlled studies are needed to confirm these findings and establish the efficacy of curcumin as an anti-obesity agent.^[18]

Safety Considerations: Turmeric and curcumin are generally considered safe when consumed as part of the diet. However, high doses of curcumin supplements may cause gastrointestinal discomfort and interactions with certain medications. Individuals with gallstone disease, bleeding disorders, or taking anticoagulant medications should exercise caution and consult with a healthcare professional before using curcumin supplements.

4.6 Fenugreek (*Trigonella foenum-graecum*)

Fenugreek is an herb commonly used in traditional medicine and culinary practices. It contains various bioactive compounds, including soluble fiber, galactomannan, and saponins, which have been studied for their potential effects on weight management.

Mechanisms of Action: Fenugreek may help regulate blood sugar levels and improve insulin sensitivity, which can have implications for weight management. It may also increase feelings of fullness and satiety, reducing overall food intake. Additionally, fenugreek may inhibit fat accumulation and promote fat oxidation.

Clinical Evidence: Several small-scale studies have explored the effects of fenugreek supplementation on body weight, body composition, and metabolic parameters. Some studies have reported modest reductions in body weight, body fat percentage, and waist circumference with fenugreek consumption. However, the overall evidence is limited and further research is needed to establish the efficacy of fenugreek as an anti-obesity remedy.

Safety Considerations: Fenugreek is generally considered safe for consumption in food amounts or as a spice. However, high doses or prolonged use of fenugreek supplements may cause gastrointestinal discomfort, maple syrup-like odor in urine, and interactions with certain medications. Pregnant women should avoid fenugreek supplementation due to potential effects on uterine contractions.^[18]

4.7 Other Herbs with Potential Anti-obesity Effects

In addition to the herbs mentioned above, there are several other herbs that have been studied for their potential anti-obesity effects. These include bitter orange (*Citrus aurantium*),

dandelion (*Taraxacum officinale*), and *Hoodia gordonii*. However, the evidence for their efficacy and safety in obesity management is limited and further research is needed.^[18]

It is important to note that while herbal remedies may offer potential benefits in obesity management, they are not a standalone solution. The effectiveness of herbal treatments can vary among individuals, and more research is needed to identify specific subgroups that may benefit the most. Furthermore, it is crucial to approach the use of herbal remedies with caution, considering factors such as lack of standardization, potential herb-drug interactions, and the need for proper regulation and quality control.^[18]

In conclusion, herbal remedies have shown promise in supporting weight loss and improving metabolic health. While scientific evidence supporting their use is still evolving, they may serve as valuable adjunctive therapies in the comprehensive management of obesity. It is recommended to consult with a healthcare professional before incorporating herbal remedies into an obesity treatment plan, especially for individuals with underlying health conditions or those taking medications.

5. Challenges and Future Directions

The use of herbal remedies as adjunctive therapies for obesity treatment presents various challenges and considerations. These challenges encompass aspects such as lack of standardization, potential herb-drug interactions, regulation and quality control, combining herbal therapies with conventional approaches, and the need for further research. Addressing these challenges and exploring future directions will contribute to a more comprehensive understanding and utilization of herbal treatments in obesity management.

5.1 Lack of Standardization

One of the challenges in herbal treatments for obesity is the lack of standardization. Herbal remedies can vary in their composition and potency due to differences in plant species, growing conditions, harvesting methods, and processing techniques. This variability can affect the bioavailability and effectiveness of the active compounds. Establishing standardized protocols for the production and quality control of herbal products is essential to ensure consistency and reproducibility in their therapeutic effects.^[19]

5.2 Potential Herb-Drug Interactions

Herbal remedies may interact with prescription medications, affecting their efficacy or safety. Some herbs can interfere with drug metabolism, leading to altered drug concentrations in the body. Additionally, herbs may have additive or synergistic effects with certain medications, potentially increasing the risk of adverse reactions. Healthcare professionals should be aware of potential herb-drug interactions and advise patients accordingly, especially those with underlying health conditions or taking multiple medications.

5.3 Regulation and Quality Control

The regulation and quality control of herbal products pose significant challenges. The herbal supplement industry is often subject to less stringent regulations compared to pharmaceuticals, resulting in variations in product quality, labeling accuracy, and safety. Inconsistent manufacturing practices and adulteration with undeclared substances can further compromise the reliability and safety of herbal products. Implementing rigorous regulations, quality control standards, and labeling requirements are necessary to ensure the safety and efficacy of herbal remedies used in obesity treatment.^[19]

5.4 Combining Herbal Therapies with Conventional Approaches

Integrating herbal therapies with conventional approaches to obesity management is an area of interest and challenge. Determining the optimal combination, dosages, and timing of herbal remedies with dietary modifications, physical activity, and pharmacotherapy requires careful consideration. Additionally, the potential synergistic or additive effects, as well as possible herb-drug interactions, need to be evaluated. Further research is needed to determine the most effective and safe strategies for combining herbal therapies with conventional approaches to maximize treatment outcomes.^[19]

5.5 Emerging Research and Novel Approaches

The field of herbal treatments for obesity is evolving, and there are several areas of research and exploration that hold promise for the future. These include:

- **Rigorous Clinical Trials:** Conducting well-designed, randomized controlled trials with larger sample sizes and longer durations to evaluate the efficacy and safety of specific herbs in obesity treatment. Standardized protocols, validated outcome measures, and appropriate control groups are essential in generating robust evidence.
- **Mechanistic Studies:** Investigating the underlying mechanisms of action of herbs in obesity management. Understanding how herbal compounds interact with metabolic pathways, appetite regulation, and fat metabolism will provide valuable insights for developing targeted therapies.
- **Personalized Approaches:** Exploring the concept of personalized medicine in herbal treatments for obesity. Identifying biomarkers or genetic markers that predict individual responses to specific herbs can help tailor treatment strategies for improved efficacy and patient outcomes.
- **Combination Therapies:** Investigating the potential synergistic effects of combining multiple herbs or herbal extracts to enhance the efficacy of obesity treatment. Evaluating the safety and efficacy of herb combinations may provide new avenues for developing more effective herbal treatments.

- **Long-Term Maintenance:** Focusing on the long-term maintenance of weight loss achieved through herbal treatments. Understanding the factors influencing weight regain and developing strategies to sustain weight loss will be crucial for successful long-term management.

Standardization, regulation, herb-drug interactions, and the integration of herbal therapies with conventional approaches require careful consideration. Future research should focus on conducting rigorous clinical trials, investigating underlying mechanisms, and exploring personalized and combination approaches to optimize the use of herbal treatments in obesity management. By addressing these challenges and advancing our knowledge, herbal therapies can potentially offer safe and effective options in the comprehensive treatment of obesity.^[20]

6. Conclusion

Obesity is a significant global health challenge that requires comprehensive and multidimensional approaches for effective management. While conventional approaches, such as dietary interventions, physical activity, pharmacotherapy, and bariatric surgery, form the foundation of obesity treatment, there is growing interest in the potential of herbal remedies as adjunctive therapies. Herbal remedies offer a unique perspective in addressing obesity due to their bioactive compounds that can influence various metabolic pathways, appetite regulation, and fat metabolism. However, it is important to approach the use of herbal treatments with caution, considering factors such as lack of standardization, potential herb-drug interactions, and the need for proper regulation and quality control. Several herbs have shown promise in scientific studies for their potential anti-obesity effects. Green tea, garcinia cambogia, cinnamon, ginger, turmeric, fenugreek, and other herbs have been studied for their mechanisms of action and clinical evidence in weight management. While the evidence supporting their efficacy is still evolving and inconsistent, they may offer complementary benefits when used in conjunction with conventional approaches. Addressing the challenges and limitations in herbal treatments for obesity is crucial for their safe and effective utilization. Standardization protocols, regulation and quality control measures should be implemented to ensure consistency and reproducibility of herbal products. Healthcare professionals need to be aware of potential herb-drug interactions and advise patients accordingly. Future research should focus on conducting rigorous clinical trials, mechanistic studies, and personalized approaches to further understand the potential of herbal remedies in obesity management. Exploring combination therapies and strategies for long-term weight maintenance will also be essential. While herbal remedies have shown promise as adjunctive therapies in obesity treatment, they are not a standalone solution. They should be integrated into a comprehensive treatment plan that includes dietary modifications, physical activity, and personalized approaches. By addressing the challenges and advancing our knowledge in herbal treatments, we can potentially optimize their role in the holistic management of obesity.¹

References:

1. Y. C. Wang, K. McPherson, T. Marsh, S. L. Gortmaker, and M. Brown, "Health and economic burden of the projected obesity trends in the USA and the UK," *The Lancet*, vol. 378, no. 9793, pp. 815–825, 2011.
2. The GBD 2015 Obesity Collaborators, "Health Effects of Overweight and Obesity in 195 Countries over 25 Years," *New England Journal of Medicine*, vol. 377, no. 1, pp. 13–27, 2017.
3. L. Landsberg, L. J. Aronne, L. J. Beilin et al., "Obesity-related hypertension: Pathogenesis, cardiovascular risk, and treatment—a position paper of the obesity society and the American society of hypertension," *Obesity*, vol. 21, no. 1, pp. 8–24, 2013.
4. P. L. Lefebvre and A. J. Scheen, "Obesity: Causes and new treatments," *Experimental and Clinical Endocrinology and Diabetes*, vol. 109, no. 2, pp. S215–S224, 2001.
5. J. E. Lambert, J. A. Parnell, J. M. Tunnicliffe, J. Han, T. Sturzenegger, and R. A. Reimer, "Consuming yellow pea fiber reduces voluntary energy intake and body fat in overweight/obese adults in a 12-week randomized controlled trial," *Clinical Nutrition*, 2015.
6. S.-L. Chu, H. Fu, J.-X. Yang et al., "A randomized double-blind placebo-controlled study of Pu'er tea extract on the regulation of metabolic syndrome," *Chinese Journal of Integrative Medicine*, vol. 17, no. 7, pp. 492–498, 2011.
7. C. Roongpisuthipong, R. Kantawan, and W. Roongpisuthipong, "Reduction of adipose tissue and body weight: effect of water soluble calcium hydroxyl citrate in *Garcinia atroviridis* on the short term treatment of obese women in Thailand," *Asia Pacific Journal of Clinical Nutrition*, vol. 16, no. 1, pp. 25–29, 2007.
8. R.-R. He, L. Chen, B.-H. Lin, Y. Matsui, X.-S. Yao, and H. Kurihara, "Beneficial effects of oolong tea consumption on diet-induced overweight and obese subjects," *Chinese Journal of Integrative Medicine*, vol. 15, no. 1, pp. 34–41, 2009.
9. G. R. Bhatti, R. Qureshi and S. M. Shah, "Ethno botany of *Calotropis procera* with Especial Reference to the People of Nara Desert," *Scientific Sindh*, Vol. 5, 1998, pp. 13–22.
10. U. Rahman and K. Zaman, "Medicinal Plants with Hypoglycemic Activity," *Journal of Ethno pharmacology*, Vol. 26, No. 1, 1989, pp. 1–55. doi:10.1016/0378-8741(89)90112-8
11. P. C. Sharma, M. B. Yelne, T. J. Dennis, and A. Joshi, "Database on Medicinal Plants Used in Ayurveda," Central Council for Research in Ayurveda and Siddha, India, 2000.
12. V. M. Gogte, "Medicinal Plants, Part III in Ayurvedic Pharmacology and Therapeutic Uses of Medicinal Plants (Dravyagunavignyan)," Bhartiya Vidya Bhavan, Mumbai, 2002.
13. Misra and R. Vaisya, "Bhavaprakasha Nigantu," Chaukamba Sanskrit Sansthan, Varanasi, 2005. [29] J. W. Yun, "Possible Anti-Obesity Therapeutics from Nature—A Review," *Phytochemistry*, Vol. 71, No. 14–15, 2010, pp. 1625–1641.
14. X. Zhang, J. Cui, Q. Meng, S. Li, W. Zhou, and S. Xiao, "Advance in anti-tumor mechanisms of shikonin, alkannin and their derivatives," *Mini-Reviews in Medicinal Chemistry*, vol. 17, no. 18, 2017.
15. X. Chen, L. Yang, J. J. Oppenheim, and O. M. Z. Howard, "Cellular pharmacology studies of shikonin derivatives," *Phytotherapy Research*, vol. 16, no. 3, pp. 199–209, 2002.
16. M. Su, W. Huang, and B. Zhu, "Acetylshikonin from Zicao prevents obesity in rats on a high-fat diet by inhibiting lipid accumulation and inducing lipolysis," *PLoS ONE*, vol. 11, no. 1, Article ID e0146884, 2016.
17. Geller F, Reichwald K, Dempfle A, Illig T, Vollmert C, Herpertz S, et al. Melanocortin-4 receptor gene variant 1103 is negatively associated with obesity. *Am J Hum Genet* 2004. Mar; 74(3):572–81.
18. Heymsfield SB, Wadden TA. Mechanisms, pathophysiology, and management of obesity. *N Engl J Med* 2017; 376(3):254–66.
19. J.-H. Wang, B.-S. Kim, K. Han, and H. Kim, "Ephedra-treated donor-derived gut micro biota transplantation ameliorates high fat diet-induced obesity in rats," *International Journal of Environmental Research and Public Health*, vol. 14, article 555, no. 6, 2017.
20. J. J. Kim, M. E. Tarnoff, and S. A. Shikora, "Surgical Treatment for Extreme Obesity: Evolution of a Rapidly Growing Field," *Nutrition in Clinical Practice*, vol. 18, no. 2, pp. 109–123, 2003.

Space-filling curves: A path tracing towards efficient miniature Antenna

Abhijeet Das

Center for Advanced Research, Department of Physics,
Rajiv Gandhi University, Arunachal Pradesh, India
Email: abhijeet.das@rgu.ac.in

Abstract

Fractals are objects that exhibit different types of self-similarity and fractional dimension. In addition, the fractal curves display space-filling properties, i.e. the curves are adept of passing through all the points in a specific space. This space-filling feature allows for the incorporation of large electrical size feature in antenna design leading to novel miniaturized antenna with optimal properties and performance. This approach offers many advantages in antenna design like compactness, versatility, economic friendliness, enhanced bandwidth, etc. along with desired performance at wide range of frequencies in various fields of applications. The article aimed to provide a mini review to the application of space-filling curves as a possible path for designing miniaturized antenna with optimal properties. Consequently, it can act as an initiation point for introduction to space-filling curve and their application in designing novel fractal antenna with augmented performance.

Introduction

I. Fractals:

Richard Feynman once quoted "I think nature's imagination is so much greater than man's, she's never going to let us relax". Still, after centuries of investigations humans have only recently started to investigate, understand, and appreciate the augmented efficiency in nature's functioning, generally arising from the extremely complex yet intricate detailed structures. Representatives include blood, and pulmonary vessels, DNA, heart beats, proteins, mountains,

A. Das., "Space-filling curves: A path tracing towards efficient miniature antenna", Horizon (Ed Book), 92-102(2023), copyright: Tinsukia College Publishing. ISBN: 978-81-19322-07-7

river networks, etc. This paradigm shift can be attributed to the development of mathematical branch of Fractal geometry. Interestingly, though fractal objects were already dealt from the 19th century, its first comprehensive description and wide applicability was discussed by Benoit Mandelbrot in his celebrated book entitled *The Fractal Geometry of Nature*. He developed fractal geometry to overcome the limitations of Euclidean geometry in description of spatial patterns of rough or fragmented objects, omnipresent in nature where, the limitation is attributed to the involvement of only locally smooth/regular or integer dimensional structures and failing to describe the intrinsic complexities and irregularities at varying spatial and temporal scales. Mandelbrot quoted "Clouds are not spheres, mountains are not cones, and bark is not smooth, nor does lightning travel in a straight line" to highlight the limitations of Euclidean geometry.

Fractal geometry studies objects known as fractals. Mandelbrot derived this term from the Greek word 'fractus', meaning broken or fragmented. In general, objects displaying self-similar characteristics and having Hausdorff-Besicovitch fractal dimension (FD) greater than their topological dimension (TD) are identified as fractals. Interestingly, fractal objects can exhibit self-similar characteristics in one of the following manner, viz.

1. **Exact self-similarity:** The fractal objects display exactly the same structure at every scale irrespective of the scaling. e.g., Geometric fractals – Cantor set, Koch curve, Sierpiński carpet, etc.

2. **Quasi self-similarity:** The objects approximate the complete pattern of the structure at varying scales and contain small copies of the whole in distorted or degenerated forms. e.g., Algebraic fractals – Mandelbrot Set, Julia sets

3. **Statistical self-similarity:** The objects exhibit pattern repetition with a random probability distribution such that statistical properties such as, correlations remain conserved across scales. They are also known as self-affine fractals. e.g., Brownian motion, randomly deposited surfaces, etc.

4. **Multifractals:** These objects display more than one characteristic scaling across scales. e.g., Nebula clusters stock market, etc.

Additionally, a characteristic feature of fractal objects, providing quantitative statistical description of their complexity, is their dimension. Interestingly, the generic term, i.e. fractal dimension is used for different fractals due to the fact that approximately similar value is provided by different computing methods such as, Box-counting and Cube-counting method, Higuchi's algorithm, etc. for a particular fractal. In addition, it had been shown that Euclidean geometry is a special case of fractal geometry. Moreover, the term signifies that fractals possesses a fractional or non-integer dimension suggesting the incompetency of topological dimension in their comprehensive description. For example, the Koch snowflake has a TD of 1 though, the curve's length between any two points is infinite suggesting the snowflake to be composed of infinite number of line segments linked at diverse angles. This in turn signifies that the snowflake is excessively detailed to be one-dimensional, but too simple

to consider as two-dimensional, i.e. it must have a dimension between 1 and 2. Captivatingly, the FD of Koch snowflake is found to approximately 1.262. Nonetheless, a comprehensive discussion of fractals and its quantitative descriptors is beyond the scope of this article. However, interested readers can find overarching details in the literatures [1-3].

Interestingly, the above mentioned indicates that FD of an object reflects its space filling properties, i.e. curves possessing FD of 1.9 and 2.9 (say) will convolute through the space nearly like a surface and volume, respectively. However, they are not strictly correlated where, space filling curves like the Hilbert curve does not follow the conventional definition of fractals as outlined above. In mathematical sense, the Hilbert curve is not homeomorphic implying that it does conserve its topological dimension, i.e. the curve of Hilbert map in $\mathbf{R}^3$ has TD of 1 but both TD and FD of image of Hilbert map in $\mathbf{R}^2$ display value of 2.

II. Space-Filling Curves

The path traced continuously by a moving point in two, three, or higher dimension is known as a curve. In a stricter mathematical sense, a continuous function with domain in the unit interval $[0, 1]$ is a curve with endpoints and a curve whose range reaches each point in a higher dimensional region, i.e. from a unit square to a n -dimensional unit hypercube is called as space-filling curve. Generally, a curve adept of passing through all the points in a specific space consequently, filling that space is known as a space-filling curve. The initiation of the concept trace back to George Cantor (1878) who aimed to show a subjective mapping from $\mathbf{R}$ to $\mathbf{R}^2$ that include all the points in a plane. The first demonstration of such curve was reported by *Guisepe Peano* in 1890. Here, each sub-square is subdivided into nine congruent sub-squares, each with a side length to $1/3$ of the previous one. A recursive subdivision is then performed in each step thereby, mapping individual interval to a specific sub-square of the identical recursion level. Hence, individual sub-square will contain a part of the Peano curve which is suitably scaled and transformed copy of the complete curve, realized from Fig. 1(a) and 1(b). Interestingly, only vertical or horizontal reflection is required in construction of the Peano curve, shown in Fig. 1(c) and 1(d), respectively with no rotation [4].

It is noteworthy that fractals and space-filling curves share close relation in the sense that the regularities in these curves are greatly linked to the self-similar property of fractals. Nonetheless, motivated by *Peano*, other curves with space-filling property were proposed by *Hilbert*, *Moore*, *Sierpiński-Knopp*, *Koch*, etc. A comprehensive overview of these curves in mathematical and additional senses is not the aim of this article. Nevertheless, a brief description to them is provided in the following paragraph.

IIa. Hilbert Curve

It is a generalization of the specific sequential ordering of grid cells in a quad tree connecting only the neighboring grid cells. The iterations here connect the mid-points of the recursively subdivided sub-squares as a polygon. In each iteration to next, the existing sub-squares are subdivided into four sub-squares and are connected via the reflection or rotation of the fundamental pattern or initiator (Fig. 2a). In addition, in moving from one to next iteration, four copies of the existing iteration are connected using reflection or rotation such that their start or end points meet. It is noteworthy that the orientation of the four iterations remains identical irrespective of the resolution of the iteration. The first three, i.e. fundamental pattern, generator, and 3rd iterations of the Hilbert curve is shown in Fig. 2(a-c), respectively [4].

IIb. Moore Curve

The union of four copies of the Hilbert curve such that the endpoints coincide forms the Moore curve. Conversely, it is a variant of the Hilbert curve. In general, the N^{th} order Moore curve in K dimension can be formed by placing 2^K copies of $N-1^{\text{th}}$ order K -dimensional Hilbert curve at individual corner of the hypercube, rotating, and connecting them via line segments, shown in Fig. 3. Thus, a 3-dimensional Moore curve, 2^3 copies of 3-dimensional Hilbert curve should be placed at the corners of a cube, rotating, and connecting them by line segments.

IIc. Sierpiński-Knopp Curve

The Peano, Hilbert, and Moore curve, respectively are based on recursive sub-structuring of the fundamental pattern into squares. Interestingly, a novel space-filling pattern via sub-structuring into triangles was proposed by *Waclaw Sierpiński* in 1912. *Konrad Knopp* provided the mathematical description of the resultant curve consequently, the space-filling curve is known as *Sierpiński-Knopp curve*. Here, the isosceles triangle is split into two sub-triangles. Thereafter, via splitting the hypotenuse the individual triangle is divided into two parts consequently, the legs of the fundamental triangle make the hypotenuse of the splitted sub-triangles. Fig. 4 displays the first few iterations via nested intervals and the corresponding sequentialisations.

IId. Koch Curve

The Koch curve (Koch snowflake, Koch star, or Koch island) was reported by *Helge von Koch* in 1904. Here, the limiting figure, i.e. Koch curve converges to ~ 1.6 times the area of the original triangle while the perimeter diverges to infinity. As a result, the figure has a finite area bounded by an infinitely long line. The construction of the Koch curve is given below;

1. Draw an equilateral triangle with side length divisible by 3.
2. Divide each side into three equal parts.

3. Draw an equilateral triangle from the middle part and measure the length of the middle third to compute the side lengths of the new triangle.

4. Divide each outer side into $1/3^{\text{rd}}$. Here, the second generation of the triangles covers parts of the first generation triangles.

5. Repeat the steps 3-4 for successive formation of new generations, e.g. 3^{rd} generation of triangles, shown in Fig. 5. The limiting figure is known as the Koch curve.

Space-filling curves have found considerable application in low-dimensional problems such as the travelling salesman problem, storage and retrieval of multi-dimensional data, organize and process raster data, antenna design, image compression, bandwidth reduction etc. Nevertheless, the aim of this article is to give a brief overview on the applicability of space-filling curves in antenna design.

III. Fractal Antenna

The modern communication system follows the trend of 4A (anyone-anything-anytime-anywhere). The communication system allows the operation of devices like Bluetooth devices, laptops, smart phones, computer peripherals, etc. Notably, antenna is significant for the transmission and reception of electromagnetic signals in these devices. Also, mobility and portability are indispensable requirements in wireless communication systems where, portability can be achieved via miniaturization of the system's component modules. Particularly, miniaturized antenna is significantly demanding for achieving portability and mobility in wireless communication systems. However, major challenges in design of antenna such as small size, economic friendliness, light weight, easy fabrication, radiation properties, and enhanced bandwidth limits the antenna's efficiency. Fascinatingly, most of these limitations can be overcome with significant size miniaturization, bandwidth, gain and directivity, multi-frequency performance, and economic friendliness along with low mutual coupling via utilizing fractal geometry in designing of antenna.

The miniaturization of antenna size has been observed to be dependent on the correlation between antenna's physical size and its corresponding operating wavelength. At constant operating wavelength, decrease in antenna size results in similar change in the electrical size. In addition, deterioration of input resistance, efficiency, radiation property, impedance bandwidth, etc. occurs as the physical size of antenna gets decreased with respect to the corresponding operating wavelength. In this regard, fractal geometrical structures becomes supportive for size miniaturization without consumption of additional volumes. Notably, the space-filling feature of fractal objects allows for the incorporation of large electrical size

fractal antenna. The first fractal antenna was reported by Nathan Cohen in 1988 from utilizing fractal geometry to loop or dipole antenna. The concept endorses decrease in antenna structure with continuous iterations while conserving the total length of the antenna [5]. Major advantages offered by fractal antenna includes i) compactness, ii) versatility, iii) economic

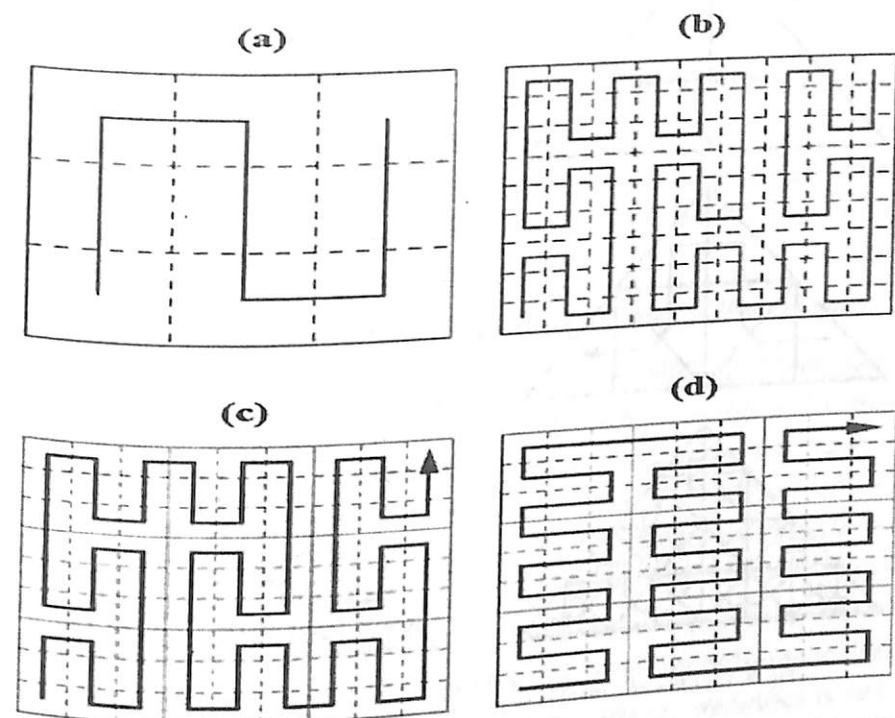


Fig. 1 Illustration of Peano curve with (a) one iteration, (b) two iterations, (c) vertical reflection, and (d) horizontal reflection [4]

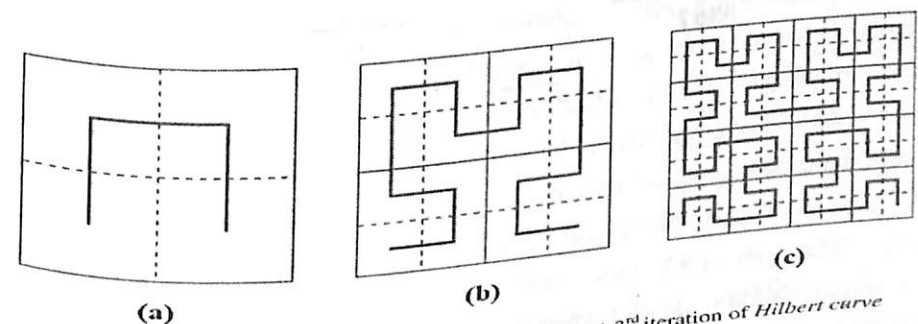


Fig. 2 Illustration of (a) initiator, (b) generator, and (c) 3^{rd} iteration of Hilbert curve

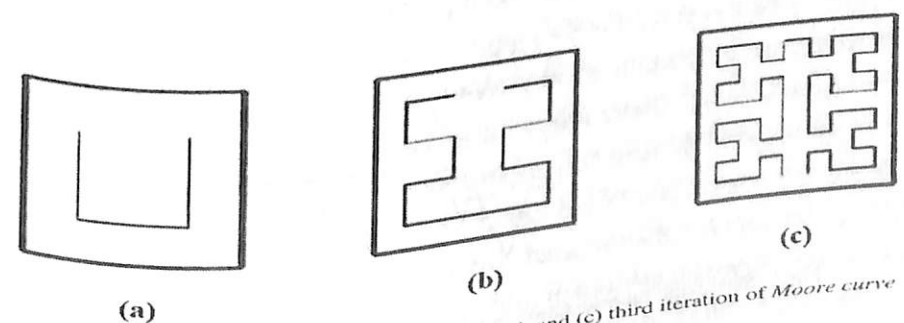


Fig. 3 Pictorial representation of (a) first, (b) second, and (c) third iteration of Moore curve

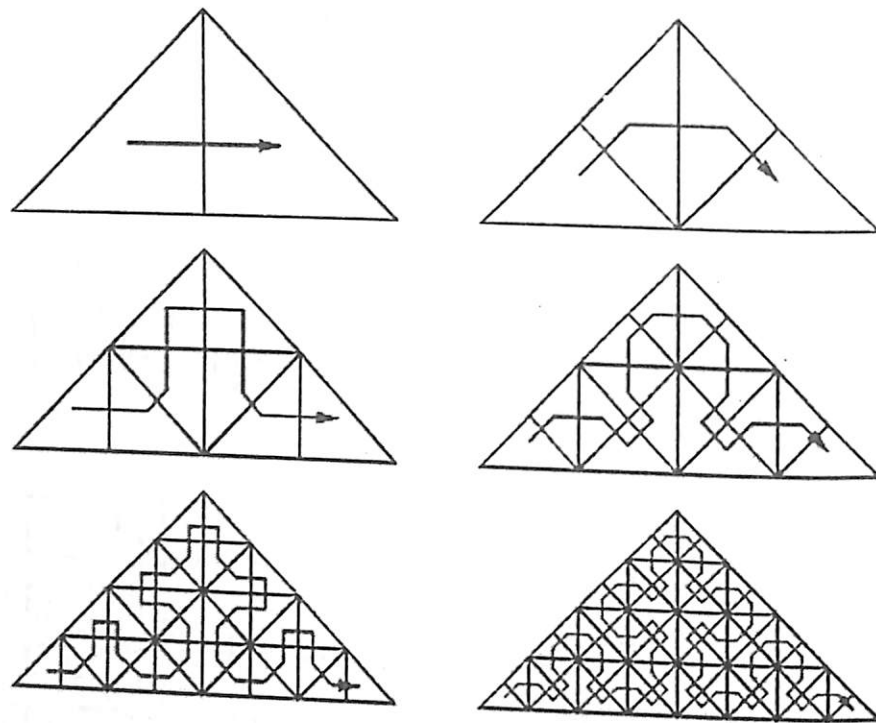


Fig. 4. First few construction steps of the *Sierpiński-Knopp curve* [4]

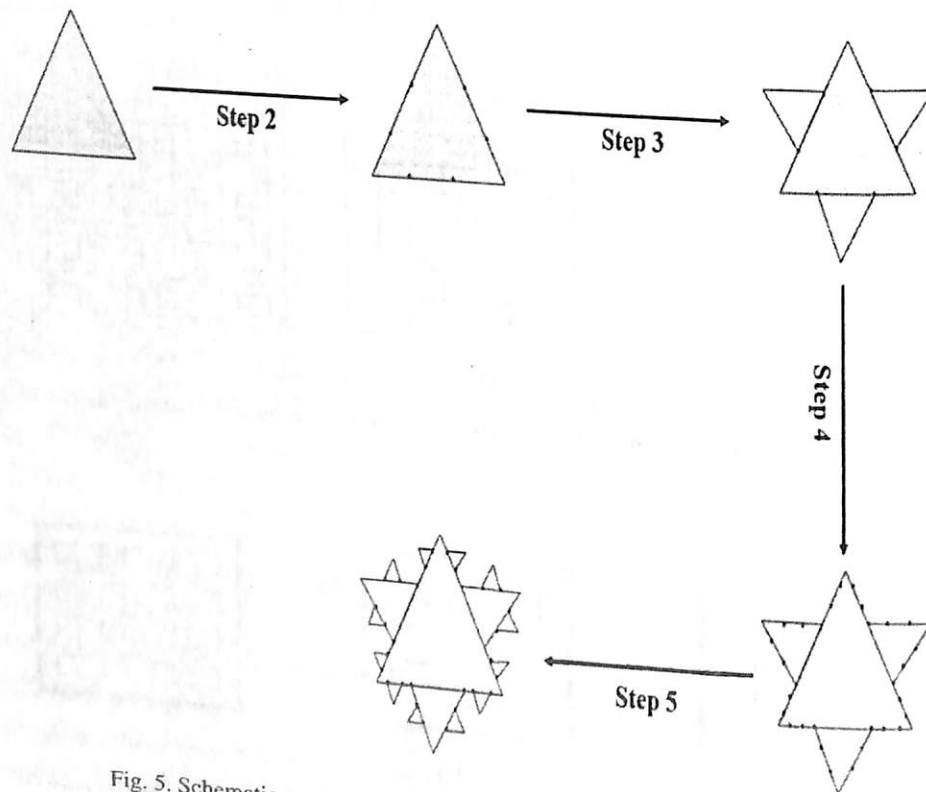


Fig. 5. Schematic representation of first few steps in formation of *Koch curve*

friendliness, iv) enhanced bandwidth, v) higher gain and directivity, vi) improved input impedance matching, vii) low mutual coupling, viii) broadband or multiband properties, ix) frequency independent properties and unbiased efficiency. Nonetheless, like any other technology, it also possess limitation like, i) complexity in design, ii) gain reduction, and iii) less advantage after few iterations. Fractal antenna displayed significant applicability in the field of wide-band antenna, cellular mobile communications, satellite communications, military applications, biomedical applications, wireless network (WLAN/WiMAX, MIMO), aviation, telemetric application, radio astronomy and space research, etc.

Design of miniature fractal antenna: Application of space-filling curves

In recent times, a lot of effort has been put for demonstrating enhanced efficiency of fractal objects with higher FD at low resonating frequencies as compared to object with any other geometry [6]. Conversely, other researchers reported similar performance of antenna with many geometries including fractal geometry for the identical wire diameter, total wire length, and occupied area [7, 8]. Space-filling curves are potential candidates for realization of miniature antenna owing to their capability of arranging long wires into small areas. The Koch curve was reported as the first example of fractal curve improving antenna characteristics in a monopole configuration [9]. The work was extended to dipole configuration utilizing two- and three-dimensional fractal trees demonstrating space-filling geometries as efficient curves in miniaturizing antenna [8]. The two-dimensional geometries were concluded to be optimal monopoles (does not require baluns) due to the easy fabrication process from conventional printed circuits manufacturing techniques. Sharma *et al.* [10] reported a novel structure of hybrid antenna using a combination of Peano-Gosper, Koch, and Minkowski curves with a modified partial ground plane with $50\ \Omega$ transmission line fed for multi-standard wireless applications. Iterated function system was implemented on to a segment of Peano-Gosper curve to obtain the hybrid antenna structure. The proposed structure displayed seven distinct frequency points with a maximum bandwidth of 5 GHz, reflection coefficient of -45.65 dB at 1.5 GHz, positive gain, and stable patterns of radiation at operational frequency bands. Space-filling, self-avoiding, and self-similar Peano curves was used to study metallic microstrip antenna with air as substrate in the ultra-high frequency band. Genetic programming was employed to find the best equation for fitting the radiation properties, like resonance frequencies and central frequencies; such that voltage standing wave ratio (VSWR) is less than 2 as a percentage of the central frequency of Peano curves. They observed the number of resonance frequency to increase with number of iterations [11]. A self-similar stair-cased fractal structure was proposed for multiband applications. Here, a tri-band response originated from the stair-cased fractal with resonance at frequencies of 3.65, 4.825, and 6.235 GHz. Antenna parameters like return loss, impedance bandwidth, and far-field characteristics at various operating bands endorses the applicability of the proposed structure in multiband applications like Wi-Max (3.5 GHz), C-band applications (4.7 GHz), and point-to-point high-speed wireless

communications (6.3 GHz) [12]. Sharma and Bhatia [13] performed a comparative analysis between different hybrid fractal antenna (Koch-Minkowski, Moore-Koch, Koch-Meander, Minkowski-Sierpiński, Meander-Minkowski, Minkowski-Hilbert, etc.) on the basis of their size, designing geometries, gain, bandwidth, number of resonant frequency bands, circular polarization, radiation efficiency, and ultimately their possible applications. In addition, they also highlighted the challenges in design of hybrid fractal antenna. Kumar and Phawaha [14] developed a modified Hilbert curve fractal antenna with parasitic patches for multiband applications. Swarm intelligence-based Dragonfly Optimization technique was employed to optimize the geometry for enhanced bandwidth. The resulting antenna displayed an augmented degree of impedance matching at all resonating frequencies with stable omnidirectional radiation pattern. They speculate the usefulness of the proposed antenna for radiolocation in ultra-high frequency band and aeronautical radio-navigation in C-band, K-band, and Ku band according to International Telecommunication Union standards. Multi-Input Multi-Output (MIMO) has emerged to be one of the significant feature for future wireless communication thanks to improvement in throughput, bandwidth, bit rate, error rate. It is a standard specified in IEEE 802.11n Wi-Fi, Wi-MAX (IEEE 802.16e), LTE (Long Term Evolution), 3GPP (3rd Generation Partnership Project), HSPA+ (High Speed Packet Access), 4G, and 5G systems. In this consideration, Pujaret *et al.* [15] overviewed the fractal microstrip MIMO antenna for wireless communication. The authors suggested introduction of parasitic element either in coplanar or stack configuration, increase of substrate thickness, patch shape modification, and slot insertion for fabrication of novel wideband microstrip fractal antenna. A complementary Sierpiński-Knopffractal antenna realized on the basic square monopole antenna with FR4 dielectric material as substrate was designed and analyzed by Iyampalam and Ganesan [16]. ANSYS electronic desktop software package was utilized to investigate the antenna metrics such as, reflection coefficient, VSWR, radiation pattern, gain and current distribution at resonant frequencies. The proposed space-filling facilitated antenna resonate at three frequencies, i.e. 2.08, 4.93, and 6.46 GHz with reflection coefficient of -20.5, -23.1, and -24.3 dB, respectively. Additionally, the antenna covers frequency bands of mobile satellite service, Public Protection and Disaster Relief communication, and INSAT-C band application. Chen *et al.* [17] reported a broadband microstrip-fed antenna realized through merging of different downscaled version of composite island-like second iteration space-filling Moore curves. The island-like curves facilitates provision of different resonant slot lengths over a wide frequency range and loss bandwidth of 10 dB in the range of 2.5-17.2 GHz. Also, the simulation results are well corroborated with experimental outcomes. In other works [18-20], concepts from fractal geometry and electromagnetic theory are combined, giving rise to the field of *fractal electrodynamics*, for investigating unique classes of radiation, propagation, scattering problems and their application to antenna theory and design. The aforementioned clearly demonstrates the significance of fractal structures in general and space-filling curves in particular in design and fabrication of miniaturized antenna with optimal properties for efficient and enhanced performance in a wide range of frequencies for desired applications.

Conclusion

This mini review aimed to demonstrate the applicability of space-filling curves like the Peano curve, Hilbert curve, Moore curve, Koch curve, Sierpiński curve, etc. in design and fabrication of miniaturized fractal antenna with efficient performance for desired applications. Space-filling curves share close relation to fractals in the sense that the regularities in these curves are greatly linked to the self-similar property of fractals. A short description to the above mentioned curves are provided. In addition, fractal antenna are introduced along with their advantages and limitations. Finally, some of the works reporting utilization of space-filling curves in various combinations for designing and fabricating of fractal antenna along with analyzing their properties at difference resonant operating frequency for preferred applications are reviewed. The article can act as an initiation point for introduction to applicability of space-filling curve in designing novel fractal antenna with augmented performance.

References

1. Benoit B. Mandelbrot, *Fractals: Form, Chance, and Dimension*, W. H. Freeman, San Francisco, 1977.
2. Benoit B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman, New York, 1983.
3. S. Banerjee, M. K. Hassan, S. Mukherjee, A. Gowrisankar, *Fractal Patterns in Nonlinear Dynamics and Applications*, CRC Press, Boca Raton, 2019. <https://doi.org/10.1201/9781315151564>
4. M. Bader, *Space-Filling Curves: An Introduction with Applications in Scientific Computing*, Springer-Verlag Berlin, Heidelberg, 2013. <https://doi.org/10.1007/978-3-642-31046-1>
5. H. Sagan, *Space-Filling Curves*, Springer, New York, 2012.
6. J. Anguera, C. Puente, E. Martinez, E. Rozan, The fractal Hilbert monopole: A two-dimensional wire, *Microwave Opt. Technology Lett.*, **36**, 102-104, 2002. <https://doi.org/10.1002/mop.10687>
7. S. R. Best, Comments on Hilbert curve fractal antenna: A small resonant antenna for VHF/UHF applications, *Microwave Opt. Technology Lett.*, **35**, 420-421, 2002. <https://doi.org/10.1002/mop.10625>
8. S. R. Best, A comparison of the performance properties of the Hilbert curve fractal and Meander line monopole antennas, *Microwave Opt. Technology Lett.*, **35**, 258-262, 2002. <https://doi.org/10.1002/mop.10576>
9. C. P. Baliarda, J. Romeu, and A. Cardama, The Koch monopole: a small fractal antenna, *IEEE Trans. Antennas Propagat.*, **48**, 1773-1781, 2000. <https://doi.org/10.1109/8.900236>
10. N. Sharma, S. S. Bhatia, V. Sharma, Peano-Gosper, Koch and Minkowski fractal curves-based novel hybrid antenna using modified partial ground plane for multi-standard wireless applications, *J. Electromagn. Waves Appl.*, **35**, 1959-1979, 2021. <https://doi.org/10.1080/09205071.2021.1925974>

11. R. P. Moroyoqui, S. Rodriguez-Romo, O. Ibáñez Orozco, Patterns of the radiation properties for Peano antennas. *Sci Rep.* **13**, 8358, 2023. <https://doi.org/10.1038/s41598-023-35185-6>
12. Z-ul. A. Jaffri, Z. Ahmad, A. Kabir, S. Sabahat, H. Bukhari, A novel compact stair-shaped multiband fractal antenna for wireless communication systems. *J. Electr. Eng.*, **72**, 306-314, 2021. <https://doi.org/10.2478/jee-2021-0043>
13. N. Sharma, S. S. Bhatia, Comparative analysis of hybrid fractal antennas: A review, *Int. J. RF Microw. Comput.-Aided Eng.*, **31**, e22762, 2021. <https://doi.org/10.1002/mmce.22762>
14. A. Kumar, A. P. S. Pharwaha, Development of a modified Hilbert curve fractal antenna for multiband applications, *IETE J. Res.*, **68**, 3597-3606, 2022. <https://doi.org/10.1080/03772063.2020.1772126>
15. N. T. Pujar, Jagadeesh S., V. S. Malagitti, Overview of fractal microstrip MIMO antenna for wireless communication. *Int. J. Sci. Eng. Res.*, **10**, 744-748, 2019.
16. P. Iyampalam, I. Ganesan, Complementary SierpinskiKnopp fractal antenna for emergency management system, *Int. J. Microw. Wirel. Technol.*, **12**, 1029-1038, 2020. <https://doi.org/10.1017/S1759078720000343>
17. Y. Chen, S. Liu, Z. Dai, Y. Wei, Broad-Band space-filling-based slot antenna, 291-293, 2010. <https://doi.org/10.1109/ICMMT.2010.5524936>
18. D. H. Werner, An overview of fractal electrodynamics research, *Proceedings of the 11th Annual Review of Progress in Applied Computational Electromagnetics (ACES) Vol. II*, Naval Postgraduate School, Monterey, CA, 964-969, 1995.
19. D. L. Laggard, Fractal electrodynamics: From super antennas to superlattices, *Fractals in Engineering*, Springer-Verlag, New York, 204-221, 1997.
20. D. H. Werner, R. Mittra (eds.), *Frontiers in Electromagnetics*, Piscataway, IEEE Press, NJ, 2000.

Galaxy Distribution Observations Predict a Smaller Hubble Constant

Dhruba Jyoti Gogoi⁽¹⁾
Jyatsnasree Bora⁽²⁾

^(1,2)Department of Physics
Dibrugarh University, Dibrugarh, India
Email: moloydhruba@gmail.com

Abstract

This paper introduces refined and accurate estimations of the Hubble constant through the integration of two primary datasets derived from observations of galaxy distribution. The datasets encompass 40 measurements of cluster galaxy gas mass fractions (GMF) and 11 measurements of baryon acoustic oscillation (BAO). By incorporating priors from Big Bang Nucleosynthesis and cosmic microwave background data, the analysis is conducted within non-flat Λ CDM cosmological model. The findings favour lower values for the Hubble constant and provide strong evidence for the accelerated expansion of the universe. This study underscores the effectiveness of combining independent datasets to constrain the Hubble constant and enhances our comprehension of cosmological parameters.

Introduction

This work presents innovative and accurate evaluations of the Hubble constant by synergistically analyzing two primary datasets derived from observations of galaxy distribution in redshifts. These datasets comprise 40 measurements of cluster galaxy gas mass fractions (GMF) and 11 measurements of baryon acoustic oscillation (BAO). To enhance the precision of the estimates, relevant priors from Big Bang Nucleosynthesis (BBN) are incorporated, enabling the calibration of the cosmic sound horizon and the local temperature of the cosmic microwave background using FIRES/COBE data.

The assessments of the Hubble constant are conducted within the non-flat Λ CDM model, which is motivated by recent discussions in the literature regarding potential tensions

D.J. Gogoi, J. Bora, "Galaxy distribution observations predict a smaller Hubble constant", *Horizon (Ed Book)*, 103-111 (2023), copyright: Tinsukia College Publishing, ISBN: 978-81-19322-07-7

in cosmological curvature [1-4]. Notably, analysis of the Planck cosmic microwave background spectra suggests a preference for positive curvature with a confidence level exceeding 99%.

The results highlight the compelling and effective nature of combining these two independent datasets to constrain the Hubble constant following Ref. [13]. Across all models examined in Ref. [13], remarkably precise estimates are obtained, aligning with conclusions drawn from Planck data, which indicate lower values for the Hubble constant. Additionally, analyses within the cosmological model provide strong evidence for an accelerating expansion of the universe at a confidence level exceeding 3σ .

This study provides valuable insights into the Hubble constant and supports the prevailing understanding of a universe undergoing accelerated expansion in accordance with modern cosmological knowledge where our primary focus will be on the non-flat Λ CDM model.

The Cosmological Model

In our quest to estimate the Hubble constant, we venture into the captivating realm of cosmological models, focusing specifically on the non-flat Λ CDM model. This model serves as a framework to unravel the dynamic intricacies of the cosmos, where two dominant forces hold sway: a perplexing form of matter known as cold dark matter (CDM) and a cosmological constant (Λ). The enigmatic CDM component, which eludes electromagnetic interaction, plays a substantial role in shaping the universe's mass distribution. On the other hand, the cosmological constant Λ is intimately associated with the concept of vacuum energy density, characterized by its unchanging value and negative pressure.

To derive the Hubble parameter within the Λ CDM framework, we venture into the profound domain of Einstein's field equations, providing the mathematical bedrock for comprehending the curvature of space-time and its intricate link to the distribution of matter and energy across the cosmos. By introducing certain assumptions, such as a constant equation of state for dark energy (symbolized by Λ) and considering a Universe governed by a homogeneous and isotropic *Friedmann-Lemaître-Robertson-Walker* geometry, we can unravel a precise expression for the Hubble parameter. This parameter encapsulates the pace at which the universe expands, offering a fundamental metric for comprehending the cosmic dynamics and evolution.

In essence, the exploration of the non-flat Λ CDM model allows us to delve into the profound interplay among dark matter, dark energy, and the fundamental forces shaping our celestial realm. Through scrupulous mathematical analysis and rigorous calculations, our aim is to unravel the enigmas surrounding the Hubble constant, thereby gaining deeper insights into the true nature of our vast and awe-inspiring universe.

The Hubble parameter in this Λ CDM framework can be expressed as:

Here H_0 is the Hubble constant and $h \equiv H_0 / (100 \text{ km s}^{-1} \text{ Mpc}^{-1})$, $\Omega_{m,0}$, Ω_Λ and $\Omega_{k,0}$ are

the current dimensionless parameter of matter density (baryons + dark matter), dark energy density and curvature density ($\Omega_{k,0} \equiv 1 - \Omega_{m,0} - \Omega_\Lambda$), respectively.

BAO Data

This sample is composed of 11 measures obtained by 7 different surveys. In this article, we skip the detailed discussion related to it and we only mention a few important points. The relevant physical quantities for the BAO data are the angular diameter distance expressed as [13]:

$$D_A(z) = \frac{1}{(1+z)} \times \begin{cases} \frac{H_0^{-1}}{\sqrt{|1-\Omega_{k,0}|}} \sin\left(\frac{\sqrt{|1-\Omega_{k,0}|}}{H_0^{-1}} \int_0^z \frac{dz'}{H(z')}\right) & \text{if } \Omega_{k,0} < 0 \\ \int_0^z \frac{dz'}{H(z')} & \text{if } \Omega_{k,0} = 0, \\ \frac{H_0^{-1}}{\sqrt{|1-\Omega_{k,0}|}} \sinh\left(\frac{\sqrt{|1-\Omega_{k,0}|}}{H_0^{-1}} \int_0^z \frac{dz'}{H(z')}\right) & \text{if } \Omega_{k,0} > 0 \end{cases}$$

We only choose the third case for a non-flat Λ CDM model. The spherically-averaged distance is expressed as,

$$D_V(z) = [(1+z)^2 D_A^2(z) \frac{z}{H(z)}]^{1/3}$$

and the sound horizon at the drag epoch has the form [4]:

$$r_s(z_d) = \frac{2}{3k_{eq}} \sqrt{\frac{6}{R(z_{eq})}} \ln \left[\frac{\sqrt{1+R(z_d)} + \sqrt{R(z_d) + R(z_d)}}{1 + \sqrt{R(z_{eq})}} \right],$$

where z_d is the drag epoch redshift, z_{eq} is the equality redshift, k_{eq} is the scale of the particle horizon at the equality epoch. The ratio of the baryon to photon momentum density can be written as [13]

$$R(z) \equiv \frac{3\rho_b}{4\rho_\gamma} = 31.5 (\Omega_b h^2) \left(\frac{T_{CMB}}{2.7 \text{ K}} \right)^{-4} \left(\frac{z}{10^3} \right)^{-1}$$

Where the symbols have usual meaning. For survey set I (6dFGS [5] and SDSS-LRG [6]), the BAO quantity is given by:

$$d_z(z) = \frac{r_s(z_d)}{D_V(z)},$$

One can define the Chi-square for this case as

$$\chi^2_{\text{BAO,I}} = \sum_{i=1}^2 \left[\frac{d_z^{\text{th}}(z_i) - d_{z,i}^{\text{ob}}}{\sigma_{d_{z,i}^{\text{ob}}}} \right]^2$$

On the other hand, the BAO quantity for survey set II (BOSS-MGS [7], BOSS-LOWZ [8], BOSS-CMASS [8], BOSS-DR12 [9] and Wiggle Z [10]) is given by:

$$\mathcal{D}(z) = \frac{D_V(z)}{r_s(z_d)} r_s^{\text{fid}},$$

and, in this specific case, the chi-square function is:

$$\chi^2_{\text{BAO,II}} = \sum_{i=1}^6 \left[\frac{\mathcal{D}^{\text{th}}(z_i) - \mathcal{D}_i^{\text{ob}}}{\sigma_{\mathcal{D}_i^{\text{ob}}}} \right]^2 + [\vec{\mathcal{D}}^{\text{th}} - \vec{\mathcal{D}}^{\text{ob}}]^T \mathbf{C}_{\text{WiggleZ}}^{-1} [\vec{\mathcal{D}}^{\text{th}} - \vec{\mathcal{D}}^{\text{ob}}],$$

$\mathbf{C}_{\text{WiggleZ}}^{-1}$ is the inverse covariance matrix, whose explicit form is [10, 13]

$$10^{-4} \begin{pmatrix} 2.17898878 & -1.11633321 & 0.46982851 \\ -1.11633321 & 1.70712004 & -0.71847155 \\ 0.46982851 & -0.71847155 & 1.65283175 \end{pmatrix}$$

We use this data for further analysis of our cosmological model.

Galaxy Cluster Gas Mass Fractions Data

The gas mass fraction quantity for a cluster is given by [11, 12]:

$$f_{\text{gas}}^{\text{X-ray}}(z) = A(z) K(z) \gamma(z) \frac{\Omega_b(z)}{\Omega_m(z)} \left[\frac{D_A^{\text{fid}}(z)}{D_A(z)} \right]^{\frac{3}{2}},$$

Where

$$A(z) = \left[\frac{H(z) D_A(z)}{H^{\text{fid}}(z) D_A^{\text{fid}}(z)} \right]^{\eta}$$

Therefore, for this sample, the chi-squared function is given by,

$$\chi^2_{\text{GMF}} = \sum_{i=1}^{40} \frac{[f_{\text{gas}}^{\text{th}}(z_i) - f_{\text{gas},i}^{\text{ob}}]^2}{\sigma_{\text{tot},i}^2},$$

with a total uncertainty given by

$$\sigma_{\text{tot},i}^2 = \sigma_{f_{\text{gas},i}^{\text{ob}}}^2 + [f_{\text{gas}}^{\text{th}}(z_i)]^2 \left\{ \left(\frac{\sigma_K}{K} \right)^2 + \left(\frac{\sigma_\gamma}{\gamma} \right)^2 + \ln^2 \left[\frac{H(z_i) D_A(z_i)}{H^{\text{fid}}(z_i) D_A^{\text{fid}}(z_i)} \right] \sigma_\eta^2 \right\},$$

We use these two data sets and obtain a relatively narrow constraint on the cosmological parameters and Hubble constant. For this purpose we use Markov Chain Monte Carlo (MCMC) method along with these two data sets.

Results and Concluding Remarks

The statistical analysis is performed by the construction of the chi-squared function,

$$\chi^2_{\text{tot}}(\{\alpha\}) = \chi^2_{\text{BAO,I}} + \chi^2_{\text{BAO,II}} + \chi^2_{\text{GMF}} + \chi^2_{\text{BBN}}.$$

The non-flat Λ CDM model reveals intriguing insights into the cosmic landscape through the examination of parameters such as $\Omega_{m,0}$, Ω_Λ , and h . These parameters play a pivotal role in shaping our understanding of the universe's composition, expansion rate, and curvature. By analyzing the joint dataset encompassing BAO and GMF, we gain a comprehensive perspective on the interplay between these parameters.

Visualizing the results, we observe the contours and likelihoods that map the regions of highest probability in the parameter space. The contours, represented by solid blue lines in Fig. 01, trace the boundaries within which these parameters are constrained. Specifically, the joint analysis of BAO and GMF allows us to explore the intricate relationship between $\Omega_{m,0}$, Ω_Λ and h .

To obtain a comprehensive understanding, we employ a process of marginalization over the Ω_Λ parameter, enabling us to focus on the $\Omega_{m,0}$ - h plane. This plane, featuring two free parameters, provides valuable insights into the range of possible values for $\Omega_{m,0}$ and h . By scrutinizing this plane, we can discern the intervals within which these parameters lie, shedding light on the dynamic characteristics of the universe and its expansion.

For this analysis, by marginalizing over the parameter Ω_Λ , we obtain from the $\Omega_{m,0}$ - h plane (with two free parameters), the intervals obtained are:

$$h = 0.644^{+0.035+0.057}_{-0.034-0.056}$$

$$\Omega_{m,0} = 0.305^{+0.024+0.042}_{-0.022-0.034}$$

In a quest to unravel the intricacies of the non-flat Λ CDM model, we delve further into the parameter space to gain deeper insights into the cosmic fabric. By employing a process of

marginalization over the parameter $\Omega_{m,0}$, we obtain from the Ω_{Λ} - h plane (with two free parameters) the constrained values:

$$h = 0.645^{+0.034+0.057}_{-0.034-0.056}$$

$$\Omega_{\Lambda} = 0.660^{+0.146+0.227}_{-0.179-0.312}$$

Within this plane, we unravel the values that emerge from the analysis, shedding light on the interplay between Ω_{Λ} and h . These values serve as crucial benchmarks, providing us with a deeper understanding of the cosmic landscape and its fundamental properties.

Expanding our exploration, we turn our attention to the $\Omega_{m,0}$ - Ω_{Λ} plane, once again featuring two free parameters. By employing the process of marginalization, this time over the parameter h , we venture into this plane to unravel further insights into the intricate interplay between $\Omega_{m,0}$ and Ω_{Λ} .

$$\Omega_{m,0} = 0.305^{+0.023+0.040}_{-0.022-0.035}$$

$$\Omega_{\Lambda} = 0.661^{+0.152+0.232}_{-0.170-0.301}$$

Within this plane, we uncover intervals that capture the range of values these parameters can assume. These intervals serve as critical guideposts, delineating the boundaries within which $\Omega_{m,0}$ and Ω_{Λ} are constrained and providing valuable clues about the composition, expansion, and curvature of the universe.

On the other hand, in order to obtain the likelihood function for the h parameter, we marginalize over $\Omega_{m,0}$ and Ω_{Λ} parameters. The likelihood function is shown in Fig. 02. From this likelihood, we obtain $h = 0.643^{+0.023+0.045}_{-0.022-0.044}$ in 1σ and 2σ confidence levels.

In summary, through the process of marginalization in the $\Omega_{m,0}$ - h , Ω_{Λ} - h and $\Omega_{m,0}$ - Ω_{Λ} planes, we unearth a wealth of information about the intricate relationship between these fundamental parameters. These findings deepen our understanding of the cosmic tapestry, illuminating its composition, expansion rate, and curvature, and propelling us further on the quest for cosmic knowledge. This investigation provides $H_0 = 64.3^{+4.5}_{-4.4} \text{ km s}^{-1} \text{ Mpc}^{-1}$ in 2σ confidence level. It suggests that within the considered data sets, the Hubble constant exhibits a smaller value in comparison to Planck's results.

Acknowledgement

DJG acknowledges the contribution and support of the European Cooperation in Science and Technology (COST) Action CA21136 - "Addressing observational tensions in cosmology with systematics and fundamental physics (*CosmoVerse*)".

Data Availability and Source Code

Data and Python source code used in the study can be provided upon request for legitimate purposes, with proper citations/mentions.

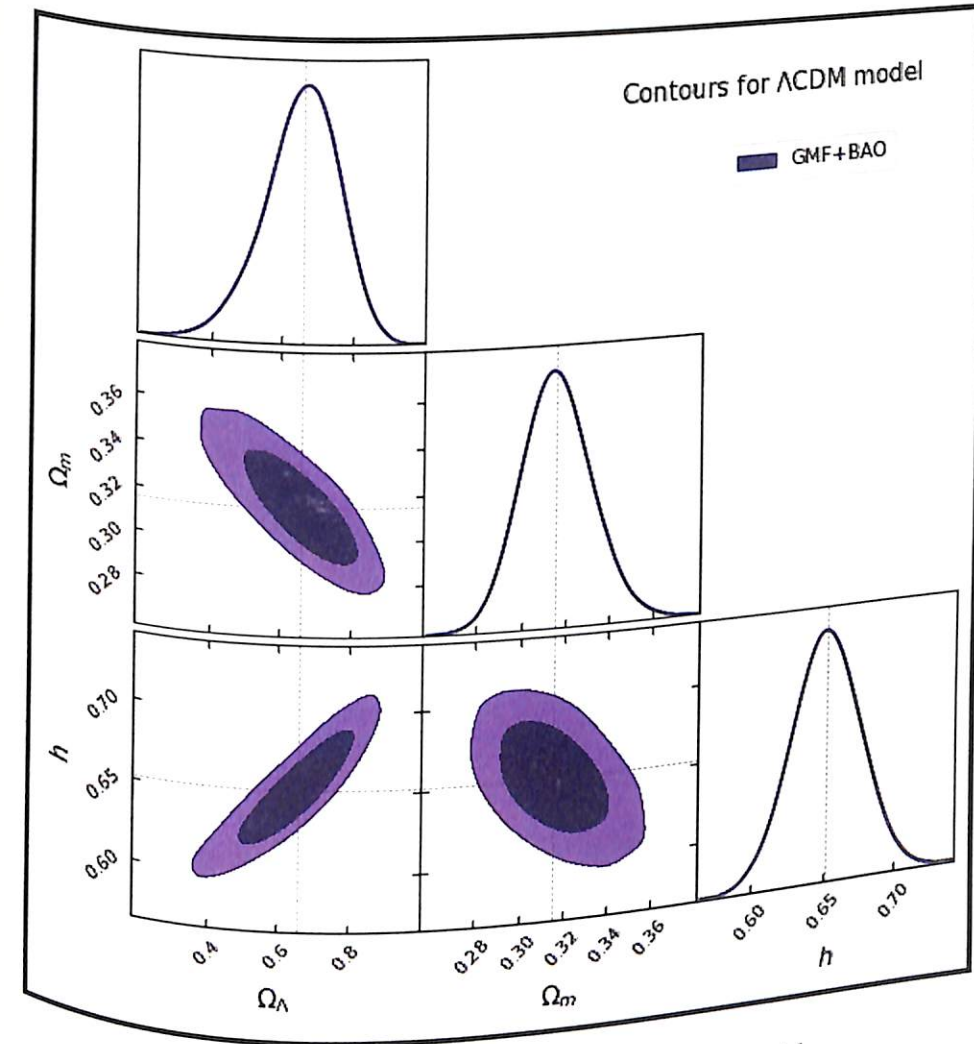


Fig. 01: 1σ and 2σ contours for the cosmological model

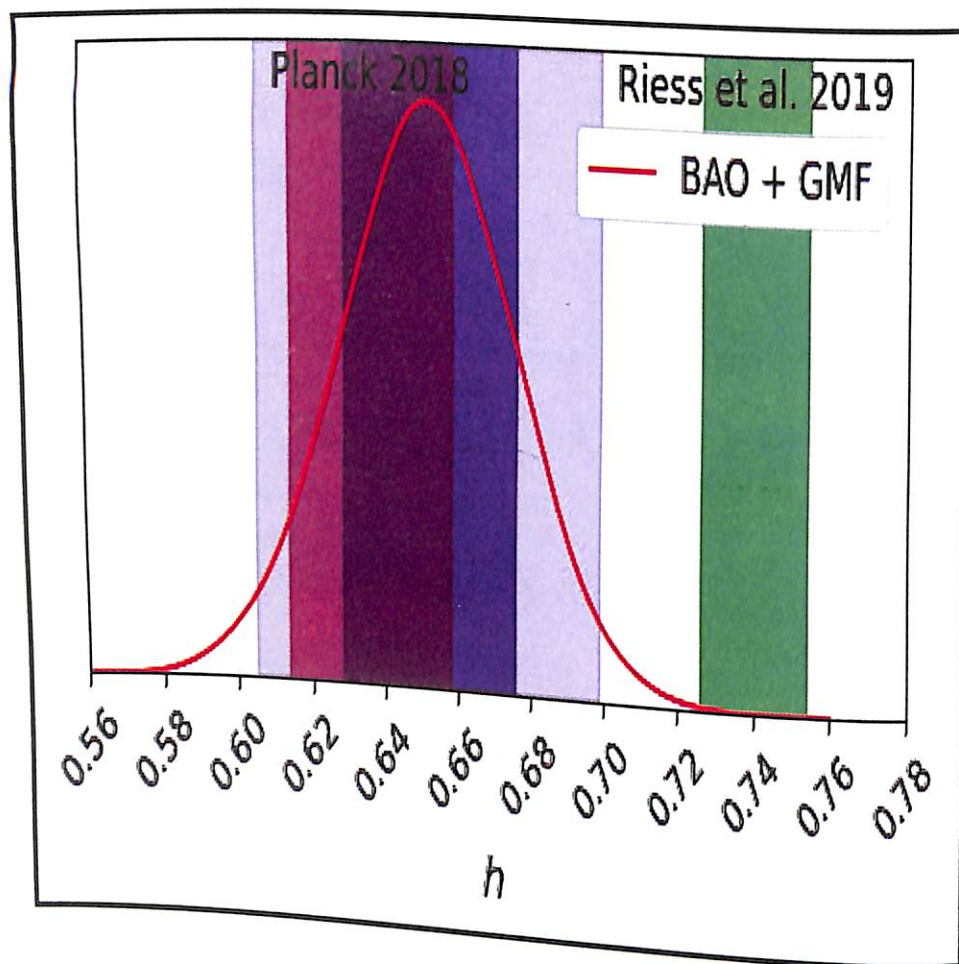


Fig. 02: Marginalized likelihoods of parameter h . Darker and lighter colour regions delimit the 1σ and 2σ confidence levels.

References

- [1] S. H. Suyu, T. Treu, R. D. Blandford, W. L. Freedman, S. Hilbert, C. Blake et al., *The Hubble constant and new discoveries in cosmology*, arXiv e-prints (2012) [1202.4459].
- [2] A. G. Riess, S. Casertano, W. Yuan, L. M. Macri and D. Scolnic, *Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond Λ CDM*, *Astrophys. J.* **876** (2019) 85 [1903.07603].
- [3] Planck collaboration, *Planck 2018 results. vi. cosmological parameters*, arXiv e-prints (2018) [1807.06209].
- [4] D. J. Eisenstein and W. Hu, *Baryonic features in the matter transfer function*, *ApJ* **496** (1998) 605 [astro-ph/9709112].
- [5] F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell et al., *The 6dF Galaxy Survey: baryon acoustic oscillations and the local Hubble constant*, *MNRAS* **416** (2011) 3017 [1106.3366].
- [6] N. Padmanabhan, X. Xu, D. J. Eisenstein, R. Scalzo, A. J. Cuesta, K. T. Mehta et al., *A 2 per cent distance to $z = 0.35$ by reconstructing baryon acoustic oscillations I. Methods and application to the Sloan Digital Sky Survey*, *MNRAS* **427** (2012) 2132 [1202.0090].
- [7] A. J. Ross, L. Samushia, C. Howlett, W. J. Percival, A. Burden and M. Manera, *The clustering of the SDSS DR7 main Galaxy sample I. A 4 per cent distance measure at $z=0.15$* , *MNRAS* **449** (2015) 835 [1409.3242].
- [8] L. Anderson, V. Bhardwaj, C. K. McBride, D. J. Eisenstein, M. E. C. Swanson, S. Escoffier et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples*, *MNRAS* **441** (2014) 24 [1312.4877].
- [9] S. Alam, S. Ho, S. Satpathy, M. V. Magaña, A. Burden, N. Padmanabhan et al., *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample*, *MNRAS* **470** (2017) 2617 [1607.03155].
- [10] E. A. Kazin, J. Koda, C. Blake, N. Padmanabhan, S. Brough, M. Colless et al., *The WiggleZ Dark Energy Survey: improved distance measurements to $z=1$ with reconstruction of the baryonic acoustic feature*, *MNRAS* **441** (2014) 3524 [1401.0358].
- [11] A. Mantz, S. Allen, R. Morris, D. Rapetti, D. Applegate, P. Kelly et al., *Cosmology and astrophysics from relaxed galaxy clusters II. Cosmological constraints*, *Mon. Not. Roy. Astron. Soc.* **440** (2014) 2077 [1402.6212].
- [12] S. W. Allen, A. E. Evrard and A. B. Mantz, *Cosmological Parameters from Observations of Galaxy Clusters*, *Annual Review of Astronomy and Astrophysics* **49** (2011) 409 [1103.4829].
- [13] R. F. L. Holanda, G. Pordeus-da-Silva, S. H. Pereira, *A low Hubble Constant from galaxy distribution observations*, *JCAP* **09** (2020) 053.

Influence of *Mimas* on Saturn's Ring System

Shyamalee Bora⁽¹⁾
DhrubaJyoti Gogoi⁽²⁾

^{(1),(2)}Department of Physics,
Dibrugarh University, Dibrugarh, India
Email: swamaleebora@gmail.com

Abstract

This article delves into the enchanting aesthetics and enigmatic intricacies of Saturn's ring system, placing particular emphasis on the profound impact exerted by its lunar companions on the rings' composition and behaviour. It explores the resonance phenomenon and introduces a sophisticated computer program designed to simulate particle orbits, meticulously analyzing the ramifications of various parameters, including particle count and the gravitational influence of *Mimas*. The findings expose the emergence of gaps and disturbances within the ring system while shedding light on the constraints of the simulation technique employed. Consequently, the study emphasizes the imperative nature of further investigations to enhance our comprehension of Saturn's resplendent rings and their intricately interwoven dynamics.

1. Introduction

One of the most strikingly beautiful and alluring objects of our solar system is the planetary rings. The four giant planets, viz. Jupiter, Saturn, Uranus and Neptune of the outer solar system contains some small particles revolving around them in a disk shaped plane called planetary rings. Many mysteries are hidden in these rings. Some are uncovered till recent times. Some have still remained unanswered. While all the gas giants in our solar system have rings, Saturn possesses the largest and most complex ring system. The wealth of information gathered by the Voyager and Cassini satellite missions has provided valuable knowledge about Saturn's rings, making them an excellent choice for modelling studies.. The Saturn ring

S.Bora ,D.J.Gogoi, Influence of Mimas on Saturn's Ring System, Horizon(Ed.Book),112-121(2023).copyright: Tinsukia College Publishing, ISBN: 978-81-19322-07-7

system is a complex interaction between many particles, moons, and Saturn itself. The Cassini mission revolutionized our understanding of Saturn, its rings, and its moons, leaving behind an invaluable legacy of scientific knowledge and breathtaking images of this majestic ringed planet and its intriguing celestial companions. After more than 13 years of exploration, the Cassini spacecraft executed its "*Grand Finale*" in September 2017. In 2017, the Cassini mission Grand Finale gave us clues about Saturn's ring system. It suggested that the rings might be about 200 million years old or even younger. However, we still don't fully understand how the mass is distributed across the rings. Many of the phenomena observed in the rings are a result of interactions between the ring particles and the orbiting moons. Initially, there were only 18 known moons of Saturn in 1997, but that number has rapidly grown to 57 and continues to increase as more discoveries are made. The reason for the late observation of these moons is that 31 of them have a diameter less than 10 km, making them difficult to detect. Despite their small size, these moons significantly influence the rings and contribute to the visible effects observed in the system. One noticeable feature we can see from Earth is the Cassini Division, a large gap (4500 km wide) between the inner B ring and the outer A ring. Different studies are going on in order to understand the different types of interactions that are helping in the formation of the gaps in between rings, like the Cassini division. Most of them are based on numerical studies, which involve the simulation of the interactions based on computer programmes (Larson, 2017; Mankovich et. al., 2019; Baillie et. al., 2018).

Kirsten Larson, in his studies published as "*The Effects of Moons on Saturn's Ring System*" (2017) created a program to simulate different aspects of the ring system and study how moons affect the orbiting particles. They considered force of gravity on each particle and collisions between them, leading to gaps forming at eight resonant positions. They also discovered that less elastic collisions cause gaps to form more quickly. By embedding a moon in the ring particles, they found that the smaller the mass ratio between Saturn and the moon, the larger the gap cleared in the ring and the wider the ringlet formed in the gap's center. They observed that interior resonances to the moon controlled the inner edge of gaps, while exterior resonances controlled the outer edge of gaps. Based on similar modelling studies, I have created a computer program using Python3 and tried to simulate the orbit of particles around the planet Saturn, specifically focusing on the moon *Mimas*.

2. Resonance in Planetary Rings

The process of resonance happening in the planetary rings can describe the moon-ring particle interactions vividly. Resonance refers to a phenomenon where the gravitational interactions between a moon (or moons) and particles within the ring system create specific stable patterns or orbits. These gravitational interactions can cause particles within the ring to cluster or create gaps, and they occur when the orbital period of a moon or moonlet is related to the orbital period of the particles in the ring in a specific way. Resonances are essential for shaping and maintaining the structure of planetary rings. They can also have significant effects on moonlet and ringlet dynamics, causing particles to collide and aggregate or forming gaps

and eccentricities in the ring system over time. Observing and studying these resonances can provide valuable insights into the dynamics of both the ring particles and the moons that create them. One of the most well-known examples of resonance in planetary rings is the Cassini Division in Saturn's rings. This gap is named after the Italian astronomer Giovanni Cassini, who first observed it in the 17th century. The Cassini Division, between two of Saturn's main rings- A and B rings is a wide gap caused by a 2:1 resonance with the moon *Mimas*. The generalized resonance equation for any resonance (m:n) where m and n are integers is given by:

$$r_p = \left(\frac{n}{m}\right)^{\frac{2}{3}} r_M$$

which works for both resonances inside and outside the orbit of the moon. r_M is the position of the moon, and r_p is the position of resonance.

3. Orbit of the moon

In this work, we have studied the planet, moon and ring interactions under purely classical consideration. Therefore, we assume the orbit of the moon to be circular. By equating the centripetal force with the gravitational force we can get the velocity of the moon:

$$M_M \frac{v^2}{r_M} = F_M = \frac{GM_S M_M}{r^2_M}$$

Where M_M is the mass of the moon, M_S is the mass of Saturn, r is the distance between Saturn and the moon, v is the velocity of the moon. The x and y positions of the moon can be found over time by:

$$x = r_M \times \cos(w_M t)$$

$$y = r_M \times \sin(w_M t)$$

4. Orbit of the Particles

The orbit of the particles is influenced by forces from both Saturn and the moon, which makes it complicated. The acceleration of the particles can be found if Newton's law of gravitation is set equal to Newton's second law:

$$ma = F = \frac{GMm}{r^2}$$

Therefore, the net acceleration of the particles from both moon and Saturn's forces is

$$a = \frac{GMm}{r^3} + \frac{GM_S}{r_s^3}$$

Where r is the distance from the particle to the moon and r_s is the distance from the particle to Saturn.

5. Implementation of Program

The program simulated the rings by orbiting the ring particles around a central planet which is the Saturn. In actual scenario there are countless number of particles in a ring, and they are of different size, shape, and mass. Here, for the program to be useful for our preliminary studies, we constrain the particles to be massless and spherical in size, and all of them are supposed to be of same size. Then we change the number of particles from 1000 to 50000, and simulate the ring. We also study the effect of another parameter called "*mimas-multiplier*" on the formation of the orbit. The variable "*mimas_multiplier*" in the provided code is a parameter that scales the gravitational influence of the moon *Mimas* on the orbiting particles. It is used in the calculation of the acceleration experienced by each particle. Specifically, the acceleration of each particle is determined by two components: the gravitational influence of the central body (Saturn) and the gravitational influence of the moon *Mimas*. The acceleration due to the moon *Mimas* is multiplied by the value of "*mimas_multiplier*", one can control the relative strength of *Mimas* gravitational influence compared to Saturn's. A value of 0 would effectively disable the influence of *Mimas*, while larger values would amplify its effect on the particles' orbits.

The constants used in the program are:

1. Gravitational Constant, $G = 6.6743 \times 10^{-11} \text{ Nm}^2 \text{ kg}^{-2}$
2. Mass of Saturn, $M = 5.6834 \times 10^{26} \text{ kg}$
3. Mass of *Mimas*, $m = 3.7493 \times 10^{19} \text{ kg}$
4. Distance between Saturn and *Mimas*, $d = 1.8552 \times 10^8 \text{ m}$
5. Cassini Division Inner Bound, $d_1 = 1.1758 \times 10^8 \text{ m}$
6. Cassini Division Outer Bound, $d_2 = 1.12217 \times 10^8 \text{ m}$

6. Assumptions

While working on the program, we consider the following assumptions:

1. The Saturn is at the origin (0,0).
2. Particles are initially evenly distributed along the x-axis between a lower and upper limit.
3. The lower and upper limits contain the Cassini Division.
4. Mass of *Mimas* is boosted in order to accelerate results.
5. Position of particles and *Mimas* are calculated for 60 and 1 second time steps respectively.

7. Observation

7.1 Effect of *Mimas-multiplier* on the orbit of the particles

Orbit of Mimas and Ring Particles around Saturn

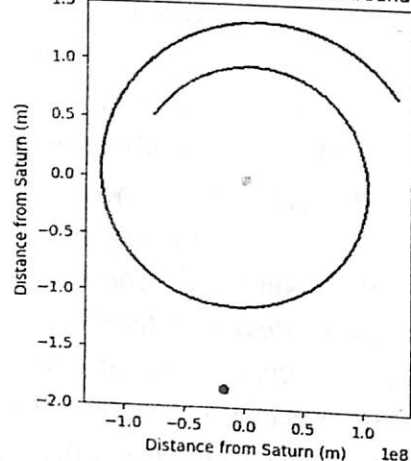


Figure 1 for *Mimas multiplier* = 100

(The yellow point indicates Saturn, Green points indicate the Mimas, and the Red particle indicates the Mimas)

Orbit of Mimas and Ring Particles around Saturn

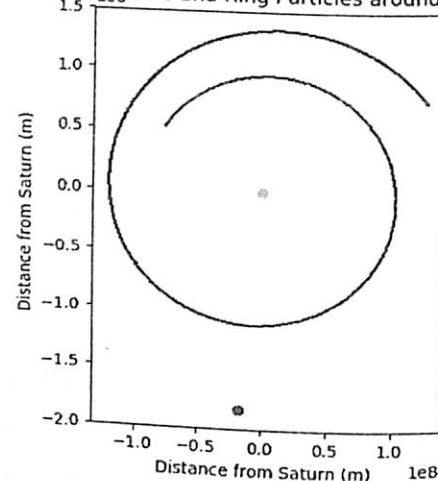


Figure 2. For *mimas-multiplier* = 1000

(The yellow point indicates Saturn, Green points indicate the Mimas, and the Red particle indicates the Mimas)

Orbit of Mimas and Ring Particles around Saturn

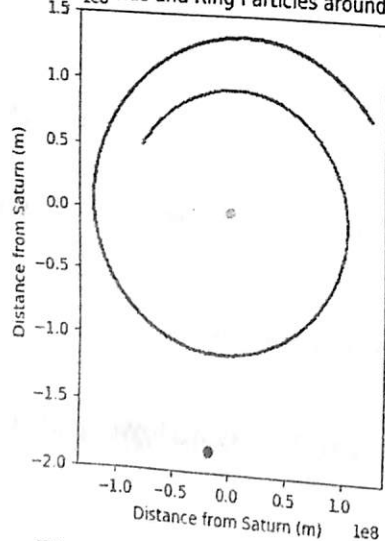


Figure 3. For *mimas-multiplier* = 5000

(The yellow point indicates Saturn, Green points indicate the Mimas, and the Red particle indicates the Mimas)

Orbit of Mimas and Ring Particles around Saturn

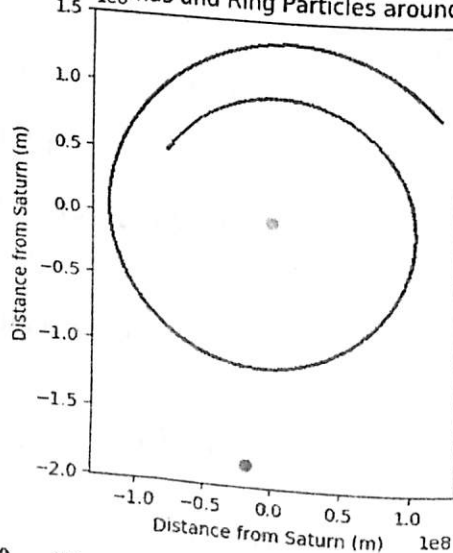


Figure 4. For *mimas multiplier* = 10000

(The yellow point indicates Saturn, Green points indicate the Mimas, and the Red particle indicates the Mimas)

7.2 Simulation of orbit by changing the number of particles

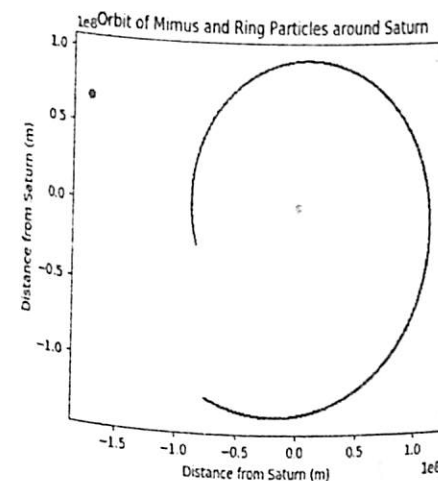


Figure: 5 For number of particles = 1000

(The yellow point indicates Saturn, Green points indicate the Mimas, and the Red particle indicates the Mimas)

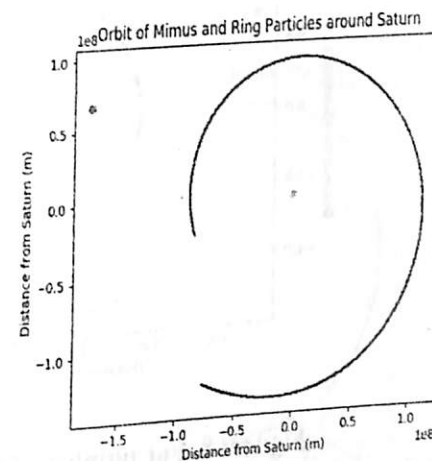


Figure: 6. For number of particles = 2000

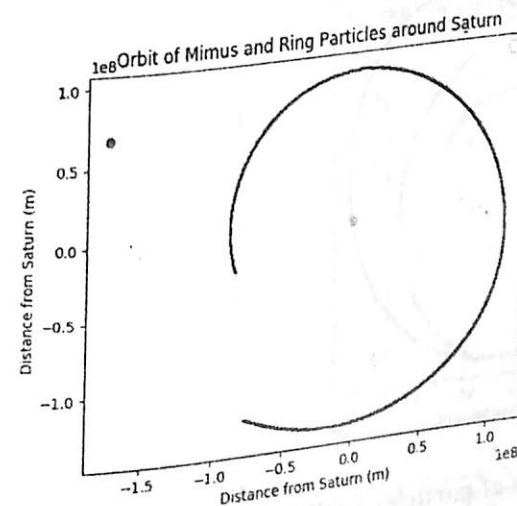


Figure: 7. For number of particles = 3000

(The yellow point indicates Saturn, Green points indicate the Mimas, and the Red particle indicates the Mimas)

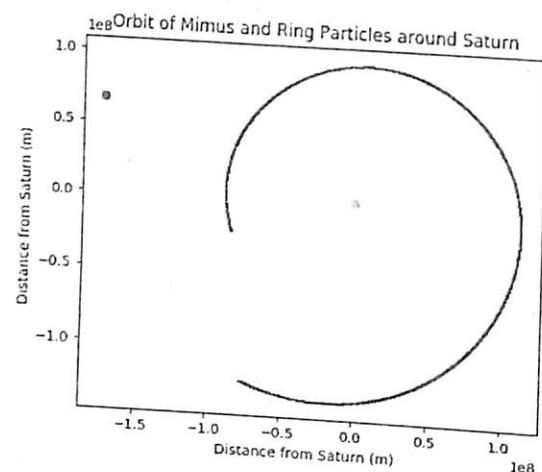


Figure 8. For number of particles = 4000
(The yellow point indicates Saturn, Green points indicate the *Mimas*, and the Red particle indicates the *Mimas*)

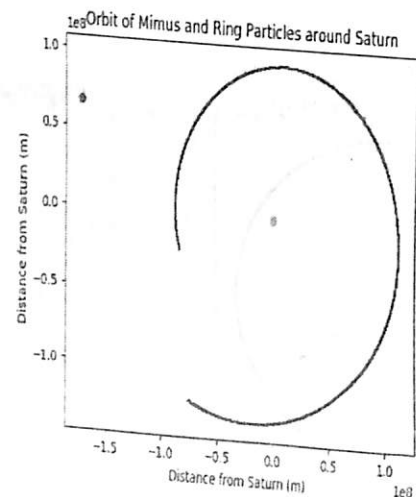


Figure 9. For number of particles = 5000

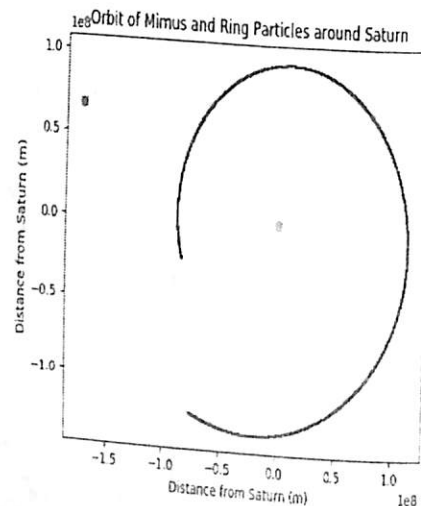


Figure 10. For number of particles = 10000

7.3 Variation of time step: Number of particles is kept constant at $p = 5000$

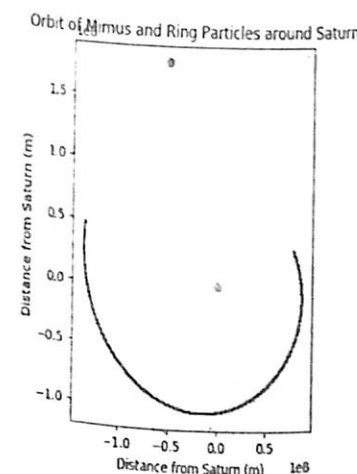


Figure 11. For time step = 400

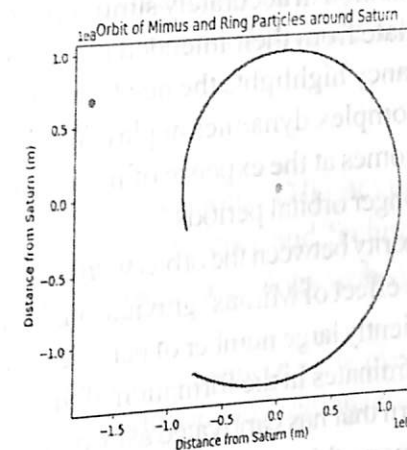


Figure 12. For time step = 600

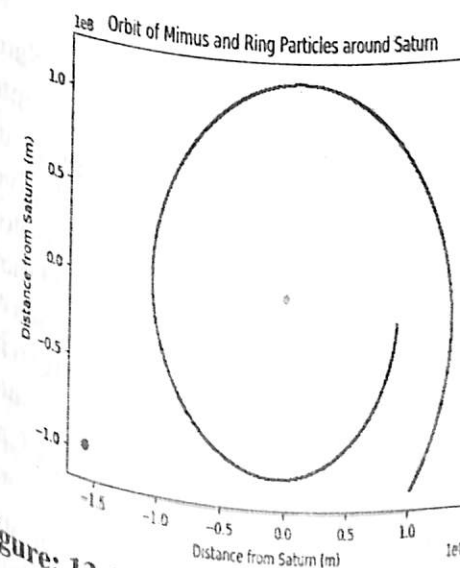


Figure 13. For time step = 800

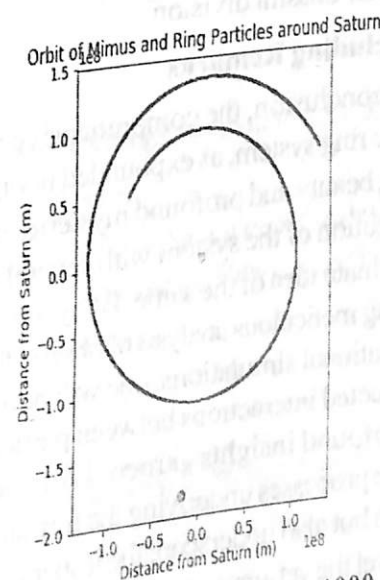


Figure 14. For time step = 1000

6. Particles are massless and are not affected by each other, only by Saturn and *Mimas*.

8. Discussion of Results

The presence of *Mimas* significantly impacts the stability of particle orbits within the ring system, leading to the emergence of gaps and perturbations throughout. The limitations of the Euler method in accurately simulating orbital motion are evident in this study, as the particles deviate from their intended paths, even without the influence of *Mimas*' gravity. This discrepancy highlights the need for more sophisticated computational techniques to capture the complex dynamics at play. While reducing the time step can partially alleviate the issue, it comes at the expense of increased computational resources, particularly when simulating longer orbital periods.

The disparity between the orbits with and without *Mimas*' gravity is striking, underscoring the profound effect of *Mimas*' gravitational force on the overall structure of the ring system. With a sufficiently large number of particles and finely tuned time steps, *Mimas*' gravitational influence culminates in the formation of the iconic Cassini division, a distinct gap within the rings of Saturn that has captivated astronomers for centuries.

In summary, this study elucidates the significant impact of *Mimas* on the stability of particle orbits within Saturn's ring system. It highlights the limitations of the Euler method in accurately simulating the complex dynamics, emphasizing the need for more advanced computational approaches. The findings underscore the remarkable influence of *Mimas*' gravity in shaping the structure of the ring system, ultimately leading to the formation of the celebrated Cassini division.

9. Concluding Remarks

In conclusion, the comprehensive investigation into the intricate dynamics of Saturn's majestic ring system, as expounded upon in this article, provides a tantalizing glimpse into the alluring beauty and profound mysteries of our cosmic surroundings. The study is preliminary investigation of the system with several assumptions to simplify the problem which gives an approximate idea of the same. By synergistically integrating a diverse range of methodologies, including meticulous analysis of observational data, sophisticated modeling studies, and intricate computational simulations, one will be able to delve deeper into the profound intricacies of the multifaceted interactions between particles, moons, and the celestial grandeur of Saturn itself. The profound insights garnered from such a study or research will not only illuminate the intricate processes underlying the formation of gaps, resonances, and the awe-inspiring Cassini division but also underscore the imperative for sustained and unwavering scientific exploration to unravel the yet-unresolved enigmas concealed within these mesmerizing celestial phenomena. As a testament to the ceaseless human curiosity and the indomitable quest for knowledge, this study serves as a preliminary investigation, propelling us little closer to the captivating and awe-inspiring realm of planetary rings.

Future Scope:

There are many parameters and many different possibilities of phenomenon to test with the program. We are considering only classical mechanical point of view. Again the ring particles are taken as massless, equal sized spherical points. Therefore, farther study may be done considering the actual scenario. The shape of the orbit is also assumed to be circular here. Our study can be further extended to simulate actual shape of the orbit. The particle dynamics includes many more aspects which are not discussed here, and will be done in near future.

Acknowledgement:

SB acknowledges Dr. B. Pathak for several discussions. DJG acknowledges the contribution and support of the European Cooperation in Science and Technology (COST) Action CA21136 — "Addressing observational tensions in cosmology with systematics and fundamental physics (CosmoVerse)".

Data Availability and Source Code: Data and Python source code used in the study can be provided upon request for legitimate purposes, with proper citations/mentions.

References

1. S. Charnoz, R. M. Canup, A. Crida, & L. Dones, (2017). The Origin of Planetary Ring Systems.
2. P. Longaretti, (2016). Planetary Ring Dynamics — The Streamline Formalism — I. From Boltzmann Equation to Celestial Mechanics.
3. M. S. Tiscareno, (2011). Planetary Rings.
4. T. Walsh, P. Zimmerman, Width of Cassini's Division in Saturn's Rings. *Nature* 230, 233–234 (1971).
5. K. Larson, (2007). The Effects of Moons on Saturn's Ring System.
6. S. Charnoz, L. Dones, L. W. Esposito, P. R. Estrada, M. M. Hedman, in *Saturn from Cassini-Huygens*, M. K. Dougherty, L. W. Esposito, T. Krimigis, Eds. (Springer Netherlands, 2009), pp. 535–573.
7. J. W. Weiss, C. C. Porco, M. S. Tiscareno, *Astron. J.* 138, 272 (2009).
8. M. C. Lewis, G. R. Stewart, *Icarus* 178, 124 (2005).
9. J. A. Burns, J. N. Cuzzi, *Science* 312, 1753 (2006).
10. J. E. Colwell et al., in *Saturn from Cassini-Huygens*, M. K. Dougherty, L. W. Esposito, T. Krimigis, Eds. (Springer Netherlands, 2009), pp. 375–412.

Variations of NDVI and Its Association with Rainfall over North East India

Anindita Borah⁽¹⁾

Niki Gogoi⁽²⁾

Rohit Gautam⁽³⁾

Binita Pathak⁽⁴⁾

Department of Physics,
Dibrugarh University, Dibrugarh, India
Email: borahanindita97@gmail.com

Abstract

Climate change has become a serious concern worldwide due to its multifaceted impact upon the physical as well as socio-economic environment (IPCC, 2013). Evaluating the quantitative relationship between vegetation patterns and climatic influence has become an increasingly important study conducted on both a regional and global scales. We analysed the spatial responses of Normalized Difference Vegetation Index (NDVI) derived from the Moderate Resolution Imaging Spectro Radiometer (MODIS) Terra data as well as the temporal relationships of the IMD-based rainfall over the region of North East India (NEI) from 2002 to 2020. In this study, linear regression was utilised to analyse the statistical correlation between NDVI and rainfall, and their trends. The average spatio-temporal patterns of NDVI indicate a well associated relationship between NDVI variability and climatic factors like temperature and rainfall. During pre monsoon season, the NDVI shows the lowest values attain its peak value in the post monsoon season. NDVI is positively correlated with rainfall during the post monsoon (ON) and winter (DJF), negatively correlated with monsoon (JJAS) and pre monsoon (MAM) over the North East India (NEI) region. Higher values of standard deviation of NDVI (0.030027) are seen during the rainy season over the region. It shows that the fluctuation of the summer monsoon rainfall may be the cause of the largest NDVI variance over the study domain. Between 2002 and 2020, regression analysis showed an increasing trend for the NDVI and rainfall, with estimated slopes of 0.0017 year^{-1} and $10.914 \text{ mm year}^{-1}$, respectively.

A. Borah, N. Gogoi, R. Gautam, B. Pathak, Variations of NDVI and Its Association with Rainfall over North East India, Horizon (Ed. Book), 122-129 (2023), copyright: Tinsukia College Publishing, ISBN: 978-81-19322-07-7

1. Introduction

According to the report of the IPCC, over the past two decades, there have been experiencing significant changes to the global climate. Developing nations like India, where the economy is mostly agro-based and agricultural production depends on summer monsoon rains are far more vulnerable to climate change. Comprehending how environmental changes affect the vegetation is becoming more and more important to understand and allows for improved predictions of global warming, habitat degradation due to biodiversity loss (Pettorelli *et al.*, 2005). As vegetation cover plays a key role in conserving the natural environment, studying the spatio-temporal trend of vegetation is crucial in identifying changes in the natural environment. The increasing or decreasing trend of vegetation can be used to evaluate the effect of mitigation measures like check dams, rainwater harvesting and irrigational programs. NDVI has been considered as a popular and widely used vegetation index which is very efficient for monitoring vegetation conditions and dynamics in several areas of agriculture, forestry, land use change detection and planning, drought monitoring etc. (Nemani & Running 1989; Stoms & Hargrove 2000; Mazvimavi, 2003; Wang *et al.*, 2004; Maselli *et al.*, 2006; Donohue *et al.*, 2009). NDVI, a measure of vegetation cover and biomass, is used as an indicator of primary productivity and crop yield (Revadekar *et al.*, 2012). Satellite data are the most effective data for studying the relationship between vegetation and climate because of their succession on both temporal and spatial scales (Myneni *et al.* 1996, 1998, Tucker *et al.* 2001, Zhou *et al.* 2001, Zhihui *et al.* 2007). Rainfall is a key factor that controls dynamic changes in vegetation growth in arid and semi-arid regions as vegetation cover in those regions are extremely susceptible to the quantity, duration and frequency of rainfall (Fravolini *et al.*, 2005; Liu *et al.*, 2012; Gamon *et al.*, 2013; Dutta *et al.*, 2015; Kundu *et al.*, 2015c). Analysing the interrelationship between rainfall and vegetation is significant in the tropical regions of the world. This climatic region is characterized by heavy rainfall which is considered as a major controlling factor of vegetation growth. Satellite-derived vegetation indices indicate increasing vegetation productivity over North East India during recent decades (Krishnan *et al.*, 2020). With increasing population and anthropogenic impacts on the environment of this region is facing significant changes. Several regional and global studies have earlier studied the inter-relationship among NDVI and rainfall in various regions of the world (Kawabata *et al.*, 2001; Wang *et al.*, 2005; Groeneveld & Baugh, 2007; Kundu & Dutta, 2011; Kundu *et al.*, 2017). According to Mendez-Barroso *et al.*, (2009), when rainfall increases, vegetation cover may not experience an enormous increase, since water is no longer the limitation to increase. The variations of rainfall have sturdy effects on the vegetation cover in yearly and decadal intervals (Snyder & Tartowski, 2006). From an Indian perspective, Prasad *et al.*, (2005) found the strongest relationship among NDVI and rainfall in the evergreen forests of North India. This study aims to assess the spatio-temporal trend of NDVI estimated through long-term satellite data and its response to rainfall.

From a climatic point of view, the climate of North East India is comprised of four climatic seasons: Winter (December, January and February; DJF), Pre-monsoon (March, April and May; MAM), Monsoon (June, July, August and September; JJAS) and Post-monsoon (October, November; ON). The spatial patterns of rainfall are strongly influenced

by the intricate orographic features of the region. The region receives the largest amount of rainfall and highest temperature in summer monsoon. So, this region shows significant variations in vegetation over different regions in different months or seasons. Many other researchers have also studied the variability of vegetation and its relation to meteorological parameters over the Indian region (Singh *et al.* 2002, 2003, Sarkar *et al.* 2004, Prasad *et al.* 2007, Krishna Prasad *et al.* 2008, Saikia 2009). The present study attempts a detailed analysis of NDVI variation and its relation to rainfall over the North East Indian region.

2. Study Area

North-East India (NEI, 22°–30°N, 88°–98°E) consists of the states of Assam, Arunachal Pradesh, Manipur, Meghalaya, Mizoram, Nagaland, Tripura and Sikkim, covering an area of 262,184 km² located at the eastern Himalayan foothills. Devoid of large industrial activity, the entire population is mostly dependent on natural resources such as two major rivers (the Brahmaputra and the Barak), various tributaries, small or medium sized wetlands, and forests are vulnerable to air quality and climate change. NEI comprises 66% vegetation and vast water bodies (Pathak *et al.*, 2022). The proposed objectives will be carried out for the entire region based on varying altitudes and different distribution of vegetation to examine the trends of NDVI and rainfall over NEI.

The region is characterized for its rich biodiversity and heavy precipitation. It occupies a unique position in relation to the Indian summer monsoon. The climate is predominantly humid subtropical with hot, humid summers, severe monsoons, and mild winters.

3. Data and Methodology

3.1 NDVI data

The NDVI is a popular spectral vegetation index acquired from remote sensing derived satellite.
$$NDVI = \frac{NIR - RED}{NIR + RED}$$
 near infrared segments of the spectrum. It is expressed as follows,

NDVI values can range from “-1 to 1 but typical values range from 0.1 to 0.7, with higher values associated with greater density and greenness of plant canopies. NDVI value of zero means no vegetation and a value close to +1 (0.8–0.9) indicates the highest possible density of green leaves. Because our objective is to investigate the large-scale features in climate and NDVI, we used the yearly and seasonal datasets in high resolution.

NDVI derived from time-series Moderate Resolution Imaging Spectro radiometer (MODIS) 16-day products (MOD13Q1) at 250 m spatial resolution for the 18 years during 2002 to 2020 (<https://appears.earthdatacloud.nasa.gov>).

3.2 Rainfall data

Rainfall data was obtained from the IMD from 2002 to 2020 (<https://indiawris.gov.in>). Using these datasets, yearly and seasonal rainfall patterns are analysed for the NEI region.

3.3 Methodology

In this study, data analysis is being done via python, excel and QGIS software. We draw conclusions in light of various findings and trends observed seasonally and annually.

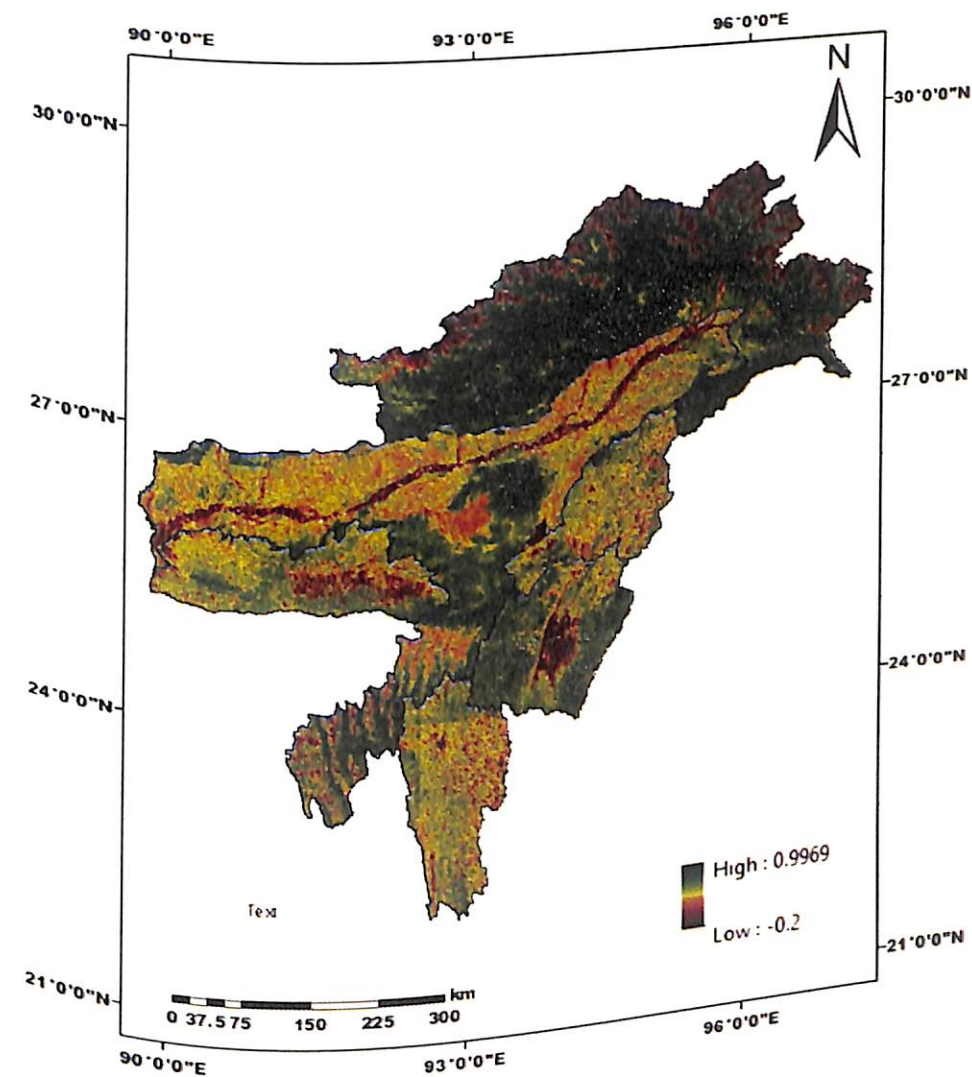


Fig: NDVI (Normalized Difference Vegetation Index) over the NEI Region

5. Conclusion

Rainfall varies from 1,000 to over 4,000 mm on average every year, with roughly 60% of it falling from June to October. (Saikia, 2009). The study adds to the limited body of knowledge about the relationship between NDVI and environmental factors in high precipitation regions. However, when only growing season rainfall and NDVI correlations were run, the significance values were relatively much higher, except for the irrigated agriculture sites. Although all climatic factors like rainfall variability shows a strong impact on the inter-annual variability of NDVI. During drought years, there is a decrease in NDVI for all of the months. This clearly suggests that the soil moisture gained during flood conditions helps the NDVI to increase. The present analysis provides a general insight into the role of climate variability on NDVI. However, it is necessary to examine the relationship of NDVI with all regional weather parameters for a better understanding of their impact. The study contributes to our understanding of the temporal response of vegetation to varying climatic conditions. The identification and prioritising of a forestation strategies in NEI is one of the study's primary consequences. On some NEI areas, we recommend the introduction and promotion of native flora-based green space development. It's also crucial to look closely at how aerosols and large-scale atmospheric processes like the Southern Oscillation and North Atlantic Oscillation affect the NDVI.

Acknowledgements

The authors are thankful to NASA for providing the temporal MODIS 250m products. We thank the National Water Informatics Centre India, New Delhi (<https://indiawris.gov.in>) for the monthly rainfall data used in this work.

References

1. Fravolini A, Hultine KR, Brugnoli E, Gazal R, English NB, Williams DG (2005) Precipitation pulse use by an invasive woody legume: the role of soil texture and pulse size. *Oecology* 144, 618–27.
2. Gamon JA, Huemmrich KF, Stone RS, Tweedie CE (2013) Spatial and temporal variation in primary productivity (NDVI) of coastal Alaskan tundra: decreased vegetation growth following earlier snowmelt. *Remote Sensing of Environment* 129, 144–53.
3. IPCC (Intergovernmental Panel on Climate Change) (2013) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. In Stocker TF, Qin D, Plattner GK et al., (eds) Cambridge University Press, Cambridge, United Kingdom and New York.
4. Kawabata A, Ichii K, Yamaguchi Y (2001) Global monitoring of interannual changes in vegetation activities using NDVI and its relationships to temperature and precipitation. *International Journal of Remote Sensing* 2, 1377–82.
5. Krishnan, R., Sanjay, J., Gnanaseelan, C., Mujumdar, M., Kulkarni, A., Chakraborty, S., *Assessment of Climate Change over the Indian Region*, Springer, 2020, pp. 74.
6. Krishna-Prasad, V., Badarinath K. V. S. & Eaturu, A. (2008) Effects of precipitation, temperature and topographic parameters on evergreen vegetation greenery in the Western Ghats, India, *International Journal of Climatology*, 28, pp. 1807–1819.

7. Kundu, A., Denis, D. M., Patel, N. R., & Dutta, D. (2018). A Geo spatial study for analysing temporal responses of NDVI to rainfall. *Singapore Journal of Tropical Geography*. 39(1), 107–116.
8. Meng, M., Ni, J., & Zong, M. (2011). Impacts of changes in climate variability on regional vegetation in China: NDVI-based analysis from 1982 to 2000. *Ecological Research*. 26, 421–428.
9. MYNENI, R.B., LOS, S.O. and TUCKER, C.J., 1996, Satellite-based identification of linked vegetation and sea surface temperature anomaly areas from 1982–1990 for Africa, Australia, and South America. *Geophysical Research Letters*, 23, pp. 729–732.
10. MYNENI, R.B., TUCKER, C.J., ASAR, G. and KEELING, C.D., 1998, Interannual variations in satellite-sensed vegetation index data from 1981 to 1991. *Journal of Geophysical Research*. 103, pp. 6145–6160.
11. Nemani RR, Running SW (1989) Estimation of regional surface resistance to evapotranspiration from NDVI and thermal IR-AVHRR data. *Journal of Applied Meteorology* 28, 276–84.
12. Pathak, B., Khataniar, A., Das, B., Upadhyaya, S., Medhi, A., Bhuyan, P. K., ... & Borah, D. (2022). Spatio-temporal diversity of biological aerosols over Northeast India: a metagenomic approach. *Environmental Science and Pollution Research*, 29(42), 64096–64111.
13. Pettorelli, N., Vik, J. O., Myrsetrud, A., Gaillard, J. M., Tucker, C. J., & Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in ecology & evolution*, 20(9), 503–510.
14. Revadekar, J. V., Tiwari, Y. K., & Kumar, K. R. (2012). Impact of climate variability on NDVI over the Indian region during 1981–2010. *International Journal of Remote Sensing*, 33(22), 7132–7150.
15. Saikia, A. (2009). NDVI variability in North East India. *Scottish Geographical Journal*, 125(2), 195–213.
16. SINGH, R.P. and KOGAN, F.N., 2002, Monitoring vegetation condition from NOAA operational polar-orbiting satellites over Indian region. *Journal of the Indian Society of Remote Sensing*, 30, pp. 117–118.
17. SINGH, R.P., ROY, S. and KOGAN, F.N., 2003, Vegetation and temperature condition indices from NOAA-AVHRR data for drought monitoring over India. *International Journal of Remote Sensing*, 24, pp. 4393–4402.
18. Snyder KA, Tartowski SL (2006) Multi-scale temporal variation in water availability: implications for vegetation dynamics in arid and semi-arid ecosystems. *Journal of Arid Environments* 65, 219–34.
19. Wang Q, Adiku S, Tenhunen J, Granier A (2005) On the relationship of NDVI with leaf area index in a deciduous forest site. *Remote Sensing of Environment* 94, 244–55

Dust Coupling Strength in presence of Asymmetric Ion Flow

Saurav Bhattacharjee

Department of Physics

Tinsukia College,

Tinsukia, Assam, India

Email: sauravtsk.bhattacharjee@gmail.com

Abstract

Interaction mechanism among dust grains immersed in plasma has been reported theoretically in presence of asymmetric ion flow and dynamical charging of dust grains. The ion flow induced particle-wake interaction in addition to repulsive Yukawa type of interaction found to play a significant role in tuning the coupling strength between charged grains and hence the vibrational mode. The transition in coupling strength is obtained as a characteristic of ion flow speed and grain size. The results show a damping in strength of interaction in presence of dynamical charging of dust. The type of interaction shows a mixed phase of Wake-Yukawa in subsonic regime and highly Yukawa dominating phase in supersonic regime of ion flow. The results are significant for understanding the thermodynamic properties and the phase behaviour of colloidal systems and nano-crystals.

1. Introduction

The study of inter grain interaction is of fundamental interest in statistical thermodynamics. The phase behavior and the dynamics of a system depend solely on the type and strength of inter grain interaction. Realization of correct and appropriate interaction potential is one of the biggest challenges in many-body systems. The complex plasma or dusty plasma is an emerging branch of research for last few decades because of its strong potentiality to replicate the strongly correlated systems at larger scale. The complex plasma consists of micron-submicron sized dust grains immersed in ambient plasma and radiative background. The dust grains are

S. Bhattacharjee, "Dust Coupling Strength in presence of Asymmetric Ion Flow", Horizon (Ed Book) 130-141(2023). copyright: Tinsukia College Publishing. ISBN: 978-81-19322-07-7

charged up to a high value of 10^3 to 10^4 electronic charges by various processes [1-2]. In laboratory dusty plasma experiments, the dust grains are commonly negatively charged due to higher electron mobility towards the grains. Because of comparatively larger size of the grains, they can be observed individually and this makes it possible to investigate various phenomena such as phase transition [3], diffusion, self organization [4] etc. at macroscopic kinetic level.

The interaction mechanism around a test dust charge in an isotropic system is caused mainly due to spherically symmetric Yukawa type of potential [5]. Under suitable conditions, when the intergrain interaction is much higher as compared to their thermal energy, the grains show some regular oriented structures, the so called plasma crystals [4,6]. Near the plasma sheath, where laboratory plasma crystals are formed by balancing the gravity and sheath electric field, there is ion streaming towards the lower electrode. This bombardment of ions over shield the spherical dust Debye cloud and elongate it in lower vertical direction. This asymmetry in charge distribution arises around the grain along the direction of ion flow. This access accumulation of positive charge in the lower side of the grain attracts other dust grains just below it to align along vertical ion flow direction [7] and shows vertical ordering of dust grains (chain formation) often observed in laboratory experiments [8-10]. This leads to an attractive type of interaction potential along vertical direction named as wake potential [11-12].

Most of the theoretical results aim to study the intergrain interaction is based on stationary charge on dust surface. However, in reality the charge on the dust grains exhibit self-consistent fluctuation in response to the oscillation in plasma currents flowing into them. The charge fluctuation on dust surface leads to some new observations [13-14] based on the damping of electrostatic wave [15].

The study of inter grain interaction is primarily based on the local plasma environment surrounding the dust. This also helps to reduce the multi-component scenario of dusty plasma to one component plasma (OCP) model of dust subsystem. In this letter, we have reported analytical expressions of intergrain interaction potential having background asymmetric ion flow towards the sheath and self-consistent fluctuation of dust charge, for two vertically placed dust subsystem. The potentials thus obtained show a damping in strength in presence of dust charging dynamics.

Moreover, the strength of interaction along the direction of ion flow is found as a characteristic ion flow speed and grain size. The type of interaction in subsonic regime of plasma flow is found in a mixed phase of repulsive Yukawa and attractive wake potential, whereas the supersonic regime is found in Yukawa dominating phase. The presence of this mixed phase in subsonic regime shows an interesting anomalous phase transition of plasma crystal reported recently by Saurav et.al. [16], which is commonly observed in solid state crystals undergo ion irradiation [17].

Since, the interaction potential is the key to understand the phase behavior of a system, the results reflect a tuned transition in crystal phase based on ion flow speed and particle size,

which is of fundamental scientific interest in metal nano crystals [18-19] because phase control is the only step to improve the functionalities for various applications.

2. Theoretical Formulation

Our theoretical model composed of an uniform dusty plasma having background Boltzmanian electrons, cold streaming ions and fluctuating negatively charged dust grains, in a collision less scenario, where the collision mean free path is greater than the dust Debye length λ_{De} near the plasma sheath [5, 20]. Dust grains near the plasma sheath encounter a downstream ion flow with uniform velocity u_{io} along z-direction towards the plasma sheath. Under the condition that dust particle radius $a < \lambda_{De}$, one can apply OML (orbital motion limited) theory [21] to describe the charging currents. The dielectric response function in presence of dust charging dynamics is given by [22]

$$\epsilon(\omega, k) = 1 + \frac{1}{k^2 \lambda_{De}^2} (1 + P) - \frac{k_{\parallel}^2}{k^2} \frac{\omega_{pi}^2}{(\omega - u_{io} k_{\parallel})^2} (1 + Q) \quad (1)$$

Where propagation vector $k^2 = k_{\parallel}^2 + k_{\perp}^2$, P and Q are normalized electron and ion attachment frequency to dust grain and are given as

$$P = \frac{|I_{eo}|}{e(\eta - i\omega)} \frac{n_{do}}{n_{eo}},$$

$$Q = \frac{|I_{eo}|}{e(\eta - i\omega)} \frac{n_{do}}{n_{io}} M \sqrt{\frac{\pi m_e}{8 m_i}} (n_{io}/n_{eo})^{3/2} \left(1 - \frac{2eQ_{do}}{m_i u_{io}^2 r_d} \right) \exp\left(-\frac{Q_{do} e}{K_B T_e r_d}\right)$$

Where $|I_{eo}|$ is the equilibrium electron current and is given by

$$I_{eo} = \pi a^2 e \sqrt{\frac{8 k_B T_e}{\pi m_e}} n_e \exp(e\phi_s / k_B T_e) \quad \text{for } \phi_s < 0.$$

Here $\phi_s = \frac{Q_{do}}{r_d}$ is the dust surface potential or floating potential with respect to surrounding plasma.

$$\eta = \frac{1}{\tau_c} = \frac{e I_{eo}}{a k_B T_e} \left[1 + \frac{1}{M} \sqrt{\frac{\pi m_e n_{io}}{2 m_i n_{eo}}} \exp\left(\frac{-e Q_{do}}{k_B T_e a}\right) \right]$$

τ_c is the dust charging time, ω is the frequency of the ion acoustic wave and $k_{\parallel}$ is the propagation parallel to ion flow (along z-direction) towards the plasma sheath.

Using the integral form of electrostatic Poisson equation and dielectric response function given in equation (1), we can write the interaction potential due to a moving test dust charge Q_{dt} based on the model described above as follows

$$\phi(r, t) = \frac{Q_{dt}}{(2\pi)^3 2\pi \epsilon_0} \int d^3 k d\omega \frac{\delta(\omega - v_t k)}{k^2 \epsilon(\omega, k)} e^{ik \cdot r'} \quad (2)$$

Where, $r' = r - v_t t$

For cold stationary dust grains, $v_t = 0$, as compatible to our physical model, the interaction potential is modified as

$$\Phi(r) = \frac{Q_{dt}}{(2\pi)^3 \epsilon_0} \int d^3 k \frac{1}{k^2 \epsilon(0, k)} e^{ik \cdot r} \quad (3)$$

In steady state and normalized scale of electron Debye length, the dielectric response function given in equation (1) can be written as

$$\epsilon(0, k) = 1 + \frac{1+P}{k^2} - \frac{1+Q}{M^2 k_{\parallel}^2} \quad (4)$$

Where it is assumed that $k_{\parallel} \gg k_{\perp}$.

Above equation can be simplified as follows

$$\frac{1}{k^2 \epsilon(0, k)} = \frac{k_{\parallel}^2}{(k_{\parallel}^2 - k_R^2)(k_{\parallel}^2 - k_{Im}^2)} \quad (5)$$

Where all k are normalized with λ_{De} . k_R and k_{Im} corresponds to real and imaginary screening modes as a function of $k_{\perp}$, in a complex $k_{\parallel}$ plane.

(6)

$$k_R = \sqrt{-A + B}$$

$$k_{Im} = \sqrt{-A - B} \quad (7)$$

Where A and B are given as

$$A = \frac{1}{2}(k_{\perp}^2 + 1 + P - \frac{1+Q}{M^2}) \quad (8)$$

$$B = \sqrt{A^2 + \frac{k_{\perp}^2}{M^2}(1+Q)} \quad (9)$$

Where $\sqrt{y^2 + z^2}$, all y and z are normalized with electron Debye length λ_{De}

The second part of the above integral can be solved by performing closed contour integration

in a complex $k_{\parallel}$ plane for two poles corresponding to $\omega = \omega_r$ (real one) gives oscillatory wake potential and $\omega = \omega_i$ (imaginary one) gives repulsive Yukawa type of potential. Finally using the saddle point integration method one can get the potentials as follows

$$\phi_w(0, z) = \frac{-Q_{dt}}{4\pi\epsilon_0\lambda_{De}} \sqrt{\frac{2\pi}{6zM}} \frac{(1+Q)^{1/4}}{\sqrt{1+P}} \sin\left[\frac{z}{M}(1+Q)^{1/2}\right] \quad (11)$$

$$\phi_D(y, z) = \frac{Q_{dt}\sqrt{\pi}}{4\pi\epsilon_0\lambda_{De}} \exp(-\sqrt{1+P}\sqrt{y^2 + z^2}) \frac{(1+P)^{7/4}}{(y^2 + z^2)^{1/4}} (z/y)^{1/2} \quad (12)$$

The expressions for corresponding interaction potential in absence of dust charge fluctuation can be obtained by putting $P=Q=0$.

The normalized equation of motion for the mode of vibration of two vertically (z-direction) placed dust grains above the plasma sheath in presence of asymmetric ion flow towards the sheath is given as

$$\frac{d^2 z_{n1}}{dt^2} = [K_z + K_{zw}](z_{n1} - z_{m1}) \quad (13)$$

Here, we have considered the vibrational mode propagates along y-direction. (n:m) represents the particle pair and $Z_{(n,m)}$ stands for their displacements from equilibrium positions.

The length and time are normalized by λ_{De} and $\sqrt{m_d\lambda_{De}^2/K_B T_d}$ respectively having K_B as the Boltzmann constant and T_d is the dust temperature. K_z and K_{zw} are the spring constant or coupling strength correspond to repulsive Yukawa and attractive wake potential respectively. Using equation (11), we can write the expression for the normalised Yukawa potential as

$$K_z = \left(\frac{d^2\phi_D}{d(y,z)^2} \right)_{y,z=a} = \frac{\Gamma\sqrt{\pi}}{2^{9/4}} \frac{(1+P)^{7/4}}{k^{3/2}} \exp(-\sqrt{2(1+P)k}) [5/4 + 2^{3/2}\sqrt{1+Pk} + 2(1+P)k^2] \quad (14)$$

Here, a is the average inter-particle distance, is the Coulomb coupling parameter [9]

and $k = \frac{a}{\lambda_{De}}$ is the screening constant.

Similarly, using equation (12), we get the spring constant for wake potential as

$$K_{zw} = -\sqrt{\frac{2\pi}{6M}} \frac{(1+Q)^{1/4}}{\sqrt{1+P}} \frac{1}{k^{5/2}} \left[\left(\frac{3}{4} - \frac{k^2(1+Q)}{M^2} \right) \sin\left(\frac{k}{M}\sqrt{1+Q}\right) - \frac{z\sqrt{1+Q}}{M} \cos\left(\frac{z}{M}\sqrt{1+Q}\right) \right] \quad (15)$$

Thus, the effective spring constant along vertical z-direction is

$$K_{eff} = K_z(\text{Yukawa}) + K_{zw}(\text{Wake}) \quad (16)$$

Considering a small oscillation in wave amplitude $z_{n1} = A_z \exp[(k_y a - \omega t)]$ and substituting this in equation (13), we get the dispersion relation for the transverse vibrational mode as

$$\omega^2 = 2(k_z + K_{zw}) \sin^2\left(\frac{k_y a}{2}\right) \quad (17)$$

3. Results and Discussion

The ion flow induced elongation of spherical dust Debye cloud and the formation of attractive wake potential in addition to repulsive Yukawa type of potential pulls the dust grain just below it to displace along the direction of ion flow. Thus, the effective strength of attraction along vertical direction is mainly depends upon the amount of distortion in Debye cloud and relative strength of both repulsive Yukawa and attractive wake potential respectively. From equation (1) it is clearly visible that the dust Debye length λ_{De} around the grain is modified in two different ways: first, by a factor $\frac{1}{\sqrt{1+P}}$ and is attributed to shift in Boltzmannian electron distribution having normalized electron attachment frequency P and second, by a factor in terms of ion flow speed M and normalized ion attachment frequency Q. Both the changes

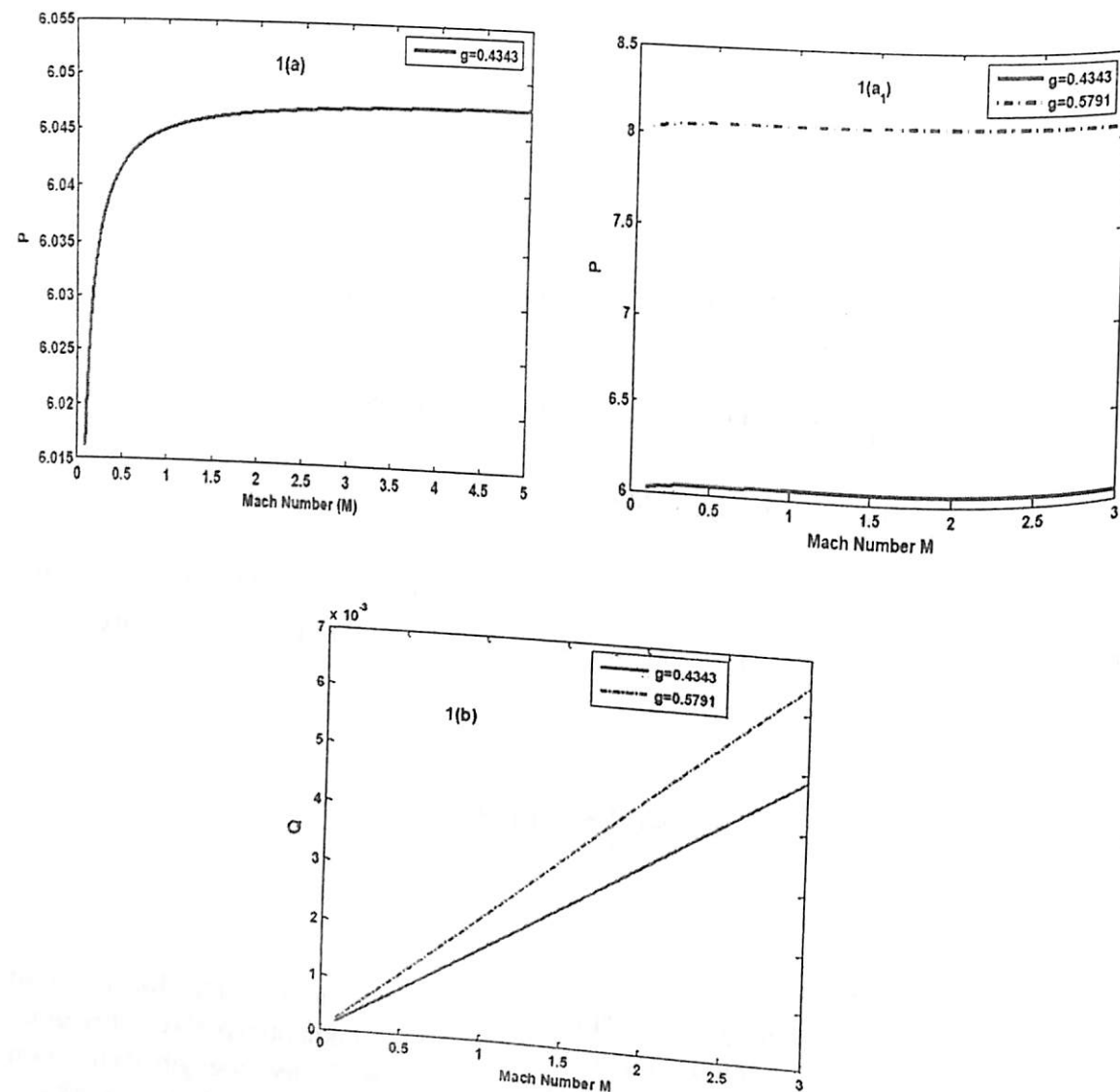


FIG. 1: The plot shows (a) the variation of electron attachment frequency P vs. M (Mach number), for a given normalized dust radius $g=0.4343$. (a1) P vs. M , for $g=0.4343$ (red solid) and $g=0.5791$ (red dash). (c) Ion attachment frequency Q vs. M , for $g=0.4343$ (red solid), $g=0.5791$ (red dotted).

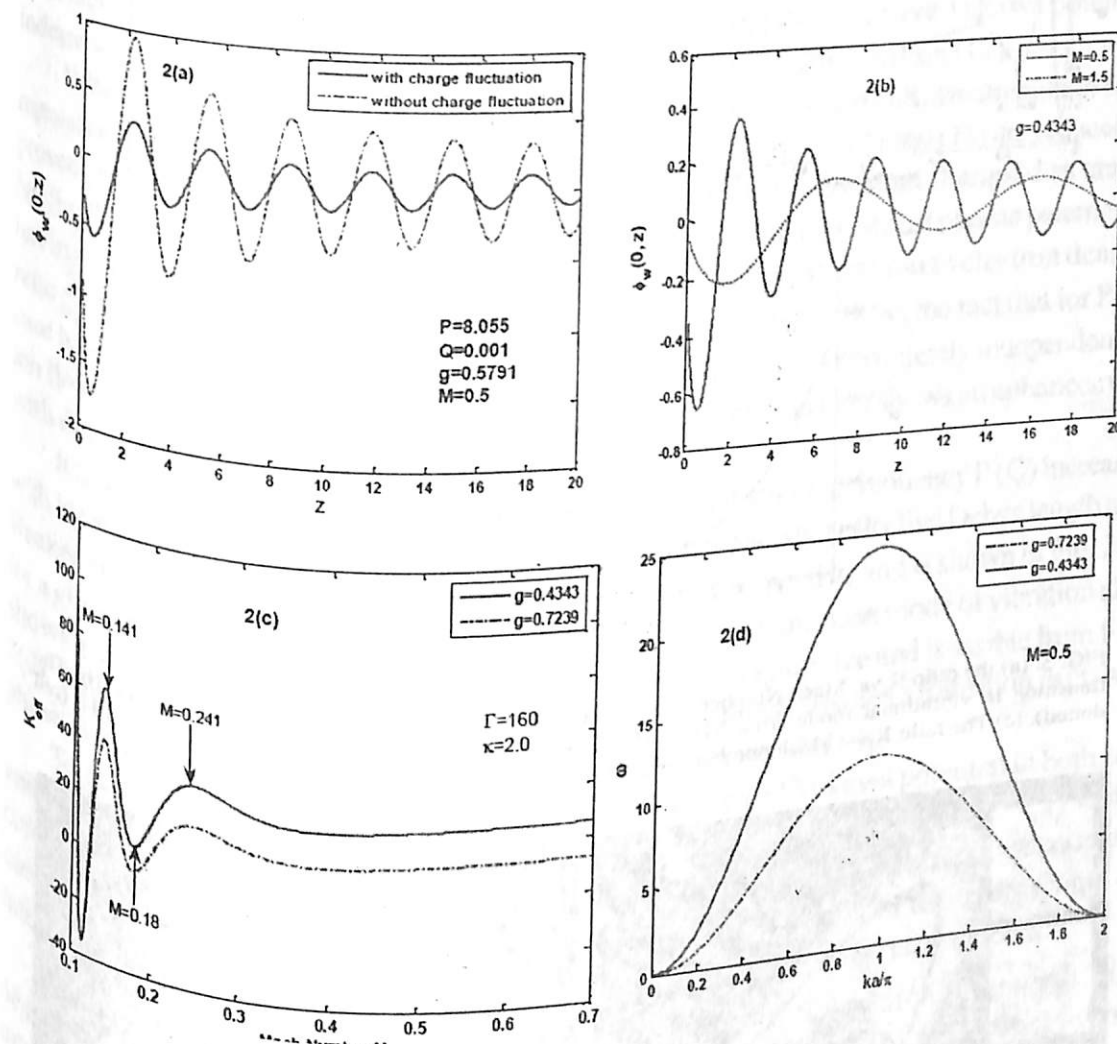


FIG. 2: The profiles show (a) Normalized wake potential in presence (red solid) and absence (red dotted) of dust charge fluctuation. (b) Normalized wake potential, for Mach number $M=0.5$ (red solid) and 1.5 (red dotted). (c) The effective vertical spring constant vs. Mach number M , having normalized grain size $g=0.4343$ (red solid) and $g=0.7239$ (red dotted), for given plasma parameters and showing a transition from $M=0.241$ to $M=0.18$ to $M=0.141$. (d) Dispersion relation for transverse vibrational modes having $g=0.4343$ (red solid), $g=0.7239$ (red dash) and Mach number $M=0.5$.

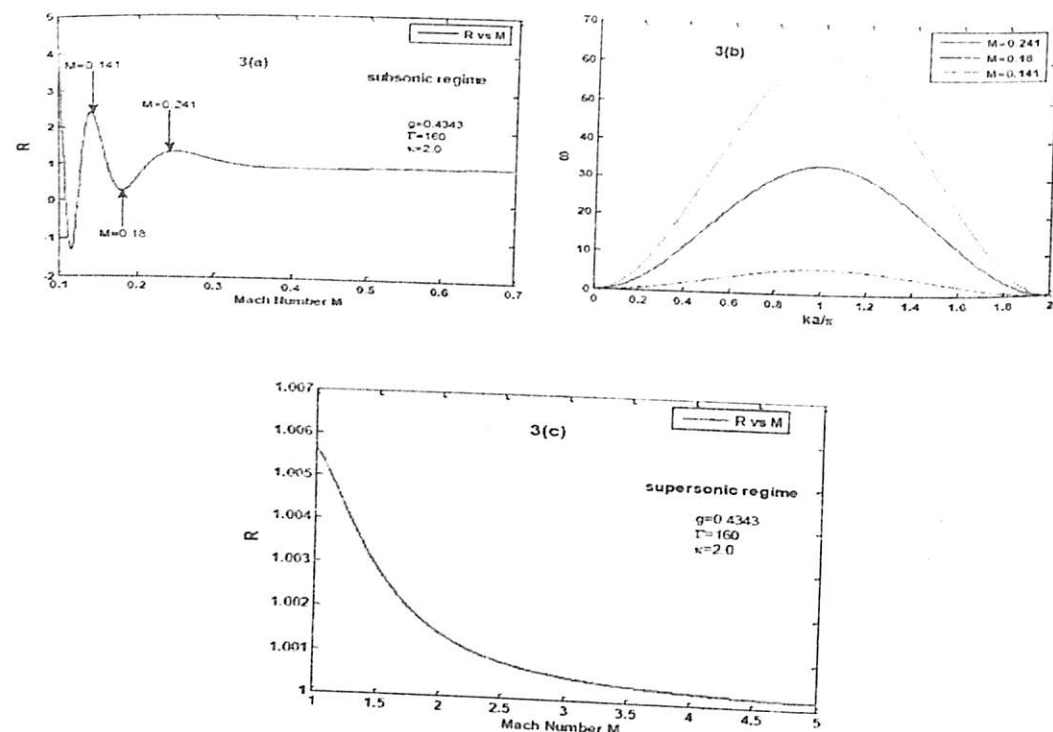


FIG. 3: (a) the ratio R vs. Mach Number M , (subsonic regime) for the given set of parameters. (b) The transition in vibrational mode from $M=0.241$ (red solid) to $M=0.18$ (red dashed) to $M=0.141$ (red dotted). (c) The ratio R vs. Mach number, (supersonic regime) for the given set of parameters.

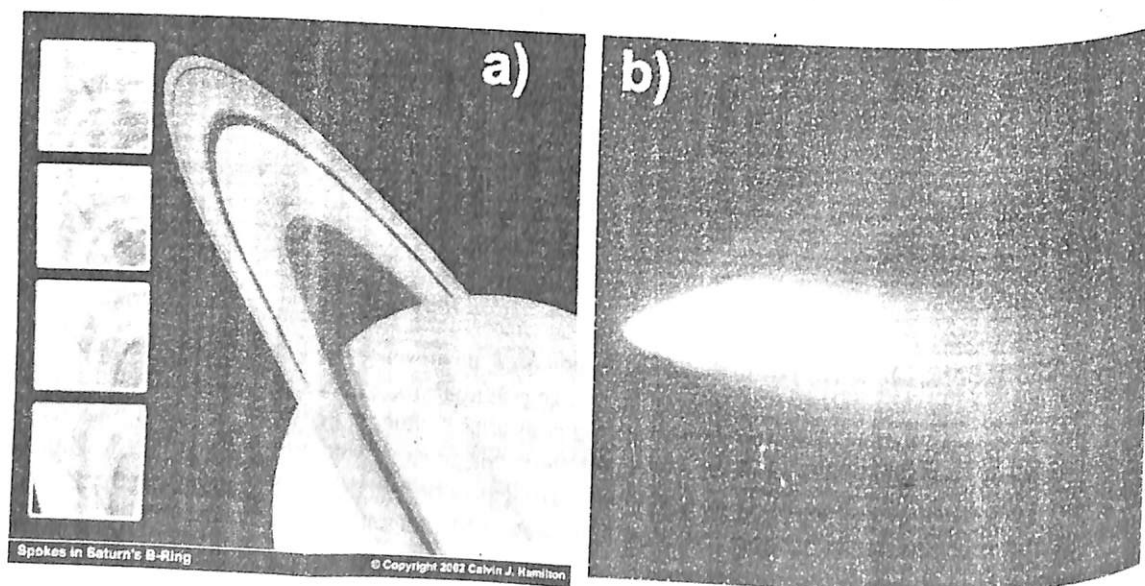


FIG 4: Dusty Plasma : a) Spokes formation in the rings of Saturn as a radially extended dark features observed by Voyager 2 in 1980. b) Plasma and dust tail of comet Hale-Bopp (1994).

contribute to two different interaction potentials given in equation (12) and (11) respectively. This shows that the modification in dust Debye cloud is caused mainly due to ion flow speed M and background plasma response towards the grain through P and Q . From Fig. 1, it is seen that P value increases with increase in ion flow speed from subsonic regime and becomes saturated approaching the supersonic regime. This makes the repulsive Yukawa potential independent of M in supersonic regime and can be realized from equation (12).

Whereas, the Q increases linearly with ion flow speed M . As a result, the strength of both repulsive Yukawa and attractive wake potential given in equation (12) and (11) are reduced in presence of dynamical charging of grains as compared to that of constant charged dust grains having $P=Q=0$ and is shown in Fig. 2(a) for wake potential, with typical dust plasma parameters having electron temperature $T_e=1.7\text{eV}$, ion temperature $T_i=0.02T_e$, ion to electron density ratio $\frac{n_i}{n_e} \approx 1$ and normalised dust radius $g = r_d/\lambda_{De} < 1$. It also clears the fact that for $P=0$, case having stationary charge on dust surface, Yukawa potential is completely independent of ion flow speed M . The strength of wake potential from equation (11) shows an enhancement with decrease in flow speed M in Fig. 2(b).

It is also shown in Fig. 1 that both the electron (ion) attachment frequency P (Q) increases with increase in grain size g . This says larger grain will have smaller effective Debye length and weaker effective strength of interaction for a given ion flow speed M and is shown in Fig. 2(c) for a given $\tilde{A}$ and k , using equation (16). The corresponding transverse mode of vibration also shows a decrease in frequency of vibration with increase in grain size and is visible from Fig. 2(d). This transition in coupling strength with grain size finally assures a transition in crystal phase, which is of recent scientific interest in metal nano crystals [18].

To get an estimation of the relative strength of Wake and Yukawa potential in both sub and supersonic regime, we defined a dimension less ratio $R = \frac{K_{eff}}{K_y}$, where K_{eff} is the combined Strength of Yukawa and Wake potential from equation (16) and K_y is the coupling constant for Yukawa potential from equation (14). The behaviour of relative strength of interaction R with subsonic ion flow speed M is plotted in Fig. 3(a) and is as follows

Subsonic regime ($0.1 < M < 0.7$): The ratio R shows a fluctuation in strength with decrease in M value and is shown in Fig. 3(a). It is clear from Fig. 2(b), that particle-wake interaction is dominant in subsonic regime of plasma flow and is also reported in recent experiment [23]. However, the effective confinement strength of particle along vertical direction in presence of both Yukawa and Wake undergo a transition from stronger to weaker and again back to stronger confinement as a characteristic of oscillatory confinement frequency of wake potential with decrease in ion flow speed M . At $M=0.249$, the ratio R , shows an enhancement. It implies that the effective interaction is repulsive in nature. On decrease of M at 0.18, the effective coupling is relatively weaker and the interaction becomes attractive in nature. The corresponding mode of vibration also decreases and is shown in Fig. 3(b). On further decrease, at $M=0.141$, the ratio R again shows an increment. As a result, the mode increases once again. The behaviour may be realized as such that on gradual decrease in flow speed M , the attractive wake potential of upstream particle starts to dominate and align the

lower dust particle along vertical direction of ion flow. On further decrease in M , when the two particle becomes very close to each other, the repulsive Yukawa dominates over there and the particle gets back to its original configuration. The presence of this mixed phase of wake and Yukawa in subsonic regime show an anomalous transition [16] in vibrational mode for given ion flow speeds and is shown in fig. 3(b).

Supersonic regime ($1.0 < M < 5.0$): In this regime the strength of wake potential is significantly reduced and it makes this region highly Yukawa dominating phase of interaction. It can be seen from Fig. 3(c), that the ratio R asymptotically approach 1, with increase in ion flow speed M . This is due to the fact that electron attachment frequency P

becomes independent of M in supersonic regime and is discussed above. This leads to the strength of Yukawa potential independent of M , for higher ion flow speed.

4. Conclusion

In this letter, we have presented an elaborate discussion on the formation of wake potential along with Yukawa potential in complex plasma and discussed the strength of interaction for a vertically placed dust subsystem, as a characteristic of ion flow speed and grain size. The observations can be summarized as follows:

- The effective plasma Debye length surrounding the dust gets distorted and reduced in presence of dust charge fluctuation in terms of ion flow speed and particle size.
- The strength of both Yukawa and wake potential is reduced with increase in grain size, in terms of electron (ion) attachment frequency $P(Q)$. The corresponding mode of vibration also decreases with increase in grain size. This implies a size dependent transition in crystal phase often observed in nano crystal.
- The type of interaction is found to be in a mixed phase of Wake and Yukawa in subsonic regime of plasma flow. In this regime the frequency of vibrational mode shows an anomaly with gradual decrease in ion flow speed M . This implies an ion flow induced anomalous transition in crystal phase, a well known phenomena in solid state physics [17]. The supersonic regime is found to be in Yukawa dominating phase.
- The study is expected to play a significant role in controlling the phase behaviour and diffusion phenomena in complex plasma in absence of magnetic field.

References

- [1] H. Fechting, E. Grun and G. E. Mor II, Planet. Space Sci. 27, 511 (1979).
- [2] O. Havnes, C. K. Goertz, G. E. Mor II, E. Grun, and W. Ip, J. Geophys. Res. 92, 2281 (1987).
- [3] S. Baruah and N. Das. Phys. Plasmas 17, 073702 (2010).
- [4] J.H. Chu and I. Lin, Phys. Rev. Lett. 72, 4009 (1994).
- [5] U. Konopka and G.E. Mor II and L. Ratke, Phys. Rev. Lett. 84, 891 (2000).

- [6] H. Thomas, G.E. Mor II, V. Demmel, J. Goree, B. Feuerbacher and D. Mnohlmann, Phys. Rev. Lett. 73, 652 (1994).
- [7] A. Melzer, V.S. Schweigert and A. Piel, Physica Scripta 61, 494 (2000).
- [8] Jie Kong, Truell W. Hyde, Lorin Matthews, Ke Qiao, Zhuanhao Zhang, and Angela Douglass, Phys. Rev. E 84, 016411 (2011).
- [9] C. Chia-Ling, I. Chong-Wai, and I. Lin, Contrib. Plasma Phys. 49, 215 (2009).
- [10] V. Steinberg, R. Sutterlin, A. V. Ivlev, and G. Mor II, Phys. Rev. Lett. 86, 4540 (2001).
- [11] N.N. Rao, P.K. Shukla and M.Y. Yu, Planet. Space Sci. 38, 4 (1990).
- [12] D.S. Lemons, M.S. Murillo, W. Daughton and D. Winske, Phys. Plasmas 7, 6 (2000).
- [13] O. S. Vaulina, S. A. Khrapak, A. P. Nefedov, and O. F. Petrov, Phys. Rev. E, 60, 5959 (1999)
- [14] N. Das and K. S. Goswami, Phys. Plasmas, 5, 312 (1998).
- [15] F. Melands, T. Aslaksen and O. Havnes, Planet. Space Sci., 41, 321 (1993).
- [16] S. Bhattacharjee and N. Das, Phys. Rev. E, 88, 043106 (2013).
- [17] Yang-Bo Zhou, Zhi-Min Liao, Yi-Fan Wang, Georg S. Duesberg, Jun Xu, Qiang Fu, Xiao-Song Wu, and Da-Peng Yu, J. Chem. Phys. 133, 234703 (2010)
- [18] R. Bardhan, L. O. Hedges, C. L. Pint, A. Javey, S. Whitlam and J.J. Urban, Nature Mater. 12, 905 (2013).
- [19] D. Szwarcman, D. Vestler, and G. Markovich, Acs Nano, 5, 507 (2011).
- [20] V.E. Fortov, A.V. Ivlev, S.A. Khrapak, A.G. Khrapak, G.E. Mor II, Physics Reports, 421, 1 (2005).
- [21] J. E. Allen, Physica Scripta., 45, 497 (1992).
- [22] S. Bhattacharjee and N. Das, Phys. Plasmas 20, 113701 (2013).
- [23] O. Arp, J. Goree, and A. Piel, Phys. Rev. E 85 046409 (2012).

A discussion on nonlinear energy exchange process through kinetic theory in Vlasov Ionospheric Plasma

Satyajyoti Gogoi

Department of Physics,
Tinsukia College, Tinsukia, India
Email: satyajyotitsk@gmail.com

Abstract

In this review article we present a brief qualitative deliberation on nonlinear wave-particle interaction energy exchange process in ionospheric plasma medium through nonlinear kinetic theory.

1. Introduction

The quasi neutral plasma is one of the thermodynamical states of matter. This phase of matter consists of negatively charged electrons, ionized atoms or molecules and neutral particles. It is considered 99% of the visible matter in the universe is composed of plasma. Plasma exists in both natural and laboratory environment. In nature, plasma is found in interstellar space, within the star, solar corona, x-ray are, solar wind, plasma tails of magnetospheres, lightning and even in planetary ionospheres.

In plasma state self-consistent electric and electromagnetic fields exist and corresponding electric and electromagnetic forces are effective. For long range charged particles interacting forces, plasma exhibits universal collective behaviours which is absent in solid and ordinary liquid states. Different physical parameters like density, temperature, pressure, external electric and magnetic field may perturb basic properties of this fourth state of matter. Interaction among plasma particles may develop oscillations and instabilities. These plasma waves and instabilities can modify behaviours of this both fully and partially ionized form of electrically

S.J.Gogoi, A discussion on nonlinear energy exchange process through kinetic theory in Vlasov Ionospheric Plasma. Horizon(Ed. Book), 142-152(2023). copyright: Tinsukia College Publishing. ISBN: 978-81-19322-07-7

conductive medium. Photoionization may be the possible dominant process for ionization in planetary atmospheric plasma. Solar wind and interstellar ultra violet radiations are the sources of energy and momenta to this wall-less weakly ionized gas. In atmospheric plasma large numbers of interesting phenomena are observed from ground based observatories and from artificial satellites. In presence of electrostatic and electromagnetic plasma waves along with plasma particles, their interactions may give some complex exotic nonlinear effects.

To understand in the details of near-Earths space environment, it is essential to investigate these phenomena through different theoretical approaches for better comprehension on planetary space environment and the universe as a whole.

2. Characterizations of Individual and Collective Phenomena of Plasmas and Graininess parameter

If inter-plasma particle distance is smaller than Debye length, these particles experience a large angle Coulomb collision which implies discrete or individual character of particles dominant over the situation. On the other hand collective character of plasma is prevail when inter-plasma particle distance is above Debye length and Debye screening is prevail under this condition.

In plasma, individual phenomena are characterised by mean free path (l_{mfp}) and collision frequency (ν_c), and collective phenomena are characterised by Debye length (λ_D) and plasma frequency (ω_p). Based on a thought experiment which was proposed by Rostoker and Rosenbluth in 1960 [1], collective characteristics of plasmas remain invariant, i.e., λ_D^2 and ω_p^2 are constants while individual characteristics of plasma are modified as both mean free path l_{mfp} and collision frequency (ν_c) approaches to zero.

The value of dimensionless graininess parameter (g) distinguishes individual and collective characters of plasma. As graininess parameter (g) approaches to zero, the collective effects are dominating the whole system and the system can consider as collisionless and Vlasov description requires for it [2].

3. Planetary Ionospheres and Ionospheric Plasma

Initially, research on ionospheric physics was focussed on origin of the ionosphere, its different layers, variability of its physical parameters with local time, latitude, seasons and on radio wave propagation. For development of experimental based observatories and theoretical frameworks, ionospheric research is mainly concern on dynamics of this region and on ionospheric plasma physics [3].

In 1901 after successfully transmitted radio signals across the Atlantic, G.Marconi experimentally established existence of electrically conductive layer in the upper terrestrial atmosphere and in 1902 a theoretical model on this electrically ionized media was proposed

by O. Lodge [45]. The name Ionosphere was proposed by R.A. Watson-Watt for this atmospheric conducting layer in a letter to the United Kingdom Radio Research Board [5].

By using ground based and with rocket-and satellite based observatories combined with remote sensing data, pictures of planetary ionosphere were obtained for different aspects. From the Pioneer -10, -11, the Voyager- 1; Voyager- 2, Galileo and Cassini spacecraft mission, ionospheres data of outer planets like Jupiter, Saturn, Uranus, Neptune, Saturn's moon Titan and Neptune's moon Triton were recorded by using radio occultation measurements and other spectrometric techniques.

From the Mariner-10 spacecraft optical observations it was found that inner planet Mercury did not have a conventional ionosphere. In situ measurements from Venera-4 spacecraft, S-band radio occultation measurements from Mariner-5 spacecraft and other mass spectrometric measurements revealed that inner planet Venus had a well-developed ionosphere during dayside and nightside duration. With the help of Retarding Potential Analyzers (RPA) of Viking-1, Viking-2, Phobos Spacecraft, Mars Express and Radio science (RS) experiment of Mars Global Surveyor (MGS), suggest that ionized environment of upper atmosphere of Mars has many similarities to that of inner planet Venus [6,7,8,9].

For presence of a neutral atmosphere and sources of ionizations, upper atmosphere of Earth has a weakly ionized ionospheric zone. Range of ionospheric altitude of the Earth starts from about 50 km to 1000 km. Characteristics of this region are known by reflection and refraction properties of electromagnetic wave. This region acts as interface in between the lower atmosphere and space of the region of upper atmosphere. Major ingredients of this radio wave reflected region are electrons, ions and few neutral particles. Photo ionization and impact ionization are major processes to produce in this region. Solar radiation and a small part from Bremsstrahlung radiation are main sources of energy and momenta for photo ionization. In daytime, Earth's ionosphere ionization occurs mainly from solar photons of extreme ultraviolet and ultraviolet wavelength range. Additionally energetic precipitating charged particle fluxes which are coming from the sun, the galaxy, the magnetosphere or from local accelerated charged particles within the ionosphere itself taking part to produce ionization by coulomb collisions in this upper atmospheric region. Behaviour of this partially ionized open system is controlled by various processes and external entities like diffusions, chemical reactions, turbulences, instabilities, gradients of pressure, temperature, density, confining fields like electric and magnetic fields, on events which are occurring in the sun, movements of neutral wind in upper atmosphere due to Lunar Orientation Movements (ILOM) and so on [10,11,12].

Earth's ionospheric region can be characterised as a weakly collisional or collisionless, low plasma beta and cold plasma. Long term stability, lack of confinement walls, constant circulating around the earth and wide region of parameter space are few major characteristics of this ionospheric plasma. Interactions among electrostatic and electromagnetic resonant and non-resonant modes and with plasma particles may give different types of instabilities in this partially ionized region [7, 13].

4. Plasma Waves in the Ionosphere

A variety of plasma waves are prevalent throughout the nonuniform ionospheres in presence of external fields. These plasma waves can perturb physical properties in the ionosphere. Plasma waves are electrostatic and electromagnetic in nature. Electrostatic plasma waves can have fluctuating electric fields and electromagnetic plasma waves can have both fluctuating electric and magnetic fields. Depending upon the circumstances, plasma waves can have amplitudes either small or large. These plasma waves can propagate in both unmagnetised and magnetised plasmas.

In the longitudinal mode, fluctuating electric fields are parallel to the propagation vector, whereas for a transverse mode, fluctuating electric fields are perpendicular to the propagation vector. In a magnetized plasma, plasma waves can propagate along or perpendicular or at an arbitrary angle to the direction of magnetic field.

To study propagation of electrostatic plasma waves, only Gauss law and the Poisson equation are relevant and electromagnetic plasma waves are governed by Maxwell's field equations. For electrostatic waves, plasma is affected due to perturbed charge density and for electromagnetic waves, plasma is affected through both the charge density and the current density.

5. Plasma Instabilities in Ionosphere

Ionospheric plasma can be considered a natural laboratory set up for theoretical and experimental investigations on wave-wave and wave-particle interactions on various observational phenomena and instabilities to gain better understanding of the behaviour of upper atmosphere. Most of the information on various processes in the ionosphere obtains from radio wave scattering or reflection experiments which are conducted from the surface of the earth along with through space based rocket and satellite-borne sensors.

From theoretical point of view in some problems, ionospheric plasma can consider as either homogeneous or inhomogeneous medium. If plasma parameters vary little over a wavelength, an ionospheric plasma is considered as homogeneous but this approximations are rarely materialized. On the other hand, due to high variations in densities, gradients in temperature and magnetic fields, corresponding ionospheric plasmas to be considered as inhomogeneous. Using the homogeneous plasma theories, it becomes very difficult to solve inhomogeneous plasma problems and for this, appropriate nonlinear theory is must.

There are diverse free energy sources in the Earth's atmosphere. For presence of input energy and momentum fluxes from these sources, spatial gradients of physical parameters and inhomogeneities develops, along with this plasma system become unstable if a small perturbation from the equilibrium state results in a growth of the perturbation. For this accumulated energy, instability is the general way of redistributing available free energy. Plasma instabilities can be treated as plasma waves which grow in amplitudes. Instabilities lead to transitions of a nonequilibrium state from some kind of one state and turbulence [14,15].

6. Nonlinearity and Turbulence in Plasma

As a physical system deviates weakly from its equilibrium phase with respect to space and time, it leaves linear behaviours and becomes nonlinear system. In general, a plasma is fundamentally a nonlinear medium for unstable collective modes and for energy exchange among plasma waves and particles. For fluctuations of physical parameters which are normally amplified with respect to time for instabilities, then a plasma system is in a turbulent state. Depending upon perturbation from the zeroth-order distribution function, both weak turbulence and strong turbulence are observed in nonlinear plasma [16,17,18]. In analytical treatment, a nonlinear plasma system is described by nonlinear partial differential equations and a set of algebraic expressions are obtained by using either Fourier transforms or Laplace transforms and obtain dispersion relations. To investigate theoretically on nonlinear effects, few analytical methods are used which are based on approximations but absence of a general algorithms one of its major limitations.

7. Turbulence Wave-Particle Interaction Theory

In plasma turbulence theory roles of coupling process between non-resonant and resonant modes of fluctuations are important to sustain the stationary turbulent state. In the nonlinear regime, some collective modes and plasma turbulence saturation will occur at a later stage for nonlinear interactions. To describe such interactions analytically, Vlasov model is used [15,16,19].

There are three types of interactions are involved in weak turbulence theory for different and distinct physical processes:

(i) Nonlinear wave-particle interaction:

Energy exchange process takes place thorough this nonlinear interaction between waves and resonant particles which are in same phase and phenomena like growth or damping of waves, instabilities, plateau or tail formation for perturbation of distribution function can occur.

(ii) Nonlinear wave-wave interaction:

With the help of this nonlinear interaction, it is possible to explain on diffusion of wave energy among many modes, parametric instability and dielectric properties.

(iii) Nonlinear wave-particle-wave interaction:

In this interaction, particles are in constant phase relative to the beat frequency of the two waves. Nonlinear Landau damping or growth, echoes, wave scattering by particle distribution etc. can explain with this kinetic interaction.

The theory of weak turbulence, also known as quasilinear theory, deals with large number of plasma waves of finite amplitudes through a statistical approach to investigate nonlinear effects in inhomogeneous plasma. This quasilinear theory mainly deals with nonlinear wave-particle interaction process. According to this theory, particle distribution is weakly affected by wave fields and this effect may self-consistently quench the instability.

Perturbation of distribution function due to this weak instability tends to grow and at large amplitudes, nonlinear effects are arise and influence on growth rate.

8. Vlasov Model

To study collective behaviour and instabilities of collisionless plasma in velocity space regime, Vlasov equation was introduced by A. Vlasov in 1945 and L.D. Landau in 1946. Not only in plasma dynamics, this Vlasov model is applicable in different fields like from semiconductors and stellar dynamics problems to quantum mechanical problems.

The Vlasov equation is valid and invariant in a plasma at the limit graininess parameter (g) approaches to zero where distribution function and electron charge density are invariant and individual collisions are negligible [2, 20].

10. Mechanism of Wave-Particle Interaction in Vlasov Plasma

In both open and closed plasma environment, nonlinear phenomena are examined through wave-particle interactions. In space plasma, some observed phenomena like formation of the magnetopause boundary layer, electro-magnetic outer zone chorus, plasmaspheric hiss emissions, interaction of plasma waves with Van Allen belt particles, ionospheric phenomena like diffuse aurora etc. are explained effectively by this interaction process [15].

In collisionless plasma, wave energy dissipation or upconversion takes place in absence of particle collisions and transfer of energy and momenta occurs in between particle velocity and plasma waves electric field. At resonance, this interaction become strong enough where waves are Doppler shifted at cyclotron frequency. Electric field of waves is not rapidly fluctuating at that moment.

In Maxwellian distribution, plasma particles having smaller velocities than phase velocity of waves and they are accelerated by waves electric field, i.e. has acquired energy from waves and in consequence, plasma waves suffer damping, known as Landau damping.

For non-Maxwellian distribution, in a given velocity interval, number of faster particles are greater than of slower particles and waves are gaining energy through these faster particles which corresponds to inverse Landau damping.

11. Maser Effect and its Physical Mechanism

In 1976 Ronmark and Steno [21] had observed that the theory of plasma turbulence could applied to the generation mechanism of high frequency plasma waves in presence of low frequency electrostatic turbulence through plasma particles and turbulent waves interaction.

Theoretical investigation on enhanced electromagnetic wave generation in an unmagnetised plasma, due to acceleration of plasma particles by the electric field of low frequency ion acoustic waves, was carried on by Nambu and Shukla [22,23].

To describe this fully developed turbulent plasma state usually weak turbulence theory was used. In previous studies [24,25,26,27,28,29], phenomena like anomalous resistivity, plasma maser, induced scattering of waves etc. were explained through nonlinear interaction between low frequency turbulence and electromagnetic waves.

Kadomstev [30] and Nambu [31] had studied on nonlinear interactions of waves through weak wave turbulence and this study had formed the foundation of plasma maser effect. From this investigation it has established that resonant electrons whose velocities are nearly equal to the phase velocities of low frequency plasma waves, are experienced a strong acceleration due to electric field of resonant waves and these accelerated electrons could generate induced bremsstrahlung radiation.

At initial stage a number of controversies arise on wave energy upconversion of Langmuir wave in presence of low frequency ion sound wave. Based on quasilinear theory with previous theoretical and experimental results, Nambu [32] had introduced the theory of Plasma maser instability which was previously known as physics of turbulent bremsstrahlung instability, in a systematic way for explanation on wave energy enhancement of high frequency plasma waves.

A nonlinear electric field was developed for coupling between resonant and nonresonant modes in an ionized medium and due to interaction between that nonlinear field and electrons caused to turbulent bremsstrahlung instability. Therefore, in plasma maser effect, which is a nonlinear kinetic interaction, energy exchange process takes place between a resonant mode and a non-resonant mode in presence of resonant electrons and the non-resonant mode gains energy.

13. The Effective Non-linear Force of Plasma Maser Effect

In a plasma system due to interaction between non-resonant and resonant plasma waves can derive a high frequency nonlinear modulation field. An effective nonlinear force which arises from this nonlinear modulation field can take part in wave energy upconversion of a high frequency non-resonant wave. For this high frequency nonlinear force, electrons which are in phase with low frequency resonant wave are accelerated. According to electrodynamics, these accelerated electrons can emit high frequency waves cause to grow unsteady high frequency waves. Therefore wave energy may transfers from resonant waves to non resonant waves through nonlinear wave-particle interactions.

Under the framework of high frequency nonlinear force, different phenomena in both astrophysical plasma and fusion plasma are explained. The effect of nonlinear forces in plasma instability was explained by Prof.M.Nambu [32].

15. Plasma maser effect in Space plasma

For interpretation of various fusion and astrophysical phenomena, different linear and nonlinear plasma theories are used for further progress in respective fields. For explanation of radiation phenomena and understanding of space environment, nonlinear plasma theories are highly relevant.

The nonlinear plasma maser effect describes wave energy upconversion of high frequency non-resonant waves in presence of low frequency resonant waves through nonlinear wave-particle interactions. From experimental observations, it has found that in space plasma, most

of the turbulence energy in the form of low frequency turbulences like drift waves, lower hybrid waves, ion sound waves, Alfvén waves etc. In different types of ionospheric and magnetospheric radiation phenomena the role of low frequency resonant waves are important. These radiation mechanisms may explain through plasma maser instability.

16. Plasma maser effect in Ionospheric plasma

To explain different types of radiation emission phenomena and plasma waves instabilities which are observing at different altitudes in terrestrial ionospheric plasma, are attempting theoretically by both linear and nonlinear methods. In nonlinear theoretical model, nonlinear mode-mode couplings are dominating interactions in wave - energy exchange process for plasma waves instabilities. To investigate energy enhancement of high frequency plasma waves, plasma maser instability is one of the prominent nonlinear wave-particle interaction energy transfer modes in both astrophysical and tokamak plasma environment. In our study, we present a briefer review on probable growth rate of high frequency plasma waves at some regions of the Earth's ionosphere through plasma maser instability.

Gogoi and Deka [33] had studied on probable amplification of electromagnetic O-mode in presence of low frequency ion sound wave turbulence by plasma maser effect in the ionosphere regime. For this problem, spatial in homogeneity of density was taken in nonuniform Earth's ionosphere. After lengthy and complex calculations, the approximate expression of growth rate for electromagnetic O-mode was found [33]. With the help of observational data on the plasma parameters, O-mode plasma wave characteristics in space, ion sound wave and ionospheric parameters from ROSE satellite observations and other sources [34, 35, 36, 37], the approximate growth rate for O-mode from the equation was found as 10^{-5} . In this investigation it had been established that high frequency O-mode was amplified in presence of ion sound wave in ionospheric plasma due to wave-particle interaction through plasma maser effect.

In another study [38] it was examined wave energy upconversion of electrostatic nonresonant lower hybrid wave through plasma maser instability in the mid-altitude Earth's ionosphere in where both gradients in density and magnetic field were taken in presence of drift wave turbulence. Using observational data of lower hybrid waves, drift waves and plasma parameters in space [39, 40, 41, 42] and considering density gradient as 0.1 in the derived equation [38], the probable growth rate of lower hybrid wave was found as 10^{-5} .

In another problem [43] probable amplification of electrostatic Langmuir wave was studied in presence of lower hybrid waves turbulence in polar ionospheric plasma, neglecting gradients parameters. Using observational data from Freja satellite observations at topside polar ionospheric regions and from other sources at the Earth's near space region [44, 45, 46], the probable growth rate was found as 10^{-2} which indicates wave energy enhancement of Langmuir wave.

In brief the major findings in these studies are plasma maser effect has demonstrated a decisive role in amplification or wave energy up conversion of nonresonant electromagnetic and electrostatic waves in presence of resonant waves in collisionless ionospheric plasma. It

has been found that electromagnetic O-mode is amplified in presence of low frequency ion sound wave where the external magnetic field acts as momentum source. At mid-altitude ionosphere region, electrons which are accelerated by resonant mode drift wave turbulent field transfer its energy to high frequency lower hybrid wave through a modulated field nonlinearly. Also due to wave-particle interaction a high frequency dissipative nonlinear force is developed which acts as driving force for the growth of non resonant mode. At polar ionosphere, it has been found the possibility of wave energy enhancement of Langmuir wave at the expense of low frequency lower hybrid wave turbulence energy under considering only such parameters which influence may be significant to the result of the estimate of growth rate. There is other subdominant processes exist in such situations but these effects are not included in our study.

17. Analytical Treatment Constraints

To determine fluctuating parts of a distribution function in presence of modulating and modulated fields due to nonlinear modes-modes interactions, and to derive dispersion relation and growth rate, usually Method of Characteristics (also known as integration along unperturbed orbit)[16], Fourier transforms, Complex residue integration, Bessels recursion relations, Dirac Delta functions, Random phase approximations and Causality principles are applying. But the major disadvantage of these complex & complicated and very lengthy calculations is that exact solutions are not obtained due to nonlinearity and non-locality of the plasma medium.

18. Conclusion

Nonlinear wave-particle energy exchange process is one of the forefront research area in plasma dynamics at present. Its important implications are investigated in astrophysical and in confined laboratory like tokamak plasmas by using observational data. In open nonlinear plasma, various forms of radiation mechanisms are analyzed and characterized by nonlinear theoretical models, weak turbulence theory is one of these. On the basis of quasilinear theory, nonlinear energy exchange between resonant and non-resonant modes can explain effectively through plasma maser effect in an open system where free energy sources are available. Various low frequency turbulences are active here in presence of inhomogeneity like density gradients, temperature gradients, magnetic field gradients etc. for supply of energy and momenta from free energy sources.

Earth's ionosphere is an open system in where exotic modes-modes mechanisms are registered due to entry of input energy and momenta through primary solar radiation and other external agents. From study on these nonlinear phenomena, useful information can assemble and helpful for further improvement on knowledge of near-Earth space environment. Based on plasma maser effect, such issues have been studied thoroughly in the magnetosphere plasma processes and satisfactorily gave interpretation of many radiation phenomena. In view of complex nonlinear waves interaction phenomena, there is a need to apply the existing nonlinear mode-mode coupling theories, such as plasma maser effect, in the frame work of ionospheric plasma regime.

References

- [1] N.Rostoker, and M. Rosenbluth, Test particles in a completely ionized plasma, *Phys.Fluids*, 3, 1, 1960.
- [2] P.Bertrand, D.D.Sarto and A.Ghizzo, *The Vlasov Equation 1, History and general Properties*, Wiley, ISBN 978-1-78630-261-8, 2019.
- [3] M.C. Kelley, *The Earths Ionosphere; Plasma Physics and Electrodynamics*, Academic Press, Inc. ISBN: 9780120884254, 2009.
- [4] O. Lodge., Mr. Marconis results in day and night wireless telegraphy, *Nature*, 66, 222, 1902.
- [5] R.A. Watson-Watt., *Weather and Wireless*, Q. J. Roy. Meteorol. Soc., 55, 273, 1929.
- [6] B.M. Smirnov, *Physics of Ionized Gases*; Wiley-VCH Verlag GmbH Co. KGaA, Weinheim, ISBN: 0-471-17594-3, 2004.
- [7] R. Schunk and A. Nagy, *Ionospheres: Physics, Plasma Physics, and Chemistry*; Cambridge University Press, ISBN: 978-0-521-87706-0, 2009.
- [8] R.P. Singhal, *Elements of Space Physics*, PHI Learning Private Limited, ISBN: 978-81-203-3710-7, 2009.
- [9] Ed. by M.G. Kivelson and C. T. Russell, *Introduction to Space Physics*, Cambridge University Press, ISBN: 0-521-45104-3, 1996.
- [10] F.W. Perkins., *space plasma physics: The study of solar-system plasmas*, National academy of sciences, 1978.
- [11] B. Zolesi, L.R. Cander., *Ionospheric Predictions and Forecasting*, Springer Geophysics, ISBN 978-3-642-38429-5, DOI 10.1007/978-3-642-38430-1, 2014.
- [12] N. Blaunstein, E. Plohotniuc., *Ionosphere and applied aspects of Radio communication and radar*, CRC press, Taylor and Francis Group, ISBN 9781138372641, 2008.
- [13] M. Takashi, *Ionosphere and Thermosphere*, *Journal of the communications research laboratory*, 49, 3, 163-179, 2002.
- [14] M.A. Uman, *Introduction to Plasma Physics*, McGraw-Hill Book Company, Library of Congress Catalog Card Number 63 21546, 1964.
- [15] R.A. Treumann and W. Baumjohann, *Advanced Space Plasma Physics*, Imperial College Press, ISBN: 978-1860940262, 2001.
- [16] N.A. Krall and A. W. Trivelpiece, *Principles of Plasma Physics*, McGraw-Hill Book Company, ISBN: 978-0070353466, 1973.
- [17] S.P. Gary, *Theory of Space Plasma Microinstabilities*, Cambridge University Press, ISBN: 0-521-43748-2, 2005.
- [18] S. Ichimaru, *Statistical Plasma Physics, Vol: 1, Basic Principles*, Taylor and Francis, ISBN 13: 978-0-8133-4178-1, 2004.
- [19] D.G. Swanson, *Plasma Waves*, Academic Press, ISBN: 978-0124122505, 2012.

- [20] D.G.Swanson, Plasma Kinetic Theory, CRC press, Taylor Francis Group, ISBN 13 978 0 4200 7581 6, 2008.
- [21] K.Ronmark and L.Steno, Planetary Space Science, 24, 904, 1976.
- [22] M.Nambu and P.K.Shukla, Physical Review A. 20, 2498, 1979.
- [23] M.Nambu and P.K.Shukla, Physical Review A. 22, 787, 1980.
- [24] J.M.Dawson, Advances in Plasma Physics, Vol. 1, 2, 3. Interscience, New York, 1976.
- [25] K.H.Spatcheck, P.K.Shukla, and M.Y.Yu, Plasma Physics, 17, 1091, 1975.
- [26] A.T.Lin, P.K.Kaw, and J.M.Dawson, Physical Review A. 8, 2618, 1973.
- [27] V.N.Tsytoich and L.Steno. Physics Letter A. 43, 7, 1973.
- [28] M.Nambu, Physics Fluids, 18, 412, 1976.
- [29] M.Nambu, Physics Fluids, 20, 459, 1977.
- [30] B.B.Kadomstev. Plasma Turbulence, academic, New York, 1965.
- [31] M.Nambu, J. Phys. Soc. Japan, 53, 1594, 1983.
- [32] M.Nambu, Laser Part. Beams, 1, 427, 1983.
- [33] Gogoi SJ, Deka PN. Physical Science International Journal. 2019;23(3):1-10
- [34] Benson RF, Mellott MM, Huffand RL, Gurnett DA. J. Geophys. Res. 1988;93:7515-7520.
- [35] Dyrud L, Krane B, Oppenheim M, Pecseli HL, Schlegel HK, Trulsen J, Wernik AW, Ann. Geophys. 2006;24:2959-2979.
- [36] Malingre M, Pottellette R, Dubouloz N, Lindqvist PA, Holmgren G, Aparicio B. Geophys. Res. Lett. 1992;19:1339-42.
- [37] Gurnett DA, Marsch E, Pilippand W, Schwenn R. Journal of Geophysical Research. 1979;84(A5):20299-2038.
- [38] Gogoi SJ, Deka PN. Journal of scientific research. 2019;11(3):339-350.
- [39] Jackel S, Eliezer S, Zigler A. Phys. Rev. A. 1981;24:1601.
- [40] Malingre M, Pottellette R, Dobouloz N, Lindqvist PA, Holmgren G, Apricio B. Geophys. Res. Lett. 1992;19:1339-42.
- [41] Gurnett DA, Frank LA, J. Geophys. Res. 1977;82:1031-1050.
- [42] Horton W. Reviews of Modern Physics. 1999;47:735-778.
- [43] Gogoi SJ, Deka PN. International Journal of Scientific Research in Physics and Applied Sciences. 2019;7(3):176-181.
- [44] Lizunov GV, Khotyantsev Y, Stasiewicz K.J. of Geophys. Res. 2001;106(A11;24):755-24; 763.
- [45] Deka PN. Ph.D. Thesis, Gauhati University Assam India; 1999.
- [46] Malingre M, Pottellette R, Dobouloz N, Lindqvist PA, Holmgren G, Apricio B. Geophys. Res. Lett. 1992;19:1339-42

Nurseries of Celestial Light: Exploring Star Formation in Young Open Clusters

Arshad Hussain

Department of Physics,
Dibrugarh University, Dibrugarh, India
Email: arshad007h@gmail.com

Abstract

Open clusters, although widely recognized celestial objects, still hold many mysteries waiting to be unravelled. These clusters offer exceptional opportunities to conduct astrophysical experiments and validate theories related to stellar evolution and star formation. In this present study, Johnson-Cousin's optical photometry (B, V) and near infrared photometric (JHK_s) data from 2MASS have been used to study the young open cluster NGC 3324 situated in the Carina Nebula in the Carina-Sagittarius spiral arm of our Milky Way galaxy. Nearly 1000 optical and 2591 near infrared (NIR) sources were selected and their color-color and color-magnitude diagrams were studied. Using theoretical age isochrones, the age of the cluster was estimated (~7 Myr). The 2MASS data were converted to Koornneef's system and plotted the J-H, H-K color-color (CC) diagram. Using Koornneef's sun reddened main sequence and giant stars, young NIR excess members within the cluster have been identified which is a strong indication that star formation may still be going on in the cluster.

1 Introduction

1.1 Star Clusters

Star clusters have been in the human eye since ancient times. They continue to be a fascinating subject with surprising new observations. They are remarkable observatories where fundamental astrophysical phenomena covering the whole star life cycle, from creation to

A. Hussain, Nurseries of Celestial Light: Exploring Star Formation in Young Open Clusters, Horizon (Ed. Book), 153-169 (2023), copyright: Tinsukia College Publishing, ISBN: 978-81-19322-07-7

death, may be studied. It has been recognized that stars are not distributed uniformly throughout the sky. They are gregarious things. They typically exhibit a social inclination and tend to congregate in groups and clusters instead of isolation. The first issue that any discussion about star clusters must deal with is how to define a cluster and distinguish it. *Zwart, McMillan & Gieles (2010)* defined a star cluster as a set of stars that are gravitationally bound to one another. According to *Krumholz et al. (2019)*, a star cluster is a gravitationally bound group of stars inside a closed tidal surface if this volume is not dark matter-dominated and contains at least 12 stars. The first condition sets star clusters apart from galaxies. The latter distinguishes it from multiple stellar systems. These stellar clusters encompass a vast age spectrum, ranging from youthful clusters (< 3 million years old) embedded in gaseous clouds to ancient clusters (> 13 billion years old), making them valuable indicators of star and galaxy formation and evolution. We can learn more about the historical evolution of the host galaxies of star cluster systems by evaluating their properties, which will help us better comprehend galactic origin and evolution.

Stars are formed in cold, thick molecular clouds that have enough gas in them to support several stars in clusters or groups. These clouds are the product of vast gas reservoirs. The stars are attracted to one another by gravitational forces, which cause the development of gravitationally bonded star clusters once any leftover gas has been heated and dispersed. It is accepted that the majority of stars do not develop independently but rather within massive star clusters before spreading out over the cosmos.

1.2 Classification

Star clusters can be classified into two main types: globular clusters and open clusters. Globular clusters, which constitute comparatively ancient assemblages of stars, were formed during the early stages of galaxy formation. These clusters often contain hundreds to millions of aged stars arranged in a tightly compacted configuration. They can be found in both the galactic disk and the halo. The term "globular" was used due to their nearly spherical symmetry. The stars within globular clusters are the most ancient inhabitants of the galaxy, consisting primarily of Population II stars, rendering them almost as old as the galaxy itself. Their age is evidenced by their scarcity of heavy elements (referred to as metals in astronomy, denoting elements beyond hydrogen and helium). These heavier elements were synthesized within stellar cores and subsequently dispersed throughout the interstellar medium through supernova explosions.

In contrast, open clusters exhibit distinct differences from globular clusters. Confined within the galactic plane, open clusters are consistently situated within the spiral arms. These clusters serve as recent nurseries for the birth of new stars, emerging from clouds of gas and dust. They typically consist of a few thousand stars within a region spanning up to approximately

30 light-years, smaller in size compared to globular clusters. Open clusters encompass a wide range of ages, with some being as old as, if not older than, our Sun. The youngest open clusters still retain associations with the interstellar matter from which they originated.

1.3 The Astronomical Importance of Star Clusters

In the realm of astronomy, the measurement of distance, mass, temperature, age, and other properties of celestial objects holds great significance. Observations primarily provide information on the brightness of objects, requiring us to deduce their other properties from these measurements. Studying nearby objects and establishing their properties serves as a reference point for understanding distant objects.

From an astrophysical standpoint, star clusters play a crucial role in various domains. This is due to the fact that stars within a cluster are believed to have formed around the same time and are located at roughly the same distance from us. These stellar clusters are assumed to have originated from a common interstellar gas cloud, resulting in similar ages and chemical compositions among their member stars. The only parameter that differentiates members within a cluster is their mass. For these reasons stellar evolution theories heavily reliant on cluster observations. Young open star clusters, in particular, offer exceptional opportunities for comprehending large-scale star formation and evolutionary processes within our galaxy. Due to their youth and relatively short life spans, open clusters serve as valuable tools for studying young stars, the dynamics of star formation, and the dispersion of material throughout our galaxy.

1.4 The Structure of Open Star Clusters

Open clusters exhibit a broad range, varying from sparsely populated clusters with only a few members to extensive collections comprising thousands of stars. Typically, they feature a distinct dense core surrounded by a more diffuse "corona" of cluster members. This extended halo of genuine cluster members encompasses the central core and extends its diameter multiple times over. The core itself usually spans around 3–4 light years, while the corona reaches approximately 20 light years from the cluster center. Star densities within the core average around 1.5 stars per cubic light year.

Our galaxy hosts over 1,100 known open clusters, although the actual count may be up to ten times higher. Open clusters are predominantly situated within the spiral arms of spiral galaxies, where gas densities are highest and star formation occurs most frequently. Consequently, these clusters are primarily concentrated near the galactic plane. In irregular galaxies, open clusters may be distributed throughout the galaxy, although their concentration is highest in regions with elevated gas density. However, open clusters are not observed in elliptical galaxies due to the cessation of star formation many millions of years ago.

Open clusters encompass stars of various spectral classes, ranging from O to M, representing high to low temperatures. However, the frequency of different spectral types

varies among clusters. High-luminosity stars are more prevalent compared to the solar neighbourhood, with the brightest stars in some open clusters shining at a magnitude 150,000 times greater than the Sun. The luminosity of the brightest stars on the upper end of the main sequence within clusters ranges from approximately -8 to -2 visual magnitude.

Although stars within an open cluster originate from the same material, they do not remain gravitationally bound over time, scattering their stars far and wide. As loose systems with shallow gravitational potential, open clusters lack organized structural symmetry, making their morphology and shape susceptible to external perturbations. (Chen, Chen & Shu 2004).

The distribution of member stars within a star cluster undergoes changes as the cluster evolves. The initial arrangement of stars is determined by the structure of the parent molecular cloud and the process of star formation. As the cluster evolves, the distribution is modified by internal gravitational interactions among the member stars. Additionally, stellar evaporation and external influences, such as the Galactic tidal force, differential rotation, and encounters with molecular clouds, can alter the spatial structure and eventually dissolve the cluster. Two types of dynamic effects impact and shape the morphology of an open cluster. The first involves the internal interactions among member stars, resulting in a spherical distribution with higher density towards the cluster's center. Simultaneously, low-mass members can acquire enough kinetic energy through these encounters to be expelled from the system (known as stellar "evaporation"). The second dynamic process entails external interactions, including tidal forces exerted by the Galactic disk or giant molecular clouds, as well as differential rotation, especially in clusters located within the inner disk region. These disruptive effects tend to elongate the shape of a star cluster, with the outer regions being particularly vulnerable.

1.5 The Hertzsprung - Russell diagram

The *Hertzsprung-Russell* (HR) diagram serves as a crucial tool for comprehending the evolution of stars with different masses, luminosities, and spectral types. By the late 1930s, astronomers had discovered that nuclear fusion in a star's core and the subsequent accumulation of heavy elements led to stellar evolution. At the end of their lifespan, stars become extinct in the form of white dwarfs, neutron stars, or black holes. This evolutionary journey is dependent upon their initial mass and, consequently, their luminosities. Stars on the main sequence adhere to a well-defined relationship between luminosity (L) and effective temperature (T_{eff}), expressed as:

$$L \propto T_{\text{eff}}^6 \quad (1.1)$$

Stars can be conceptualized as black bodies that absorb all electromagnetic energy falling upon them without reflecting or scattering any radiation. The black-body radiation curve for different temperatures exhibits a peak at distinct wavelengths. *Wilhelm Wien's*

discovery established that as temperature increases, the spectrum of black-body radiation shifts toward shorter wavelengths. Wien's displacement law formally states that the wavelength at which maximum power is emitted can be calculated using the equation:

$$\lambda_{\text{peak}} = b/T \quad (1.2)$$

where T represents absolute temperature. The constant b is known as Wien's displacement constant, equal to 2.89×10^{-3} mK.

In this context, λ refers to color, with very hot stars emitting most of their visible light in the blue wavelength, while cooler stars emit predominantly in the red. Therefore, the color of a star provides an indication of its surface temperature. Astronomers typically measure a star's apparent brightness through filters to determine its exact color. While various filters can be used, a commonly employed set includes U (ultraviolet), B (blue), and V (visual). The brightness measured through each filter is expressed in magnitudes, and the difference between any two magnitudes is referred to as the color index. Astronomers assigned a color index of 0 to a star with a surface temperature of approximately 10,000 K, such as Vega. Since magnitudes decrease with increasing brightness, a smaller B-V value indicates a bluer and hotter star, while a larger B-V value suggests a redder and cooler star.

In 1911, Danish astronomer *Ejnar Hertzsprung* plotted the luminosities of stars against their colors. In 1913, Henry Norris Russell constructed a diagram correlating stellar luminosities with spectral types of stars. Both color and spectral type serve as measures of a star's effective temperature. These diagrams came to be known as *Hertzsprung-Russell* (HR) diagrams, and the location of stars within these diagrams indicates the phase of evolution in which they reside.

1.6 NGC 3324

The Milky Way, our galaxy, assumes the form of a barred spiral with a central crossbar and bulge from which several spiral arms extend. These structures consist of billions of stars and thousands of gas clouds. The *Carina-Sagittarius* Arm stands out prominently in the Milky Way due to its abundance of HII regions, young stars, and giant molecular clouds. Within the *Carina-Sagittarius* Arm lies the Carina constellation, home to the *Carina Nebula* (catalogued as NGC 3372). This nebula comprises a vast and intricate expanse of bright and dark nebulae, captivatingly visible within the southern Milky Way. NGC 3324, a very youthful open star cluster, resides within a region of active star formation. It occupies a position within a partial ring of structured nebulosity situated northwest of the immense *Carina Nebula* (Walborn 1982).

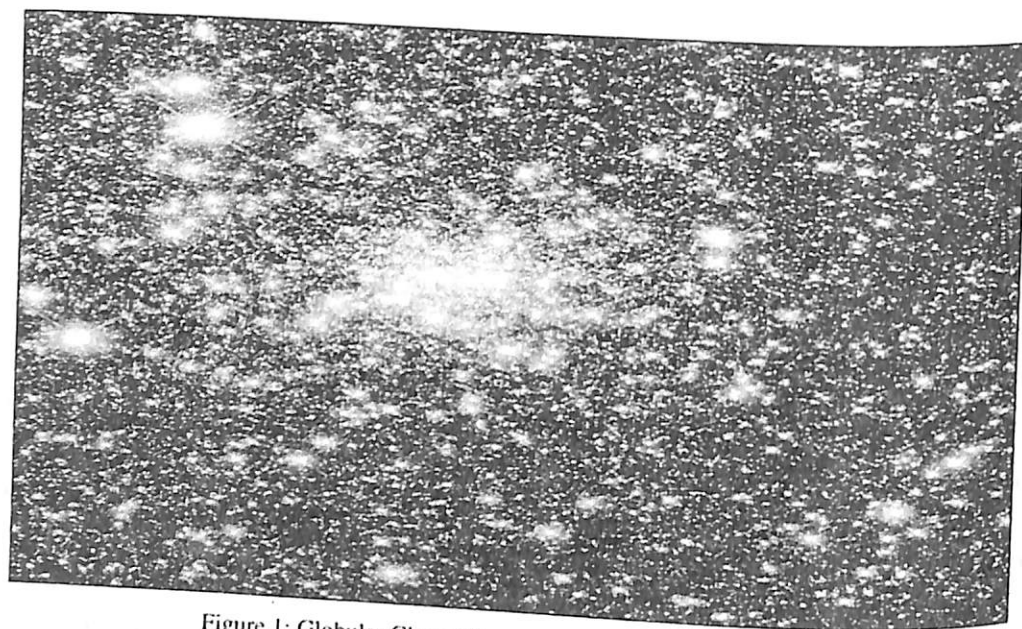


Figure 1: Globular Cluster Terzan 4 (ESA/Hubble & NASA).

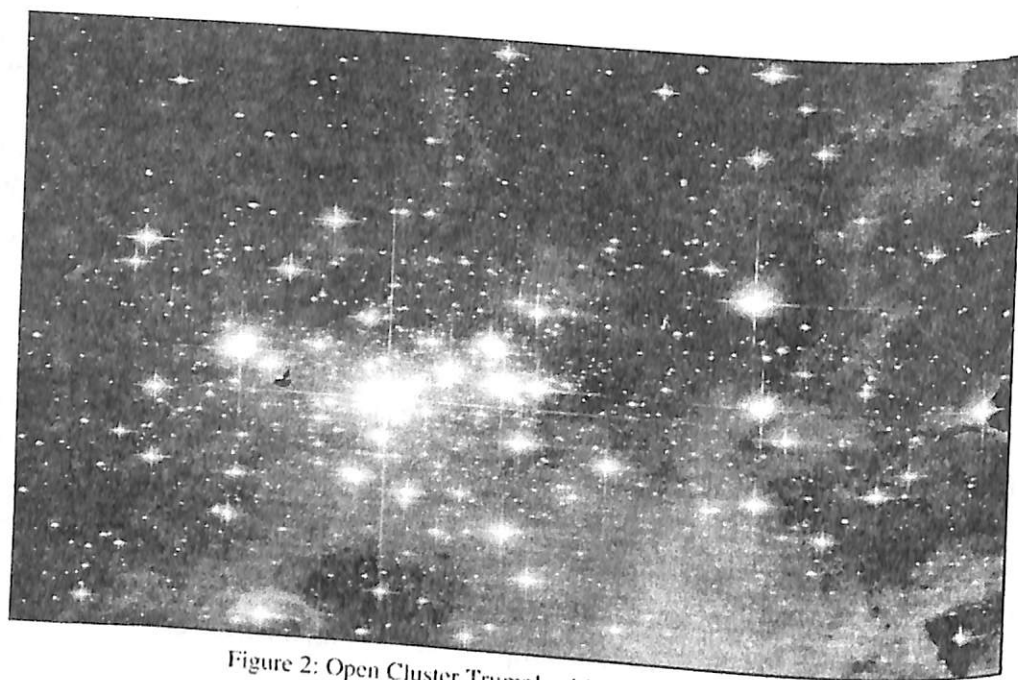


Figure 2: Open Cluster Trumpler 14 (ESA & NASA).



Figure 3: NGC 3324, DSS Image: 12 x 12 arcminutes.

This open star cluster is relatively small and compact, and its significance stems from its presumed extreme youth and significant pre-main sequence population. Consequently, it provides an excellent opportunity to gain insights into the early stages of star formation within that specific region.

The intense ultraviolet radiation emitted by these young, hot stars causes the gas cloud to emit a glowing light and has gradually sculpted a void within the surrounding gas and dust. In fact, the stars are gradually eroding the gas cloud over time. The areas of

darkness visible in the image represent regions where layers of dust obstruct the light emanating from the illuminated gas in the background. Astronomers often assign informal names to nebulae based on their shapes and resemblances to earthly objects. The edge of the gas and dust wall on the right side bears a striking similarity to the profile of a human face, with a distinctive "bump" resembling a nose. Consequently, NGC 3324 is commonly referred to as the *Gabriela Mistral* nebula, in honor of the Nobel Prize-winning poet from Chile.

The study of this cluster involved observations in the optical and infrared spectrum, as well as determination of the fundamental parameters. Color-magnitude and color-color diagrams were also constructed, and the age of the cluster was estimated through the use of isochrones. Additionally, the presence of objects exhibiting infrared excess was identified through the analysis of color-color diagrams.

2 Methodology

2.1 Optical CCD Photometry

One of the most widely used photometric systems is the UBV (Ultraviolet, Blue, Visual magnitudes) system, which was established by *Harold Johnson* and *William Morgan* in 1953. This system, also known as the Johnson system or *Johnson-Morgan* system, covers a broad range of the electromagnetic spectrum, with the B and V regions corresponding roughly to the visual responses captured by photographic methods, and the U region covering the violet and ultraviolet range. The filters used in this system are as follows:

U = Central wavelength 365 nm, Bandwidth 70 nm

B = Central wavelength 440 nm, Bandwidth 100 nm

V = Central wavelength 550 nm, Bandwidth 90 nm

It was the first standardized photometric system known. Despite its popularity, the UBV system has certain drawbacks. The Earth's atmosphere absorbs the shorter wavelengths in the U response, while the B and V response curves are relatively unaffected by atmospheric absorption. The scales are calibrated in a manner that the magnitudes obtained through all three filters are equal to each other ($U=B=V$) for A0 V stars like α Lyrae (Vega). This is referred to as the zero points of this system.

To investigate properties such as distance, age, and extinction reddening law of the NGC 3324 cluster, optical CCD UBV photometry was conducted. The B and V magnitude data were obtained from the WEBDA archive.

2.2 2MASS NIR Photometry

The "Two Micron All-Sky Survey" (2MASS) is an extensive astronomical survey

of the entire sky conducted in the infrared spectrum. Observations were carried out using 1.3 m telescopes at *Mount Hopkins* in Arizona for the northern survey and at *Cerro Tololo* in Chile for the southern survey. The data from both telescopes were processed, calibrated, and subjected to quality control at the *Infrared Processing and Analysis Center* (IPAC). Both the northern and southern 2MASS telescopes utilized similar optics, filters, and detectors. The survey focused on the short-wavelength infrared range with three distinct frequency bands (J, H, and K) centered around 2 micrometers, which gives the survey its name (Carpenter 2001). The wavelengths of the filters used in 2MASS are as follows:

J-band = 1.235 μ m

H-band = 1.662 μ m

K_s-band (K-short) = 2.159 μ m

The *Near Infrared* (NIR) JHK_s data for NGC 3324 was extracted from the 2MASS point source catalog available in the Infrared Science Archive (IRSA). Using the JHKs colors obtained, various color-magnitude diagrams were constructed.

All NIR photometric systems currently in use are considered natural systems. Measurements obtained with specific instruments (filters, detectors, etc.) are typically published without any color transformations and often lack a precisely defined zero point. Such results, when published in a natural system, become impractical to use. It is essential to understand the photometric properties of a system in relation to others and transform the data to a common system (Koornneef 1983).

2MASS provides JHKs band photometry for millions of galaxies and stars. While the 2MASS data alone can contribute significantly to scientific studies, their value can be further enhanced by comparing them with existing photometric measurements. However, a challenge arises when directly comparing 2MASS photometry with other NIR observations. These comparison data are often obtained using filters that differ in various aspects from those used in the 2MASS survey. Such variations can lead to differences in observed stellar colors, particularly for objects with extremely red spectral energy distributions or unusual spectral line features. Therefore, detailed comparisons require both sets of photometry to be placed on a common photometric system. To address this, the JHKs colors were transformed from the 2MASS system to the *Koornneef* system using the relations provided by Carpenter (2001):

$$(J-H)_{2MASS} = 1.019(J-H)_{Koornneef} + (-0.038) \quad (2.1)$$

$$(H-K)_{2MASS} = 0.779(H-K)_{Koornneef} + 0.030 \quad (2.2)$$

$$(J-K)_{2MASS} = 0.974(J-K)_{Koornneef} + (-0.011) \quad (2.3)$$

$$K_{2\text{MASS}} = K_{\text{ornneef}} + (-0.046) + 0.036 (J-K)_{K_{\text{ornneef}}} \quad (2.4)$$

Using these transformed colors, different color-magnitude diagrams were plotted. Additionally, color-color diagrams were also constructed in the 2MASS system.

2.3 Color-Magnitude or H-R Diagrams

Figure 2.1 shows the B-V versus V CMD (Color-Magnitude Diagram) for all the sources found in optical from WEBDA.

Determining the ages of stars is a crucial aspect of observational studies aimed at understanding stellar and galactic evolution. One of the primary methods used on a large scale to assess stellar ages is the application of theoretical stellar evolutionary sequences, specifically through isochrone fitting (*Jorgensen and Lindegren, 2005*). An isochrone refers to a curve on the *Hertzsprung-Russell* (H-R) diagram that represents a population of stars sharing the same age. As stars evolve, they undergo changes in their positions on the HR diagram, and isochrones depict the evolutionary path they follow. By knowing the initial mass function of an open cluster, isochrones can be computed for any given age. This involves taking each star within the initial population and simulating its forward evolution to the desired age using numerical methods. Subsequently, the star's luminosity and magnitude are plotted on the HR diagram, resulting in an isochrone. This isochrone can then be compared with the observed color-magnitude diagram to assess the level of agreement. If the match is close, it indicates that the assumed age of the isochrone aligns well with the actual age of the cluster.

The Figure 2.2 shows the best fitted Geneva isochrones. A correction for the binary fraction in the cluster was done with a probability of 0.5. Binary fraction indicates the ratio of binary to single stars in a cluster which is chosen to be 0.5 as indicated by WEBDA. The different colored lines indicate isochrones for different metallicity for the same age. The fainter lines are the corrections for binary fraction.

The Figure 2.3 shows the best fitted *Padova* isochrones. A correction for the binary fraction in the cluster was done with a probability of 0.5.

The Geneva stellar evolution code has been primarily developed and utilized for the purpose of computing models of massive stars. Extensive efforts have led to the publication of grids of stellar models encompassing a broad range of masses. Within the model code, considerations for mass loss through anisotropic winds and rotational effects have been incorporated by *Meynet and Maeder (1997)*. These codes are also employed to calculate the properties of stars at various stages of their evolutionary lifespan. On the other hand, the *Padova* isochrones represent a combination of distinct stellar models. The foundational models used in these theoretical isochrones were computed by *Bertelli (1994)* and *Girardi (1996)*.

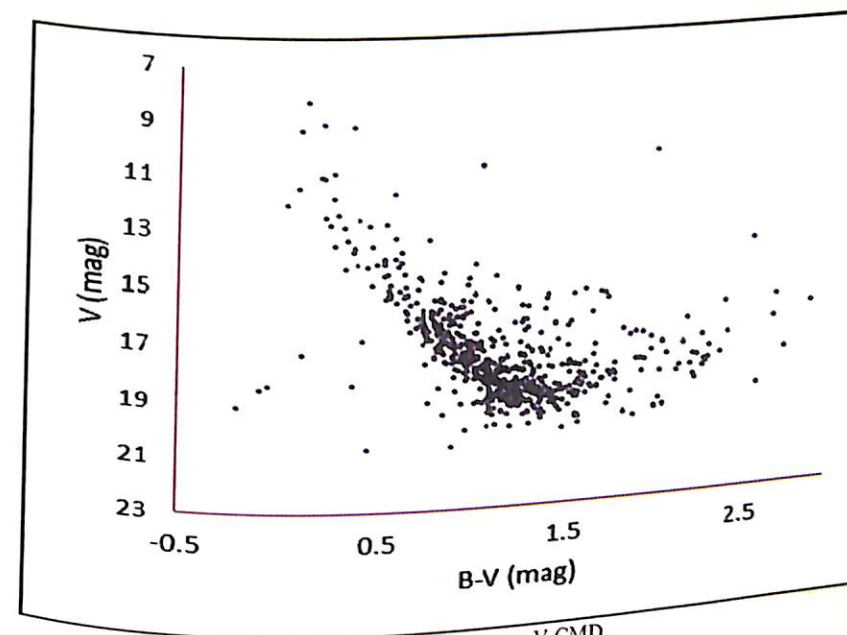


Figure 2.1: B-V versus V CMD.

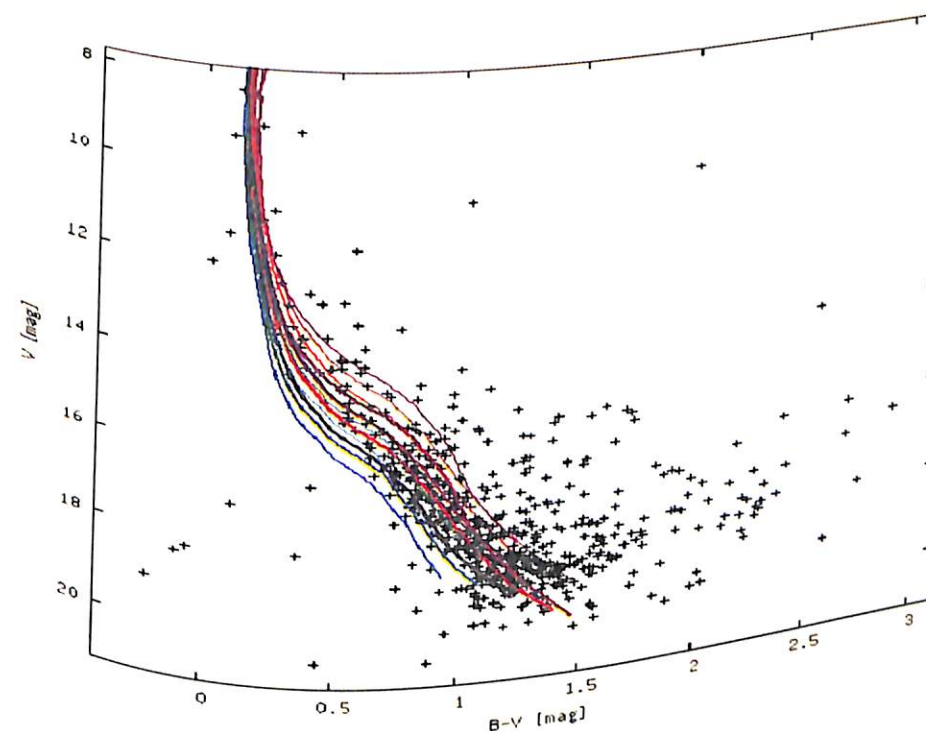


Figure 2.2: Best fitted Geneva isochrones.

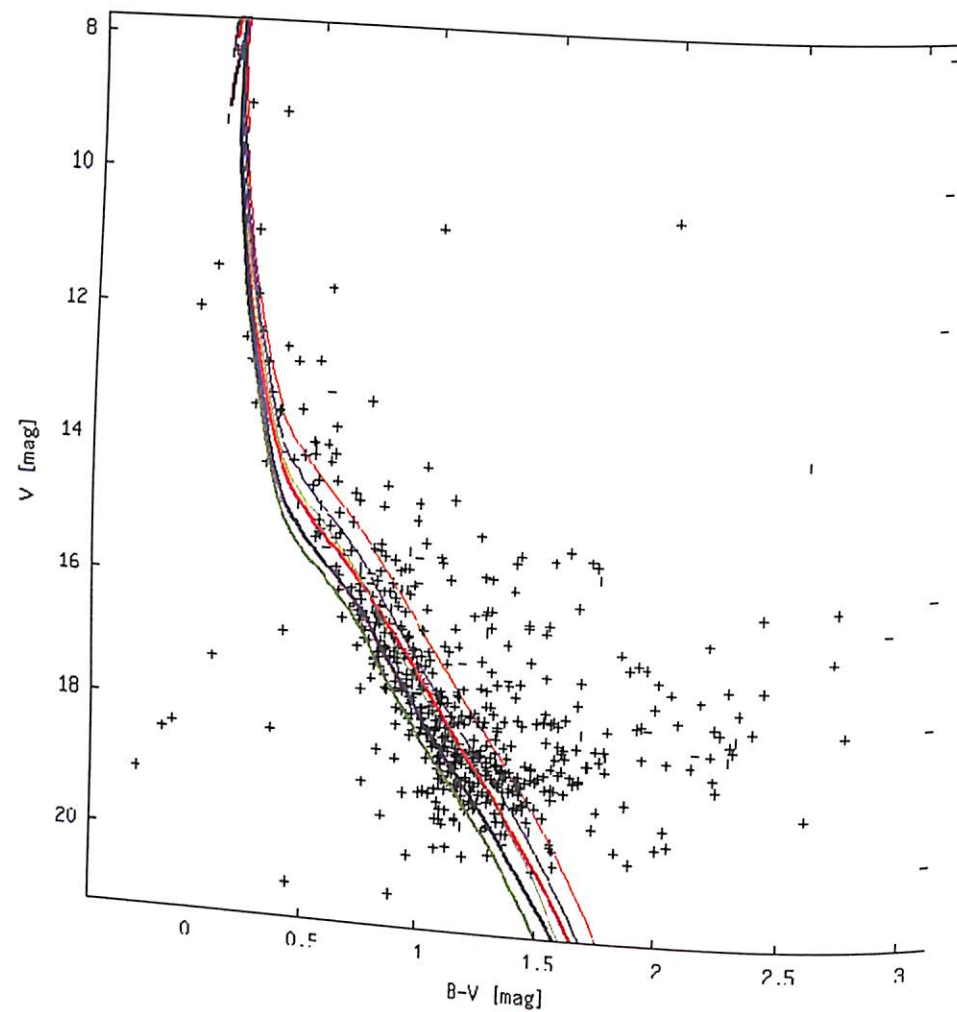


Figure 2.3: Best fitted Padova isochrones.

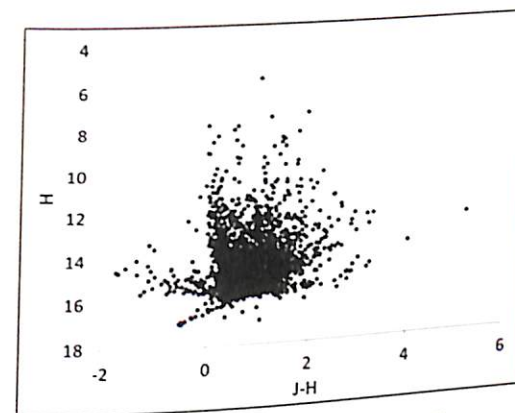


Figure 2.4: J-H versus H CMD (2MASS)

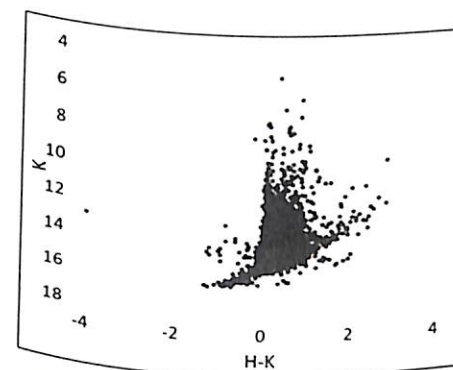


Figure 2.5: H-K versus K CMD (2MASS)

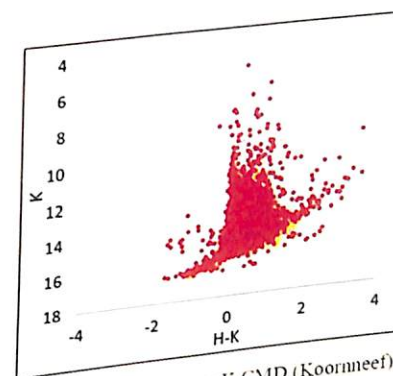


Figure 2.6: H-K versus K CMD (Koomneef)

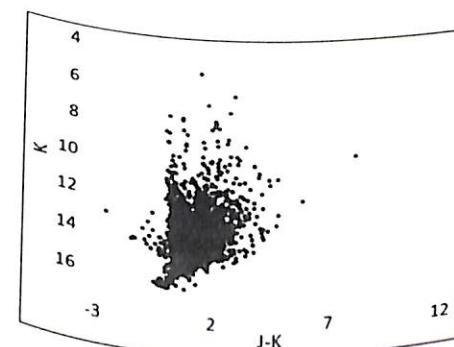


Figure 2.7: J-K versus K CMD (2MASS)

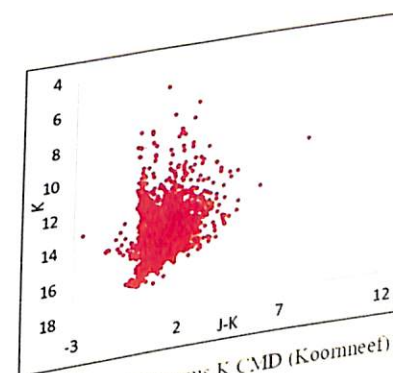


Figure 2.8: J-K versus K CMD (Koomneef)

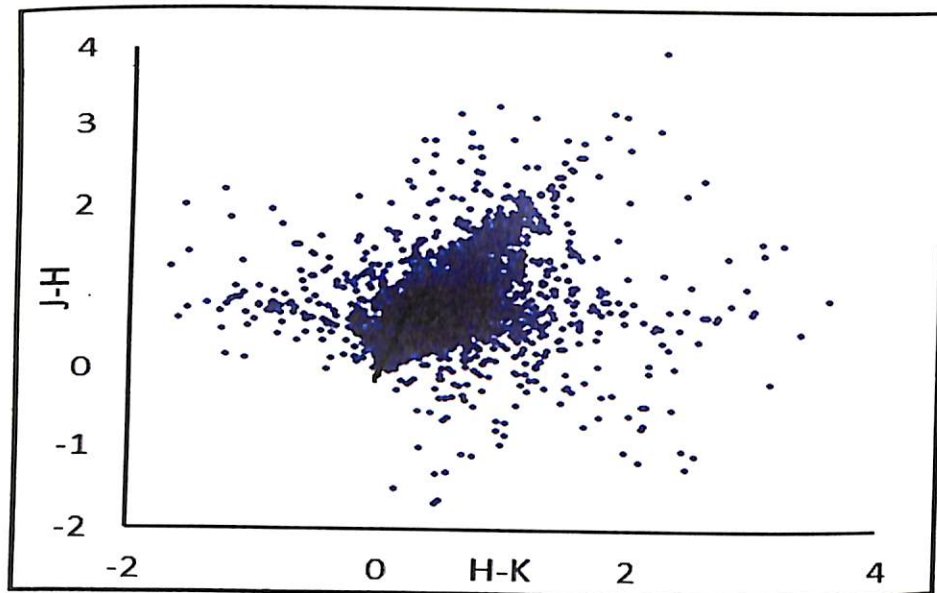


Figure 2.9: J-H versus H-K color-color diagram (converted to Koornneef's system).

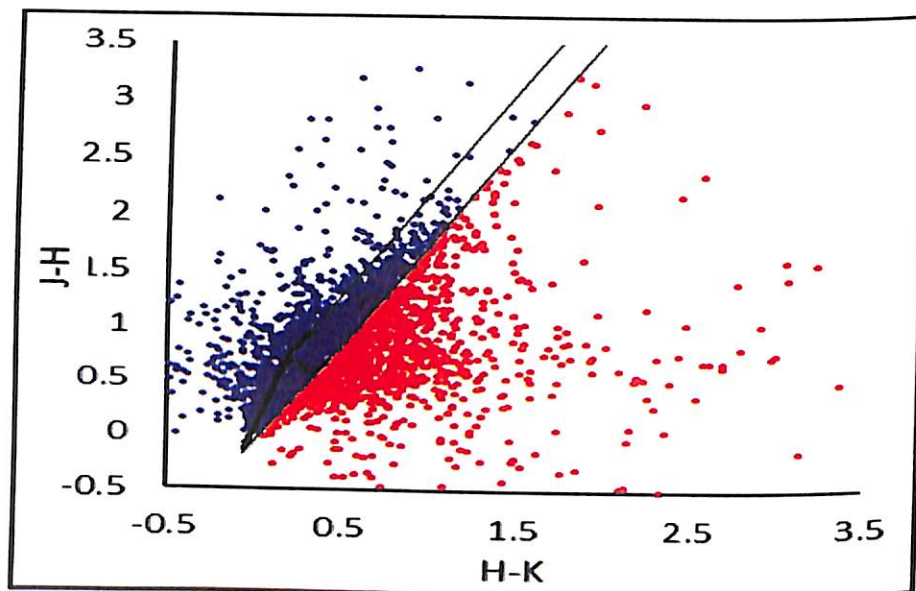


Figure 2.10: To the right of the reddening band, there the NIR excess sources (red circles).

The color-magnitude diagrams (CMDs) provide an estimated logarithmic age of 6.75 million years (Myr), which corresponds to a slightly lower age estimate of 5.62 Myr compared to the 6.75 Myr calculated by *Kaltcheva and Golev* (2012). Regarding metallicity, the Geneva model suggests a value of $Z=0.040$, while the Padova model indicates $Z=0.019$.

Metallicity serves as an indicator of the chemical composition of a star cluster or galaxy. In terms of mass fractions, where X represents the fraction of hydrogen, Y represents the fraction of helium, and Z represents the fraction of other elements, these fractions are related by the equation:

$$X + Y + Z = 1 \quad (2.5)$$

A metallicity estimate of $Z=0.040$ (Geneva) or $Z=0.019$ (Padova) suggests that the cluster has a relatively low metal content.

In the near-IR, the J-H versus H (obtained from 2MASS) and H-K versus K and J-K versus K (obtained from 2MASS and transformed to *Koornneef*) CMD for all the sources have been plotted (Figure 2.4 to 2.8).

2.4 Color-Color Diagrams

Figure 2.9 shows a color-color diagram (J-H, H-K) of the cluster, where the colors have been transformed to the *Koornneef* system. It is assumed that the reddening law established by *Rieke & Lebofsky* (1985) is applicable to the cluster NGC 3324, providing a reasonable approximation of the near-infrared (NIR) extinction caused by the associated molecular material. The solid lines in the figure indicate the positions of unreddened main-sequence, giant, and supergiant stars, as determined by *Koornneef* (1983). Using the extremes of these curves, two dashed lines parallel to the *Rieke & Lebofsky* (1985) interstellar reddening vector has been plotted. The region enclosed between these lines, known as the reddening band, represents as the infrared excess region (*Lada & Adams 1992*), corresponds to the presence of pre-main sequence stars. However, naked-T Tauri stars, post-T Tauri stars, and certain class I vectors do not exhibit any near-IR excess and can be found between the two reddening vectors in such a diagram (*Maheswar et al. 2006*).

In the color-color diagram (CCD), 1081 sources have been identified that exhibit clear NIR excess emission (depicted as red circles in Figure 2.10). Within the region of NIR excess sources, two distinct populations can be observed. The first set consists of sources with $J-H < 0.7$, while the second set comprises sources with $J-H > 0.7$. Among the 1081 NIR excess sources, 540 fall into the first set, whereas 541 belong to the second set. The NIR excess sources in the first set exhibit a redder color, indicating that they may be very young (*Maheswar et al. 2006*).

3 Results and Discussion

In this study, both optical and infrared investigations of the NGC 3324 cluster have been conducted. The analysis was focused on determining key parameters such as the cluster's age

and metallicity. Based on optical isochrones, the age of the cluster was estimated to be 5.62 Myr, slightly different from the previously reported value of 6.75 Myr (Kaltcheva & Golev 2012). This discrepancy could be attributed to the use of different stellar evolution codes for generating the theoretical age isochrones. Additionally, metallicity was found to be very low ($Z_{\text{Geneva}}=0.040$ and $Z_{\text{Padova}}=0.019$), indicating that NGC 3324 is a young open cluster.

By analyzing the 2MASS JHK_s measurements, it was identified that a significant number of NIR excess sources is present within NGC 3324. Utilizing the J-K versus H-K color diagram and assuming the applicability of Rieke & Lebofsky's reddening law in the region, the positions of unreddened main sequence and giant branches have been plotted based on Koornneef's work (1983). This analysis enabled to identify sources located in the NIR excess region, which are indicative of very young stars. These findings suggest that ongoing star formation processes may be occurring in this region.

It is worth noting that the available literature on the NIR domain study of NGC 3324 remains limited. Given the significant number of NIR excess sources identified within the cluster, a more comprehensive investigation in the NIR range could provide valuable insights into its morphology and ongoing star formation activity.

References

1. 2MASS All-Sky Catalog of Point Source
2. Adams, F. C., & Myers, P. C., ApJ, 553, 744 (2001)
3. Baumgardt, H., & Kroupa, P., MNRAS, 380, 1589 (2007)
4. Bertelli, G., Bressan, A., Chiosi, C., Fagotto, F., Nasi, E., A&AS, 106, 275 (1994)
5. Bressan A. MNRAS, 427, 127-145 (2012)
6. Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., MNRAS, 427, 127 (2012)
7. Carpenter, J. M., AJ, 121, 2851 (2001)
8. Carraro, G., Patat, F. & Baumgardt, H., A&A, 371, 107-114 (2001)
9. Chandar, R., Fall, S. M., & Whitmore, B. C., ApJ, 711, 1263 (2010)
10. Chen, W. P., Chen, C. W. & Shu, C. G., AJ, 128, 2306-2315 (2004)
11. Claria, J. J., A&AS, 27, 145 (1977)
12. Cutri R.M., Skrutskie M.F., Van Dyk S., et al. (2003)
13. Froebrich, D., Scholz, A. & Raftery, C. L., MNRAS, 374, 399-408 (2007)
14. Garcia, B., & Mermilliod, J. C., A&A, 368, 122 (2001)
15. Gieles M., Baumgardt H., Heggie D. C., Lamers H. J. G. L. M., MNRAS, 408, L16 (2010)
16. Gieles M., Portegies Zwart S. F., Baumgardt H., Athanassoula E., Lamers H. J. G. L. M., Sipior M., Leenaarts J., MNRAS, 371, 793 (2006)
17. Girardi, L., Bressan, A., Chiosi, C., Bertelli, G. & Nasi, E., A&AS, 117, 113-125 (1996)
18. Handbook of CCD, Steve B. Howell
19. Hurley, J. & Tou, C. A., MNRAS, 300, 977-980 (1998)
20. Jorgensen, B. R. & Lindegren, L., A&A, 436, 127-143 (2005)
21. Joshi, Y. C., Balona, L. A., Joshi, S. & Kumar, B., MNRAS, 437, 804-815 (2014)
22. Kaltcheva, N. T. & Golev, V. K., Astronomical Society of the Pacific, 124, 128-139 (2012)
23. Khalaj P., Baumgardt H., MNRAS, 457, 479 (2016)
24. Kharchenko, N. V., Piskunov, A. E., Schilbah, E., Scholz, R.-D., A&A, 440, 403 (2005)
25. Kitchin, C. R., Astrophysical Techniques, Institute of Physics Publishing
26. Koornneef, J., A&A, 128, 84-93 (1983)
27. Kroupa, P. 2001, MNRAS, 322, 231
28. Lada, C. J. & Adams, F. C., ApJ, 393, 278-288 (1992)
29. Lada, C. J., & Lada, E. A., ARA&A, 41, 57 (2003)
30. Lejeune, T. & Schaerer, D., A&A, 366, 538-546 (2001)
31. Maheswar, G., Saurabh, S., Medhi, B. J., Pandey, A. K & Bhatt, H. C., MNRAS, 000, 000-000 (2006)
32. Massey, Philip, Johnson, Kelsey E., Degioia Eastwood, Kathleen, 1995, ApJ, 454, 151
33. McKee, C. F., & Ostriker, E. C. 2007, ARA&A, 45, 565
34. Meynet, G. & Maeder, A., A&A, 321, 465-476 (1997)
35. Palla, F., & Stahler, S. W., ApJ, 540, 255 (2000)
36. Portegies Zwart, S. F., & McMillan, S. L. W., ApJ, 576, 899 (2002)
37. Ramadhani Putri Ayu, R. P. et al., The Korean Astronomical Society, 30, 275-278 (2015)
38. Rieke, G. H. & Lebofsky, M. J., ApJ, 288, 618-621 (1985)
39. Selim, I.M., Essam, A., Hendy, Y.H.M. & Bendary, R., NRIAG Journal of Astronomy and Geophysics, 5, 16-22 (2016)
40. Theory of Star Formation, Palla and Stahler
41. Tognelli, E., Moroni, P. G. P. & Degl'Innocenti, S. (2018)
42. Walborn, N. R., AJ, 87, 9 (1982)
43. WEBDA archive: webda.physics.muni.cz

On the pulsar-like compact stars: Strangeon stars

Jyatsnasree Bora⁽¹⁾

DhrubaJyoti Gogoi⁽²⁾

Shyamalee Bora⁽³⁾

^{(1),(2),(3)}Department of Physics

Dibrugarh University, Dibrugarh, India

Email: jyatsnasree.borah@gmail.com

Abstract

This article explores one of the hypothetical compact objects known as strangeon stars. From different astrophysical phenomena, it is pointed that pulsar like compact objects may be comprised entirely of strangeons (quark-clusters with three-light-flavor symmetry) and a small amount of electrons to maintain charge neutrality. Detailed analyses of these stars are discussed here including their formation, identification or manifestation. Such stars have the key ingredient known as strangeon matter.

Introduction

The study of the stellar objects that are born after the death of a normal star has fascinated mankind for a long. There is a sea of mysteries to date related to this enchanting area of strong gravity. This study is even more accelerated after the detection of gravitational waves. This paramount confirmation of Einstein's theory of relativity has been able to draw significant attention from the scientific community. The study of pulsars and neutron stars also belongs to this area of astrophysics. With some recent missions like NASA's NICER mission, investigations on such high gravity regimes are going on. The exploration of the core of neutron stars is a fascinating topic of current-day research as we are lacking proper knowledge of the core of

J. Bora, D.J. Gogoi, S. Bora, On the pulsar-like compact stars: Strangeon stars, Horizon (Ed. Book), 170-176 (2023), copyright: Tinsukia College Publishing, ISBN: 978-81-19322-07-7

these compact stars. For such a mysterious core, it is possible that there might exist free-flowing deconfined quarks known as quark matter. Such a conventional neutron star with a core of quark matter is dubbed as hybrid star. It is further proposed for strange quark stars in which the quark matter constitutes the entire star. These stars are more compact than the neutron stars. They are composed of strange quark matter. The strangeon matters (quarks localized almost in a certain unit, called strangeon) are similar to the strange quark matter as for both of these matter the constituting elements are the u,d,s quarks. The term 'strangeon' is coined by combining "strange nucleon" for the sake of simplicity. This term, which is also regarded as an acronym of strange nucleon, was used to characterise a novel type of component identical to nucleons but with strangeness. Though the constitutes are similar yet these two stars have distinctive characteristics. For the strangeon matter the quarks are localized as clusters and they form a state more like a solid. The stars made up of strangeon matter are known as the strangeon stars and have intrinsic stiff equation of state. Due to their large compactness than the strange quark stars and neutron stars they are proposed to support massive pulsars ($> 2M_{\odot}$).

A great challenge for physicists regarding the study of compact stars is the state equation of dense matter around nuclear density. It is proposed that strangeness might be relevant to the solution to this challenge [1]. There are a few proposed equations of state for such strangeon matter. Important equations of state to describe such matter are discussed later in this article. The nature of pulsar-like compact stars is fundamentally a core problem of the strong interaction at low energy scale (described in quantum chromodynamics), whose solution is still a problem despite enormous efforts. The recent advancement of various Earth and space-based missions primarily focused on pulsars and neutron stars have been able to draw different characteristics of these compact objects. However, the core physics of neutron star is not unveiled yet. It may be conjectured that the core of such a star might be solid. Theoretically, such a possibility cannot yet be ruled out through a conventional neutron star is fluid-like except for a solid crust [4]. For the so called strangeon stars having nucleon-like units with strangeness, should be in a globally solid state due to the massive and strongly coupled strangeons. These strangeons may form in bulk supra-nuclear matter produced during core-collapse supernova [4,5,6].

It is generally believed that strangeness would play a significant role in understanding the puzzling state, to be probably the first major problem solved in the era of gravitational-wave astronomy. Furthermore, it is proposed that pulsars might be strange quark stars with mass and radius similar to that of conventional neutron stars, making strange quark stars a potential candidate model for this class of compact objects. Quarks would be a strange star's fundamental building blocks, but strangeons might instead exist if three-flavored quarks localise in strangeons as they do in nucleons in the two-flavored scenario. It is to be noted here that the concept of strangeon stars are successful in explaining many phenomena of pulsar-like stars. Among these, subpulse-drifting, the glitches interpreted with star-quakes, the Optical/UV excess in

X-ray dim isolated NSs, as well as massive pulsars ($\approx 2M_{\odot}$) proposed before discoveries can be explained with the concept of strangeon stars. Additionally, the tidal deformability and light curve from GW170817 data are in agreement with the strangeon star model. In the future, the hypothesis of this compact stars may potentially be examined by looking for gravitational-wave echoes connected to strangeon stars [7].

Literature survey confirms that research on strangeon stars is reported in a plethora of articles by various scientists in the past few years [2, 3, 4]. Possibilities for the release of huge free energy besides the kinematic energy of rotation of pulsar-like compact objects are recently reported in Ref. [4]. It is further reported therein that the huge amount of free energy ($>10^{46}$ erg) could arise due to starquakes, under an extremely small anisotropy. The free energy of pulsar-like compact object depends unquestionably on the equation of state of bulk matter at supra-nuclear density.

We have arranged the rest of this article as follows: in Section:2, the formation of strangeon stars is discussed. The distinctive behaviour of the strangeon stars from other compact stars is discussed in Section:3. The equations of state to describe strangeon matter are added in Section:4. In Section:5 a brief summary and concluding remark is made.

Formation of strangeon matter

A strangeon is, in principle, a colour-singlet N_q - quark state with $q = 6, 9, 12, 15$, and 18, as well as the equal numbers of u, d, and s quarks. According to numerous research, a strangeon may not be stable or just weakly bound in vacuum since those multi-quark states are not observed. Strangeon matter, on the other hand, may become stable if strangeons are packed closely together because of the strong attraction interactions, which might create compact stars known as strangeon stars. De-electronization, or the killing of electrons, is necessary to establish the stellar collapse state of gravity-compressed atomic matter. Strangeonization would be an alternative to neutronization, which is often mentioned as the method of de-electronization from Landau's period. It is important to note that weak interaction is the mechanism for both strangeonization and neutronization [2]. When all of the collapsing-core nuclei contact and combine to create a massive nucleus (i.e., strong matter), electrons, which normally contribute very little energy in ordinary electromagnetic matter (i.e., atoms and ions), turn out to be very relativistic. In this case, bigger is different: a macro-nucleus is distinct from a micro-nucleus. In Fig. 01, two potential de-leptonization scenarios are evaluated, and as a result, two models (neutron star and strangeon star) are suggested [2,3].

Distinction of strangeon stars from different pulsars

In Fig. 02, different models of pulsar's interior structure are shown. Among these models the neutron stars (hadron star and hybrid/mixed star) are gravity-bound. The other two star

models: strange quark stars and strangeon stars are self-bound on surface by strong force [2]. In neutron stars, the core should have a boundary and it maintains equilibrium with the ordinary matter as the star has a surface composed of ordinary matter. At each boundary of such stars, the inner and outer cores and the crust keep chemical equilibrium resulting in gravity bound configuration [2]. Given that the star's surface is made up of ordinary stuff, the core should have a limit and be in equilibrium with it. However, there isn't any conclusive observable proof of a neutron star's surface. Strangeon stars are similar to conventional strange quark stars in that their internal structure is almost identical to that of their surface. This similarity makes it possible to use the strangeon matter surface to explain a variety of observations. However it is interesting to note that strange quark stars are bound by bag-like strong interaction, and strangeon stars are bound by residual strangeon interaction.

The equation of state of strangeon matter

Using equations of state, any material's characteristics may be characterised. For the case of supra-nuclear matter, the equation of state is still under hot debate. The strangeon stars can be described using the phenomenological Lennard-Jones model [5]. It assumes an interaction potential between two strangeons which can be described by the Lennard-Jones potential as

$$u(r) = 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6] \quad (1)$$

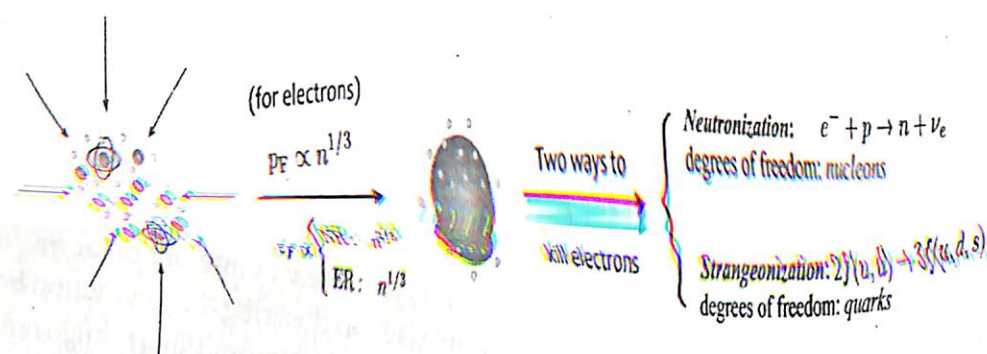
here, ϵ represents the depth of the potential, σ is the range of interaction or the distance where $u(r)$ becomes zero. This model was initially used to describe the interaction between two inert gas molecules, with the property of long-range attraction and short-range repulsion. To employ this model for strangeon stars one may note that such potential helps quark matter to crystallize and to form solid strange stars. As for short distances ($r \leq 0.5$ fm), the lattice QCD reveals that there is a strong repulsive core of a few hundred MeV surrounded by an attractive well at medium and long distances [4]. The total energy density and pressure for strangeon matter by considering simple cubic lattice structures, ignore the surface tension and vibration energy can be given as [4]

$$p = 2\epsilon(A_{12}\sigma^{12}n^5 - A_6\sigma^6n^3) + nN_q m_q c^2 \quad (2)$$

$$p = n^2 \frac{d(p/n)}{dn} = 4\epsilon(2A_{12}\sigma^{12}n^5 - A_6\sigma^6n^3) \quad (3)$$

In these expressions for energy density and pressure, ϵ represents the number of quarks and n is the number density of strangeons [4].

Neutronization V.S. Strangeonization



Bigger is different: compression of normal baryonic matter by gravity

Energetic electrons inside a gigantic nucleus

Neutron star or Strangeon star?

Fig 01: A comparison between neutronization and strangeonization, the key way to de-electronize [2]

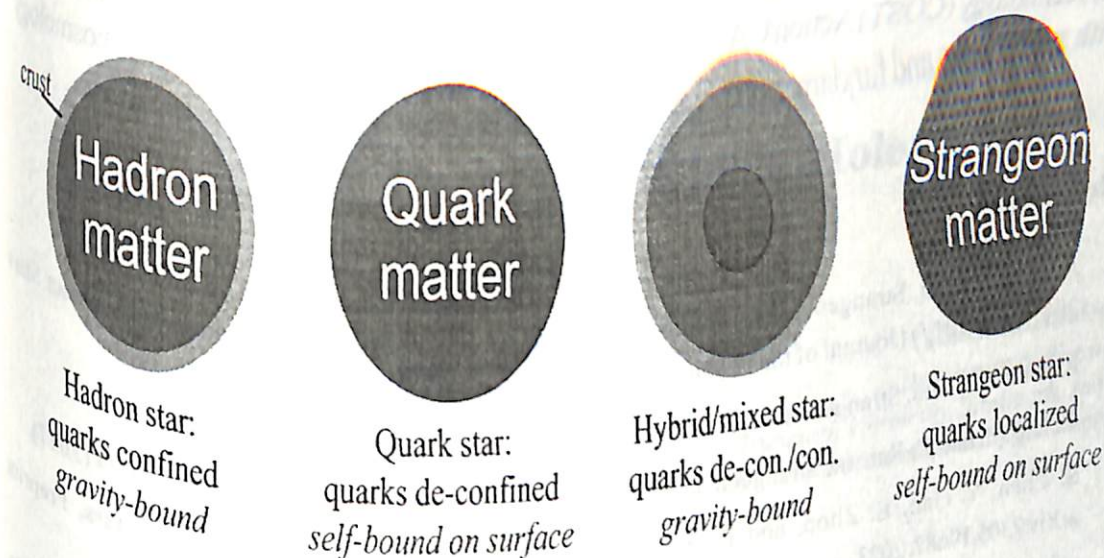


Fig 02: Different pulsar's nature. Hadron star and hybrid/mixed star belong to conventional neutron stars. For the three-light-flavor (u, d and s) symmetry strangeness plays an important role for strange quark star and strangeon star [3].

Conclusion and Outlook

Normally, pulsars are believed to be typical neutron stars, however, if Witten's hypothesis (the true ground state of matter is strange quark star rather than Fe^{56} is true, they may actually be strange quark stars. Unfortunately, because of the non-perturbative behaviour of strong interactions, it is currently difficult to demonstrate or refute this claim. If the coupling between quarks is still strong enough, 3-flavoured quarks might be gathered in strange-clusters or strangeons, where a self-bound strangeon star could form. Furthermore, the properties of strangeons in high density medium are less known [8]. A detailed investigation of such matter will definitely illuminate the physics of these highly compact objects. From the observational point of view, it can be conferred that various manifestations of pulsar-like compact objects could be interpreted in the regime of strangeon stars. Such stars in future can be expected to get detected by future advanced facilities like Five-hundred-metres Aperture Spherical Telescope (FAST), Square Kilometres Array (SKA), and enhanced X-ray Timing and Polarization (eXTP) [8]. These proposed missions are one of the key sciences to know the equations of state of dense matter inside compact stars. With this present and future mission, more knowledge about pulsar-like compact stars will be accumulated. Also, we can expect to test the strangeon star model further.

Acknowledgement

DJG acknowledges the contribution and support of the European Cooperation in Science and Technology (COST) Action CA21136 — “Addressing observational tensions in cosmology with systematics and fundamental physics (*CosmoVerse*)”.

References

1. J. Lu and R. Xu, Strangeon Stars, in Proceedings of the Workshop on Quarks and Compact Stars 2017 (QCS2017) (Journal of the Physical Society of Japan, Kyoto, Japan, 2018).
2. R. Xu and Y. Guo, Strange Matter: A State before Black Hole, in (2017), pp. 119–146
3. L. Xiaoyu and X. Renxin, Strangeon and Strangeon Star, J. Phys.: Conf. Ser. 861, 012027 (2017).
4. S. Chen, Y. Gao, E. Zhou, and R. Xu, Free Energy of Anisotropic Strangeon Stars, Preprint arXiv:2305.19687, 2023.
5. X. Lai, C. Xia, and R. Xu, Bulk Strong Matter: The Trinity, Advances in Physics: X 8, 2137433 (2023).
6. X. Y. Lai and R. X. Xu, Lennard-Jones Quark Matter and Massive Quark Stars, Monthly Notices of the Royal Astronomical Society: Letters 398, L31 (2009).
7. C. Zhang, Y. Gao, C.-J. Xia, and R. Xu, Rescaling Strangeon Stars and Its Implications on Gravitational-Wave Echoes, Preprint, arXiv:2305.13323 (2023).
8. Z.-Q. Miao, C.-J. Xia, X.-Y. Lai, T. Maruyama, R.-X. Xu, and E.-P. Zhou, A Bag Model of Matter Condensed by the Strong Interaction, Int. J. Mod. Phys. E 31, 2250037 (2022).

The Theory of Black Hole

Ronit Karmakar

Department of Physics,
Dibrugarh University, Dibrugarh, India
Email: ronit.karmakar622@gmail.com

Abstract

Black Holes are among the most elusive objects in the universe and has attracted the attention of the scientific community for decades. Predicted first by the Einstein's theory of General Relativity in 1915, these objects give us opportunity to study gravity and space-time in the extreme regimes. The detection of Gravitational waves emitted by spiralling black holes provided the much awaited confirmation of the existence of these objects recently.

Introduction

The term 'Black hole' was first coined by John Wheeler in 1967, but it was predicted much earlier by eminent physicist Albert Einstein in 1915 via his General Theory of Relativity (GR), which is basically a theory of space-time and matter interaction.

According to his theory, planetary objects follow the geodesics or paths which are curved out in space-time due to the presence of massive object. Elaborating this point, we can say that a more massive object produces a deep dent in the fabric of space-time and other smaller objects follow the curved paths thus created in space-time. Even light is bound to follow these curved paths, due to which another spectacular phenomena called Gravitational Lensing was discovered. It is the phenomena of bending of light due to the presence of a massive object.

Our sun is indeed a big star. But some other stars are far bigger, many are millions to billions times larger than our sun. Such stars are very far away from us so that we cannot appreciate their large size. When the fuel of such a star runs out, then the outward radiation pressure due to fusion decreases sharply, while gravitational collapse becomes dominant and

R. Karmakar, The theory of Black Hole, Horizon (Ed. Book), 177-183 (2023), copyright: Tinsukia College Publishing, ISBN: 978-81-19322-07-7

the star is squeezed to a point. This point is so dense and extremely massive that it produces huge curvature in the space-time fabric. Its influence around a certain limit is so extreme that even light cannot exit its strong pull. This point of no return is called the event horizon of the black hole. The study

of event horizon is extremely important to physicist as it is virtually impossible to know what's inside the event horizon as no information can escape from inside the event horizon. Indeed, this is a classical perspective of black hole. The concept of the event horizon can be traced to the solution of the Einstein's equation of GR, which was first solved by German physicist *Karl Schwarzschild* in 1916 for vacuum case. In 1958, *David Finkelstein* interpreted it as an region of space-time from which nothing can escape. It showed the existence of an hypothetical non-rotating massive object in space time, which created a zone of no return around it. In 1931, however, *S. Chandrasekhar* calculated a limiting mass above which a star can collapse to a black hole. In 1939, *Tolman-Oppenheimer-Volkoff* limit gave the limiting mass as around $3 M_{\odot}$. Observations of GW from the event GW170817 refined the limiting mass of a star to form black hole as $2.17 M_{\odot}$. Later, more complex cases were studied, assuming charged black holes and rotating black holes, which will be discussed in a later work. Rotating black holes are called Kerr black holes and charged black holes are called *Reissner-Nordstrom* black holes. Basically, we assign three parameters which are capable of describing a black hole, viz. mass, charge and spin (angular momentum). So black holes are called the cleanest objects of the universe. The following figure 1 is a processed image of a black hole accretion disk created by encircling matter around the black hole.

First Observational Proof

Black holes are really black, i.e. they cannot be directly visible in the electromagnetic spectrum. Indirect observational techniques have been developed to detect these mysterious objects. *Gravitational waves* (GW) produced by spiralling black holes are ripples in space time fabric that produces elongation-compression movement of space time along with objects in it. This principle has been utilized by the *Laser Interferometer Gravitational Wave Observatory* (LIGO) situated in USA at two widely separated locations to counter any noise detection or falsely generated signal. On September 14th of 2015, the first-ever detection of GW produced by two collapsing black holes of mass $36 M_{\odot}$ and $29 M_{\odot}$ was reported by LIGO team and scientifically reported in 2016 in an article published in scientific journal. Later, LIGO-Virgo collaboration reported the detection of several black holes and neutron stars.

The first ever direct image of black hole at the centre of M87 galaxy was published in April 2017 by the *Event Horizon Telescope* (EHT). EHT observes the active environment next to the black holes at the centres of galaxies like our Milky Way galaxy and M87*. This image is representing the disc of matter rotating at relativistic speed around the event horizon of the black hole. A processed image of the black hole is shown in figure 2.

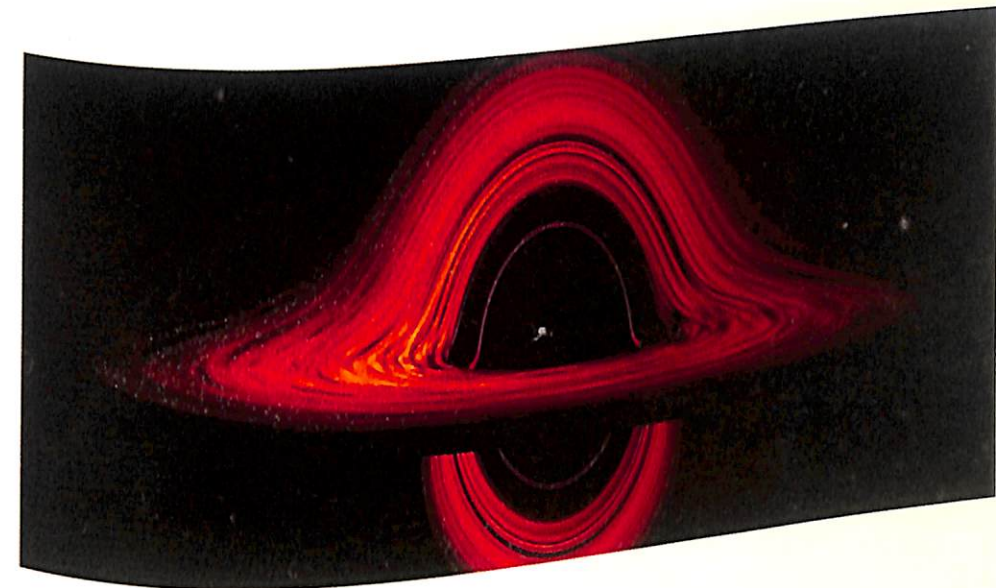


Fig1. Visualization of a black hole with its accretion disc.

Image credit: nasa.gov



Fig2. Processed image of black hole M87*.

Image credit: nasa.gov

Mathematical solution of Black Hole

Solving the Einstein equation for a particular case like vacuum or presence of matter (energy-momentum tensor $T_{\mu\nu}$), we arrive at the metric function, which is perceived as the black hole solution. This metric function denoted by $f(r)$ encodes various information about the event horizon, the Hawking Temperature and surface gravity along with various other properties of a black hole. All we need is the Einstein field equations and we are good to go. We start from a spherical symmetric background metric of Schwarzschild given by

$$ds^2 = -f(r) dt^2 + f(r)^{-1} dr^2 + r^2 d\Omega^2.$$

Here, as $f(r)=0$, the radial component blows up and we say that there is a singularity which exists for the *Schwarzschild black hole*. From this metric, the components of the metric tensor $g_{\mu\nu}$ are easily identified as

$$g_{tt} = -f(r).$$

$$g_{rr} = f(r)^{-1}.$$

$$g_{\theta\theta} = r^2.$$

$$g_{\phi\phi} = r^2 \sin^2 \theta.$$

From these components we can easily compute the *Christoffel's symbols* of second kind as

$$\Gamma^c_{ab} = \frac{1}{2} g^{dc} (g_{ad} / n x^b + g_{bd} / n x^a - g_{ab} / n x^c).$$

The surviving *Christoffels symbols* can be easily calculated and resulting Ricci tensor R_{mn} and Ricci scalar R can be accordingly found. The *Einstein field equation* is given by

$$G_{ab} = R_{ab} - \frac{1}{2} g_{ab} R = 0.$$

Putting the values of the *Ricci tensor* components and the Ricci scalar into the vacuum field equation, we have

$$1/r^2 (r f'(r) + f(r) - 1) = 0.$$

$$1/r^2 (r f'(r) + \frac{1}{2} r^2 f''(r)) = 0.$$

Solving these two equations lead to the solution of the metric function or the black hole solution as

$$f(r) = 1 - 2M/r.$$

Thus the *Schwarzschild metric* can be written as

$$ds^2 = -(1 - 2M/r) dt^2 + (1 - 2M/r)^{-1} dr^2 + r^2 d\Omega^2.$$

Here, $r=2M$ corresponds to the singularity condition for the *Schwarzschild black hole*, however there is another singularity at $r=0$, which is non-removable by simple coordinate transformation.

Similarly, we can theoretically deal with matter content with the introduction of the energy-momentum tensor (T_{mn}) which will compensate for the matter surrounding the black hole in

the field equations. Solving such equations lead to the metric function for that particular black hole. Few such examples are the string clouds field or quintessence dark energy field or any exotic matter content surrounding the black hole. In literature we find works on black hole physics where authors have considered various surrounding fields to the black hole and studied various properties of the black hole like *Quasi normal modes*, shadow, light deflection, thermodynamics etc. In this note, we have shown black holes which are very simple and have only one hair or property, that I mass M . there can be in fact many different types of black holes having several hairs, like charge and spin. Charged black holes are called *Reissner-Nordstrom black holes* and spinning black holes are called *Kerr black holes* after their discoverers.

Some Properties of Black Hole

From the metric function, one can obtain many information regarding the black hole. We will discuss some major properties which can be associated with the metric. The concept of black hole temperature was first proposed by *Bekenstein* in 1973 where he introduced thermodynamic quantities related to black hole such as entropy. Another prominent physicist was S. Hawking who calculated an expression for *Hawking radiation*, a form of radiation of particles from the event horizon of a black hole, which ultimately lead to the evaporation of the black hole. Such ground breaking ideas led to a flood of research in this field. Some of the properties discussed above are shown below with their expressions.

i) Event Horizon

Event horizon is the region in space time near a black hole where, if anything once enters it cannot leave the strong gravitational attraction of the black hole. Event horizon provides us with many information regarding the properties of the black hole and thus is the first quantity which we want to analyse while studying black hole. Event horizon area (A) is the surface area of this horizon. This quantity A is calculated by simple law of area of circle $A = \pi r_H^2 = 4\pi M^2$.

Another interesting concept regarding the surface area is the classical interpretation that a black hole's surface area never decreases. This is self-explanatory as the black hole will only absorb matter inside it and as a result, its size will only increase and never decrease. However this school of thought was later challenged by quantum mechanical interpretation of emission or radiation from a black hole.

ii) Surface Gravity

Surface Gravity of a black hole is a measure of the pulling strength of the black hole at the horizon. It is defined as $\kappa = (1/2) n f(r) / n r$ at $r=r_H$. Here r_H stands for horizon radius which can be found by setting $f(r) = 0$ and solving for r .

iii) The Hawking temperature

The Hawking temperature of a black hole corresponds to the temperature of the black hole as inferred by the process of emittance of radiation from the horizon of the black hole. This quantum mechanical concept of emission of radiation from a black hole is explained in the following manner. Suppose a particle-antiparticle pair is created in vacuum due to vacuum

energy fluctuations near the horizon of a black hole and suppose one of them falls into the black hole, while the other just escapes from it. It will seem to any far away observer that a flow of parties is coming out of the black hole.

iv) Entropy

The Entropy of a black hole is a measure of all the information that falls into the black hole. As the black hole engulfs more matter, the information is not lost but is encrypted within the black hole as an increase in the surface area of the event horizon. Mathematically, the entropy of a black hole is calculated as

$$S = A/4 = \pi r_H^2/4.$$

Conclusion

Black holes are ultra-compact objects which can create a event horizon beyond which there is no escape. Black hole physics has been an active area of research, specially after their confirmation from the Gravitational wave detection by LIGO-Virgo collaboration in 2015. In the recent time, we have witnessed another astonishing achievement in this field in the form of first ever digital photograph of the event horizon of a black hole by the *Event Horizon Telescope (EHT)* which is an array of a number of telescopes in different regions of earth. Theoretical development has also gone parallel. It needs to be mentioned that many theories of gravity exists today and in order to determine the true theory of gravity, we need many filters and tests which will justify or falsify some of the theories. Since Einstein's theory of relativity has been infected with many problems, it is necessary to come up with meaningful alternatives and black hole physics can help us in this regard. With more sophisticated instruments and techniques, we will be able to know more secrets of the universe and understand it more deeply.

Acknowledgement

I would like to thank Prof Umananda Dev Goswami for useful discussions regarding black holes and general relativity. I would also like to thank Dr. Dhruba Jyoti Gogoi for his helpful discussions in the field of black hole physics. I would like to acknowledge Dibrugarh University Department of Physics for providing me with the facility to carry out my research related works there.

References

- [1] J. Bekenstein, (1973) Phys. Rev. D 7, 2333.
- [2] S. W. Hawking, (1976) Phys. Rev. D 13 191.
- [3] S. W. Hawking, (1972) Commun. Math. Phys. 25 152.
- [4] D. J. Gogoi and U. D. Goswami, (2022) JCAP 06 029.
- [5] D. J. Gogoi, R. Karmakar and U. D. Goswami, (2023) IJGMMP 20 2350007.

Propagation of UHE Cosmic Rays and its Signatures

Swaraj Pratim Sarmah

Department of Physics,
Dibrugarh University, Dibrugarh, India
Email: swarajpratimsarmah16@gmail.com

Abstract

The source of the ultrahigh energy cosmic ray is yet unknown. However, significant progress has been made in the last several years thanks to the high-quality data collected by the most recent cosmic ray observatories. One of the most recent successes is the Pierre Auger observatory's acquisition of reliable observational data regarding the extragalactic origin of the most intense cosmic rays. On the other hand, the intergalactic medium is thought to be filled with a non-null turbulent magnetic field. Therefore, the intergalactic magnetic field's presence may significantly impact how far ultra high energy cosmic rays travel through space, which may have implications for how to interpret experimental data. In this work, we study the diffusion effects on the propagation of UHECRs. We see that for the higher value of the magnetic field, source distance, and coherence length, the enhancement peak of the UHECRs propagation is decreased.

Introduction

Despite the tremendous advancements made recently, particularly in the area of experimental study, it is still unknown the origin of the ultrahigh energy cosmic rays (UHECRs, $E \sim 10^{18}$ eV) and what they are made of. The southern hemisphere-located Pierre Auger Observatory and the northern hemisphere located Telescope Array are now collecting very high-quality data. Additionally, the existing data are quite vast, particularly the one pertaining to Auger, who has currently amassed a significant amount of exposure. The UHECR flux has been measured by Auger and Telescope Array with previously unheard-of data. It has two primary characteristics: suppression and the so-called ankle stiffening at $10^{18.7}$ eV. Auger observed this suppression at an energy of $10^{(19.78 \pm 0.06)}$ eV [1]. Additionally, the values taken by the Auger spectrum are smaller than those corresponding to the Telescope Array. By adjusting both experiments' energy scales within their systematic errors, the differences between the two observations can be reduced. But the suppression region still shows some variances [1]. On the other hand,

S. P. Sarmah, *Propagation of UHE Cosmic Rays and its Signatures*, Horizon, (Ed. Book), 184-190 (2023), copyright Tinsukia College Publishing, ISBN: 978-81-19322-07-7

composition data indicate that cosmic rays are primarily light at EeV energies, consisting primarily of H and He, whereas, above a few EeV, their average mass grows progressively heavier [2]. The *Greisen-Zatsepin-Kuzmin effect* [3], or the attenuation they experience from photopion production off cosmic microwave background (CMB) photons, has been proposed as one scenario for the UHECRs. In this scenario, pure proton sources have maximum energies above a few hundred EeV [4]. In this instance, in the so-called dip scenario, the ankle feature has been linked to a propagation effect connected to the threshold for pair generation with the same background photons. The observed composition change is not explained in this scenario, even though the major characteristics of the spectrum may be explained naturally. Additionally, the source characteristics needed to accelerate protons to such high energies are highly demanding. The most energetic component of the UHECR flux is widely accepted to have an extragalactic origin. This is because galactic sources are known to be ineffective at accelerating particles at higher energies. Additionally, Auger has just recently discovered substantial observational evidence from the analysis of the distribution of the arrival directions of cosmic rays with primary energy exceeding $10^{18.9}$ eV concerning their extragalactic origin. A dipole-like anisotropy with an amplitude of about 6.5% was discovered [5] when this data set was taken into account. The significance of this discovery exceeds 5.2σ and the direction of the dipole rules out a scenario in which the flux is dominated by a galactic component. Alternatively, the transition between a sharply declining galactic component at lower energies and a developing harder extragalactic component has been linked to the spectral hardening at the ankle. The fundamental issue with this kind of scenario is that if those CRs were of galactic origin, a high anisotropy towards the galactic center and the galactic plane would be predicted, something that is not observed [6]. This is due to the comparatively light composition estimated at EeV energies. Furthermore, no natural Galactic source candidates exist to accelerate protons to energies below a few EeV, and the transition between sources that are primarily galactic and those that are extragalactic is more likely to occur at about 0.1 EeV. When the spectrum and composition measurements from the Pierre Auger Observatory are fitted in terms of evenly distributed extragalactic sources with power-law spectra [7], a lower cutoff value for the maximum cosmic-ray rigidities, $E/Z \sim 5$ EeV, is frequently chosen to increase the average mass above the ankle energy. One unsolved issue in high energy astrophysics is the change from galactic to extragalactic components. According to the Auger results, an isotropic flux is consistent with the broad-scale distribution of cosmic ray arrival directions in the energy range from 10^{18} eV to the ankle. This conclusion conflicts with the idea that the light component, which appears to dominate the flow in this energy range, has a cosmic origin [8]. Consequently, it is anticipated that the transition between these two components will occur below 10^{18} eV. As a result, Auger's large-scale arrival direction analysis and the composition data at lower energy demonstrate that the extragalactic component of the UHECR flux appears to be dominant.

The distance travelled by charged particles when subject to the field's effect concerning the scattering length is determined by the formula $l_D = 3D/c$, where c is the speed of light and D is nothing but the diffusion coefficient. The propagation is ballistic over travelled distances much smaller than l_D , which typically corresponds to the area near the source. The propagation

is diffusive for travelled lengths that are significantly greater than l_D .

We here to study the effects of diffusion and obviously the effect of source distance, coherence length, and magnetic field on the propagation of UHECRs.

Diffusive Propagation of UHECRs

It is challenging to develop a model for the extragalactic magnetic fields because there are so few empirical constraints on them [9]. They are investigated in galaxy clusters by measuring the Faraday rotation impact on the light of embedded and background galaxies as well as by radio emission from diffuse synchrotron sources, both close to and on the edges of the clusters. In the cluster center regions, these large-scale magnetic fields have recorded amplitudes that vary from a few to tens of μG [10]. This implies that enormous large-scale magnetic fields, with strengths that could vary from nG to μG , should also be present in cosmic structure filaments and sheets, however observations of these are yet missing. We'll take a look at a condensed description of the intergalactic magnetic field in this region, which could range from a few tens of nG to a few hundred nG, and having a coherence length with typical values of order $l_c = 0.01\text{--}1\text{ Mpc}$. By using a power law in Fourier space, one can characterize the distribution of the magnetic energy density on various length scales. We shall focus on the instance of a *Kolmogorov turbulence* spectrum where $m = 5/3$.

The critical energy E_c , defined such that, $rl(E_c) = l_c$, is a vital variable to describe particle diffusion and it is given by,

$$E_c = ZeBl_c \simeq 0.9Z \frac{B}{\text{nG Mpc}} EeV$$

Where Ze is the particle charge and r_L is the Larmor radius. A fit to the diffusion equation is given by [11],

$$D(E) = \frac{cl_c}{3} \left[4 \left(\frac{E}{E_c} \right)^2 + a_l \left(\frac{E}{E_c} \right) + a_L \left(\frac{E}{E_c} \right)^{2-m} \right].$$

For *Kolmogorov* spectrum for the turbulent magnetic field ($m = 5/3$), the resulting coefficients are $a_l \approx 0.23$ and $a_L \approx 0.9$, while for *Kraichnan* spectrum $m = 3/2$, $a_l \approx 0.42$ and $a_L \approx 0.65$. The diffusion length is defined as $l_D \equiv 3D/c$ and corresponds to the distance after which the total deflection of the particles is about 1 rad. In an expanding universe, the diffusion regimes obey the equation,

$$\frac{\partial n}{\partial t} + 3H(t)n - b(E, t) \frac{\partial n}{\partial E} - n \frac{\partial b}{\partial E} - \frac{D(E, t)}{a^2(t)} \nabla^2 n = \frac{Q(E, t)}{a^3(t)} \delta^3(r - r_s)$$

Where $H(t) = \frac{\dot{a}}{a}$ is the Hubble constant. The general solution of the above equation is given by [12,13],

$$n(E, r_s) = \int_0^z dz \left| \frac{dt}{dz} \right| Q(E, z) \frac{\exp[-r_s^2/4\lambda^2]}{(4\pi\lambda^2)^{3/2}} \frac{dE_g}{dE}$$

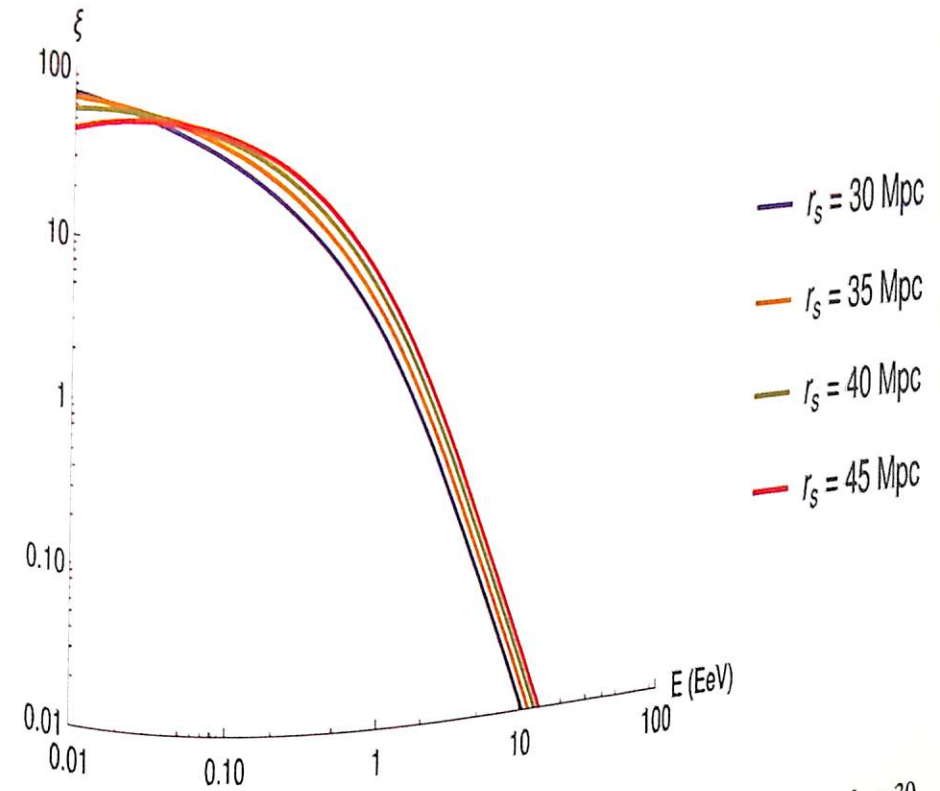


Fig. 1: Density Enhancement as a function of energy with a variety of source distances from 30 Mpc to 45 Mpc.

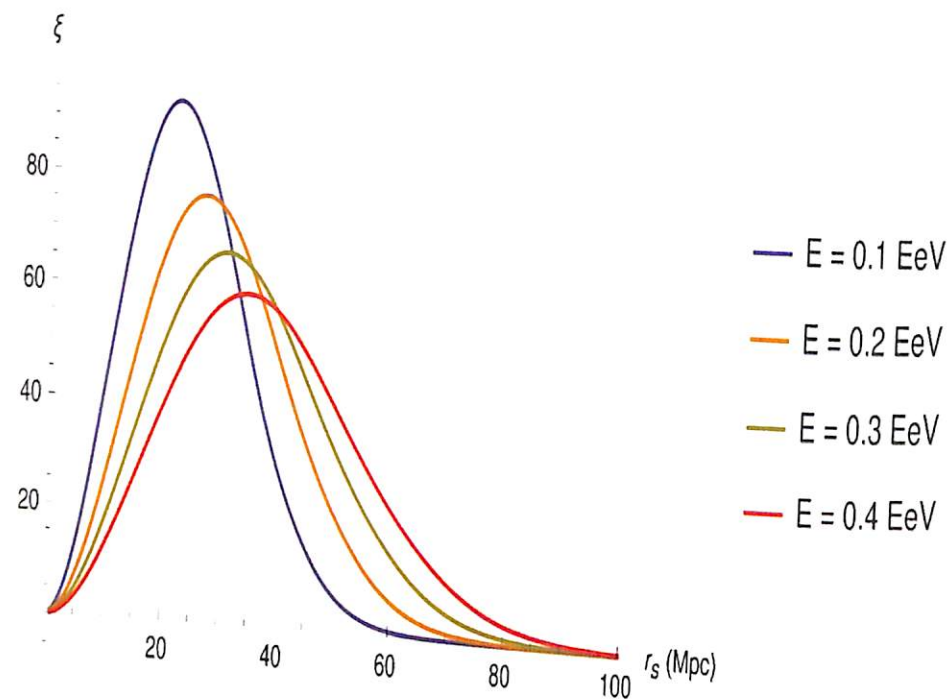


Fig. 2: Density Enhancement as a function of source distance with a variety of energy range from 0.1 EeV to 0.4 EeV.

Where λ^2 is the typical distance from where the cosmic ray diffuses from the sources and it given by,

$$\lambda^2(E, z) = \int_0^z dz \left| \frac{dt}{dz} \right| (1+z)^2 D(E)$$

Where dE_g/dE is the energy degradation and it is given by,

$$\frac{dE_g}{dE} = (1+z) \exp \left[\int_0^z dz \left| \frac{dt}{dz} \right| \left(\frac{\partial b_{int}}{\partial E} \right) \right].$$

In an expanding Universe,

$$\left| \frac{dt}{dz} \right| = \frac{1}{H_0 (1+z) \sqrt{(1+z)^3 \Omega_m + \Omega_\Lambda}}$$

We have considered here $H_0 = 67.4$ km/s. Mpc, $\Omega_m = 0.3$ km/s, and $\Omega_\Lambda = 0.7$. The enhancement factor is given as,

$$\xi = \frac{4\pi c r_s^2 n}{Q(E)}$$

In the next section we will discuss the corresponding results of the enhancement factor that will describe the propagation of UHECRs.

Results and Discussions

We have calculated the density enhancement of UHE cosmic rays as a function of energy and source distance and plotted them in Fig.1 and Fig.2 respectively. In Fig.1, we have seen that at about 0.04 EeV the enhancement is the same for all source distances value. And from that point, it gradually decreases. For higher value enhancement is lower. For all the calculation we have taken the critical energy value $E_c = 4.5$ EeV and $l_c = 0.1$ Mpc. In Fig.2, we have plotted the density enhancement as a function of source distance r_s and taking the energy value from 0.1 EeV to 0.4 EeV. We have seen that when $E = 0.1$ EeV the density enhancement peak is at the top. Then when we increase the energy value, then the peak gradually decreases but it gives a better distribution of density enhancement. So, we can say that the enhancement peak is inversely proportional to the energy value but its distribution is definitely proportional to the energy. Hence the source distance, energy, and magnetic field play a vital role in the description of the propagation of CRs particles.

References

- [1] D. Ivanov, PoS ICRC2017(2018) 498.
- [2] A. Aab et al. (Pierre Auger Collaboration), Phys. Rev. D 90 (2014) 122005; ibidem Phys. Rev. D 90 (2014) 122006.
- [3] K. Greisen, Phys. Rev. Lett. 16 (1966) 748; G.T. Zatsepin and V.A. Kuzmin, Pisma Zh. Exprim. Theor. Phys. 4 (1966) 114.
- [4] V. Berezhinsky, A.Z. Gazizov and S.I. Grigorieva, Phys. Lett. B 612 (2005) 147.
- [5] Pierre Auger collaboration, Science 357 (2017) 1266.
- [6] P. Abreu et al. (Pierre Auger Collaboration), Astrophys. J. Lett. 762 (2013) L13.
- [7] A. Aab et al. (Pierre Auger Collaboration), JCAP04 (2017) 038, Erratum: JCAP03 (2018) E02.
- [8] Pierre Auger collaboration, Astrophys. J. Suppl. 203 (2012) 34.
- [9] J. L. Han, Annual Review of Astron. and Astrophys. 55 (2017) 111.
- [10] L. Feretti et al., Astron. Astrophys. Rev. 20 (2012) 54.
- [11] D. Harari, S. Mollerach and E. Roulet, Phys. Rev. D 89 (2014) 123001.
- [12] V. Berezhinsky and A.Z. Gazizov, Astrophys. J. 643 (2006) 8.
- [13] V. Berezhinsky and A.Z. Gazizov, Astrophys. J. 669 (2007) 684.

Transition metal Mediated Oxidation of Alcohols using Environmentally Benign Oxidant

Biplab Banik

Department of Chemistry,
Tinsukia College,

Tinsukia, Assam, India

Email: biplab.tsk@gmail.com

Abstract

The transition metal catalysed selective oxidation of alcohols is another powerful protocol in organic chemistry because carbonyl compounds serve as important intermediates in the fine chemicals and pharmaceutical industries. Traditional oxidation methods produce considerable amounts of toxic waste with a high E-factor e.g. use of Cr(VI), Mn(VII) etc. in stoichiometric amount. With the ever-increasing environmental concerns, in the last few years, much attention has been paid to catalytic oxidation reactions in the presence of environmentally benign oxidant. Among the oxidants, hydrogen peroxide is considered as one of the most sustainable oxidants for industrial-scale applications because it produces water as the sole byproduct and it is of low cost. Over the years, there has been an exponential rise in the number of publications in alcohol oxidation reaction, but majority of these systems often required drastic conditions like high catalyst loading, high temperature, biphasic solvent system, use of additives, high concentration of base, etc. So, there is plenty of scope for development of efficient catalytic systems that can promote alcohol oxidation reactions with low metal loading under mild reaction condition. Hence in order to enlarge the scope of oxidation catalysts further, the catalytic activity of some new metal complexes are explored for alcohol oxidation reaction using H_2O_2 as oxidant.

B. Banik, "Transition metal mediated oxidation of alcohols using environmentally benign oxidant", Horizon (Ed. Book), 191-198(2023). copyright :Tinsukia College Publishing. ISBN: 978-81-19322-07-7

1. Introduction

Selective oxidations of primary and secondary alcohols to the corresponding carbonyl compounds is considered as one of the most useful and fundamental reactions in industrial processes due to the wide ranging utility of these products as precursors and/or intermediates for numerous drugs, vitamins and fragrances.^[1] Traditional methods involve small organic molecule-based reagents, e.g., Dess-Martin periodinane, Swern, Moffatt, Corey-Kim oxidants, $\text{SO}_3/\text{pyridine}$, some of them being moisture-sensitive or stoichiometric metal-based systems (such as Jones, Collins, Oppenauer reagents, pyridinium chlorochromate (PCC), pyridinium dichromate, barium permanganate, manganese dioxide, ruthenium tetroxide, silver carbonate).^[2-3]

2. Modern Oxidation of Alcohols using environmentally benign Oxidants

Standard organic text books^[4] still recommend classical oxidation methods using stoichiometric quantities of inorganic oxidants, which are highly toxic and environmentally polluting. Although such methods have broad scope, they are very atom inefficient and involve the use of hazardous reagents generating excess amounts of waste. Hence, the traditional oxidation methods are no longer acceptable and great efforts have recently been devoted to develop sustainable methods utilizing green oxidants and minimizing the effects of toxic wastes. An ideal oxidizing reagent should: (i) oxidize in good yield a broad variety of alcohols bearing different functionalities, (ii) have the ability to be used for scale-up procedures and (iii) be as green as possible, considering worker safety, ecotoxicity and atom economy.

Now-a-days, a chemical process for the production of a desired product has to be evaluated according to the atom economy principle established by Trost in 1991.^[5] Applying this concept to the field of oxidation reactions implies that oxidants with higher active oxygen percentage content, i.e., the mass amount of oxygen transferred to the substrate with respect to the total mass of the oxidant, should be privileged. A number of common oxidants are reported in Table 1, ordered according to their active oxygen content. It is evident that O_2 is the ideal oxidant to be used. Immediately after, with a 47% oxygen content H_2O_2 comes, which also releases 1 equivalent of H_2O . Both of these oxidants fulfill the requirement of "green chemistry" as they are inexpensive and produce water as the sole by-product. Consequently, there is an on-going quest to mediate the selective catalytic oxidation of alcohols using O_2 or H_2O_2 ^[6] as the primary environmentally friendly oxidant.

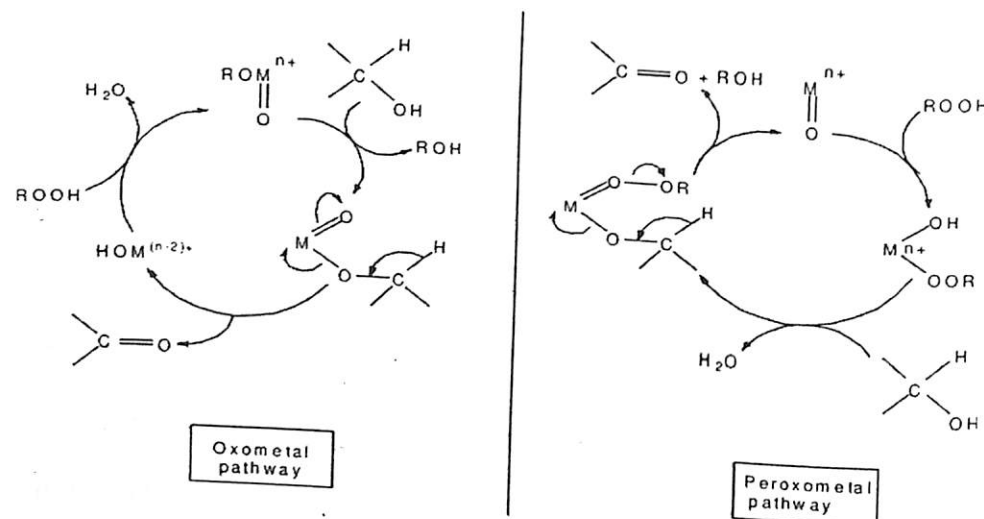
Table 1: Common oxidants ordered by active oxygen content

Oxidant	Active oxygen content (wt.%)	Waste product
O_2	100.0	Nothing or H_2O
$\text{O}_2/\text{redactor}$	50.0	H_2O
	192	

H_2O_2	47.0	H_2O
N_2O	36.4	N_2
O_3	33.3	O_2
KMnO_4	30.4	Mn(II) salts
HNO_3	25.0	NO_x
CrO_3	24.0	Cr(III) salts
NaOCl	21.6	NaCl
CH_3COOH	21.1	CH_3COOH
$t\text{BuOOH}$ (TBHP)	17.8	$t\text{BuOH}$
$\text{C}_5\text{H}_{11}\text{NO}_2$ (NMO)	13.7	$\text{C}_5\text{H}_{11}\text{NO}$ (NMM)
KHSO_5	10.5	KHSO_4
$m\text{-ClC}_6\text{H}_4\text{COOOH}$ ($m\text{-CPBA}$)	9.3	$m\text{-ClC}_6\text{H}_4\text{COOOH}$
NaIO_4	7.5	NaIO_3
PhIO	7.3	PhIO

2. Mechanism of Metal-Mediated Oxidation of Alcohols

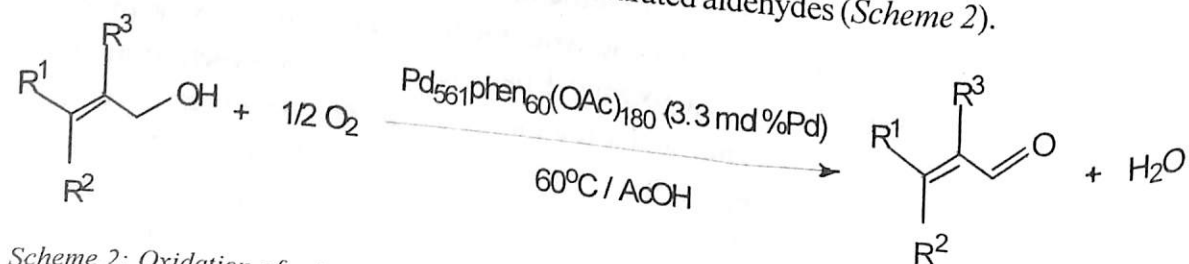
Metal-catalyzed oxidations of alcohols with peroxide reagents can be conveniently divided into two categories involving peroxometal and oxometal species, respectively, as the active oxidant (Scheme 1). In the peroxometal pathway, the metal ion remains in the same oxidation state throughout the catalytic cycle and the reaction involves the formation of a hydridometal species. In contrast, oxometal pathways involve a two-electron change in the oxidation state of the metal ion, and stoichiometric oxidation is observed, with the transition metal ions with ad^0 configuration while oxometal pathways are characteristic of late transition elements and first row transition elements and involve an oxometal intermediate.



Scheme 1: Oxometal versus peroxometal pathways in metal-catalyzed alcohol oxidations

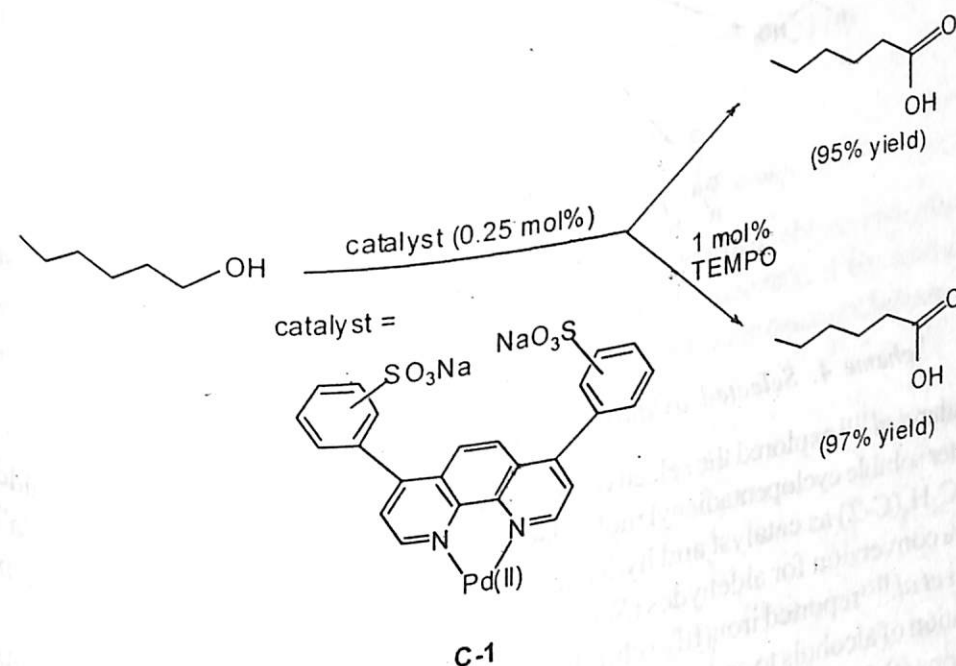
3. Oxygen as oxidant

An ideal oxidant for any large scale oxidation reaction should be easily accessible, cheap and non-toxic. As it happens, the best oxidant to fit this description is dioxygen.^[7] It is easily available since it is present in air and the only by-product produced from its decomposition is water. Oxygen is among the cheaper and less polluting stoichiometric oxidants, since it produces water as the sole by-product. The implementation of a catalyst in combination with molecular oxygen represents an emerging alternative to the traditional procedures. In the development of transition metal-catalyzed aerobic alcohol oxidations, several challenges exist, as the need of low pressures of O_2 especially in flammable organic solvents, mild reaction conditions, low catalyst loadings, and avoidance of costly or toxic additives. Blackburn and Schwarz^[8] first reported the $PdCl_2$ - $NaOAc$ -catalyzed aerobic oxidation of alcohols in 1977. However, activities were very low, with turnover frequencies of the order of $1\ h^{-1}$. Subsequently, much effort has been devoted to find synthetically useful methods for the palladium-catalyzed aerobic oxidation of alcohols. For example, the giant palladium cluster, $Pd_{561}phen_{60}(OAc)_{180}$, was shown to catalyze the aerobic oxidation of primary allylic alcohols to the corresponding α, β -unsaturated aldehydes (Scheme 2).



Scheme 2: Oxidation of primary allylic alcohols catalysed by palladium cluster, $Pd_{561}phen_{60}(OAc)_{180}$

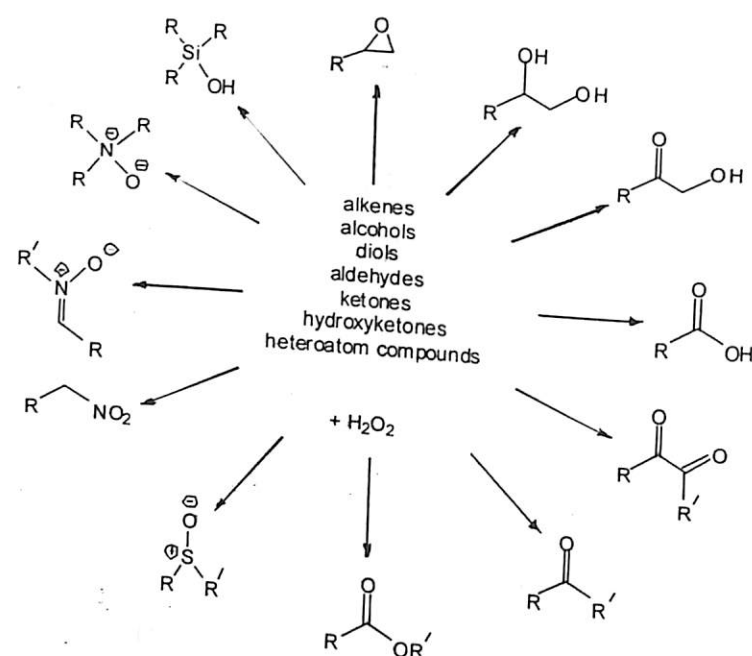
A much more active catalyst is represented by a water-soluble palladium(II) complex of sulfonated bathophenanthroline (C-1) introduced by Sheldon and co-workers.^[9] This stable, recyclable catalyst allowed oxidation in a two-phase aqueous-organic medium in 5 h at $100^\circ C/30\ bar$ air with 0.25 mol% catalyst (Scheme 3). Primary alcohols afforded the corresponding carboxylic acids via further oxidation of the initially formed aldehyde, e.g. 1-hexanol afforded 1-hexanoic acid in 95% yield. In contrast, when 1 mol% TEMPO was added the aldehyde was obtained in high yield; for example, 1-hexanol afforded 1-hexanal in 97% yield. (Scheme 3).



Scheme 3: Sheldon's Pd -catalyzed aerobic oxidation of alcohols

4. H_2O_2 as oxidant

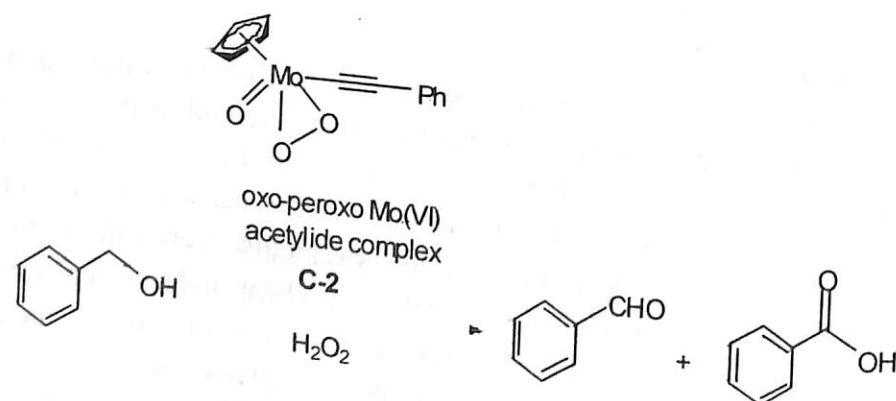
Although both O_2 and H_2O_2 stem as the oxidants of choice, O_2 is more difficult to be activated, requiring often harsh conditions of temperature and/or pressure. Moreover, oxidations with O_2 often occur with poor selectivity. For these reasons, H_2O_2 is a more practical oxidant and its use for organic transformations has been investigated intensively in the recent years. Hydrogen peroxide is a versatile oxidant discovered almost two centuries ago by the French chemist Louis Thenard. It is relatively cheap and is widely used in various industrial processes such as paper, pulp and textile, wastewater treatment, metallurgy and chemical synthesis. H_2O_2 is an effective oxidant with a high oxidation potential.^[10] Under appropriate conditions, it is able to perform a multitude of oxidative transformations by use of transition metal derivatives (Scheme 4).



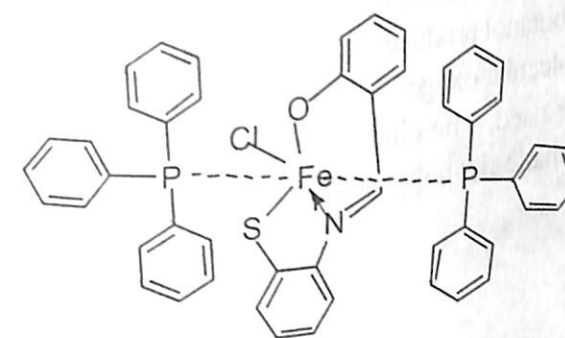
Scheme 4: Selected oxidative transformations performed by H_2O_2

Biradaret *et al.*^[11] explored the selective oxidation of various aromatic alcohols to aldehydes using water soluble cyclopentadienyl molybdenum acetylide complex, $CpMo(CO)_3(Ca''CPh)$; $Cp = \eta^5-C_5H_5$ (C-2) as catalyst and hydrogen peroxide as environmentally benign oxidant with 90% conversion for aldehydes (Scheme 5).

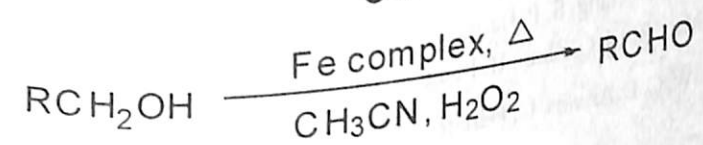
Rani *et al.*^[12] reported iron(III)-schiff base-triphenylphosphine complex (C-3) catalyzed the oxidation of alcohols to carbonyl compounds in presence of hydrogen peroxide with high conversions (Scheme 6).



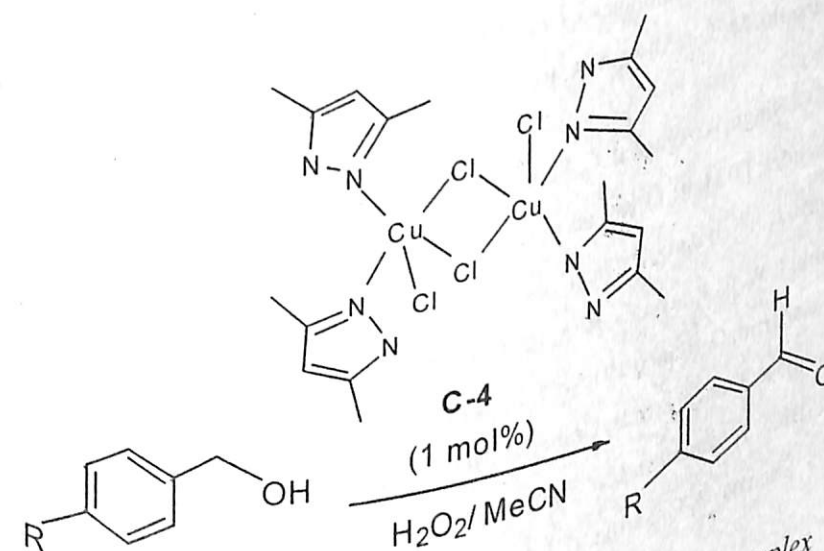
Scheme 5: Oxidation of aromatic alcohol by C-2 complex



C-3



Barooah *et al.*^[13] reported the dimeric copper(II) complex di- μ -chloro-bis[chloro(di-3,5-dimethylpyrazole)copper(II)] (C-4) as efficient catalyst for the oxidation of benzylic alcohols to their corresponding carbonyl compounds with hydrogen peroxide oxidant (Scheme 7).



Scheme 7: Oxidation of benzylic alcohols by C-4 complex

CONCLUSION

Oxidation of aliphatic alcohols with hydrogen peroxide in the presence of homogeneous catalysts shows low efficiency and is, therefore, not interesting for industrial applications. Oxidation with *tert*-butyl hydroperoxide gives much higher efficiencies in the presence of

homogeneous catalysts since this oxidant is more reactive. However, its higher price and the need to utilize the *tert*-butanol produced make this oxidant unattractive for industry. The ideal oxidant is certainly molecular oxygen, which is cheap and quite selective if temperatures not higher than 343K are used. The challenge is to find active catalysts for the oxidative transformations of aliphatic alcohols, which need to be further studied.

References

- [1] (a) R. A. Sheldon, I. W. C. E. Arends, A. Dijkstra, *Catal. Today* 2000, 57, 157-166.
(b) J. Zhang, S. Lu, Y. Xiang, S. P. Jiang, *ChemSusChem*, 2020, 13, 2484-2502. (c) D. Tang, G. Lu, Z. Shen, Y. Hu, L. Yao, B. Li, G. Zhao, B. Peng, X. Huang, *J. Ener. Chem.* 2023, 77, 80-118.
- [2] R. A. Sheldon, I. W. C. E. Arends, U. Hanefeld, *Green Chemistry and Catalysis*, Wiley-VCH, Weinheim, Germany, 2007.
- [3] S. Albonetti, R. Mazzoni, F. Cavani, in *Transition Metal Catalysis in Aerobic Alcohol Oxidation*, (Eds: F. Cardona, C. Parmeggiani), ch. 1, 2014, pp. 1-39.
- [4] J. Clayden, N. Greeves, S. Warren, P. Wothers, *Organic Chemistry*, Oxford Univ. Press, New York, 2001.
- [5] B. M. Trost, The atom economy: a search for synthetic efficiency, *Science* 1991, 254, 1471-1477.
- [6] (a) O. A. Kholdeeva, O. V. Zalomaeva, *Coord. Chem. Rev.* 2016, 306, 302-330.
(b) C. B. Vilanculo, M. J. daSilva, A. A. Rodrigues, S. O. Ferreira, R. C. DaSilva, *RSC Adv.* 2021, 11, 24072-24085.
(c) R. Goyal, O. Singh, A. Agarwal, C. Samanta, B. Sarkar, *Catal. Rev.* 2022, 64, 229-285.
- [7] R. Ray, S. Chandra, D. Maiti, G. K. Lahiri, *Chem. Eur. J.* 2016, 22, 8814-8822.
- [8] T. F. Blackburn, J. J. Schwartz, *J. Chem. Soc., Chem. Commun.* 1977, 157-158.
- [9] G. -J. ten Brink, I. W. C. E. Arends, R. A. Sheldon, *Adv. Synth. Catal.* 2002, 344, 355-369.
- [10] J. M. Campos-Martin, G. Blanco-Brieva, J. L. G. Fierro, *Angew. Chem. Int. Ed.* 2006, 45, 6962-6984.
- [11] A. V. Biradar, M. K. Dongare, S. B. Umbarkar, *Tetrahedron Lett.* 2009, 50, 2885-2888.
- [12] S. Rani, B. R. Bhat, *Tetrahedron Lett.* 2010, 51, 6403-6405.
- [13] N. Barooah, S. Sharma, B. C. Sarma, J. B. Baruah, *Appl. Organometal. Chem.* 2004, 18, 440-445.

Synthesis of Some Natural and Synthetic Adhesives and Their Applications

Saswati Phookan⁽¹⁾
Shauvik Chandra Das

⁽¹⁾Department of Chemistry,
Tinsukia College, Tinsukia, Assam
Email: saswatiphookan1973@gmail.com

Abstract

Adhesives are those substances that are applied to surfaces of two separate items to bind them together. They may be either found naturally or produced synthetically. There are several advantages as well as disadvantages of adhesives. Adhesives have been in use since around 4000 BC and from then it had come along a long way to the present day adhesives. A tremendous change can be seen from the use of natural adhesives to the synthetic ones. Different forms of adhesives are available such as resins, rosins, mucilage, glues and gums. Based on the source from where they are obtained, polymers can be natural and synthetic polymers. Adhesives are formed by polymerization reactions, i.e. either by addition polymerization or condensation polymerization. Adhesives are classified into two broad categories, viz. — Natural glues and Synthetic resins. The study carried out was on synthesis of Protein glues (a sub-category of natural glues) and Amino-plastic resins (a sub-category of synthetic resins). Three natural protein-based glues were prepared— Blood Albumin Glue, Casein Glue and Soya bean Glue and three synthetic amino-based resins were prepared— Urea Formaldehyde resins, Melamine Formaldehyde resins and Melamine Urea Formaldehyde resins. Along with the preparation of the six types of adhesives their applications were also tested on five different types of materials, viz. — cloth, glass, paper, plastic and wood. The properties, applications, advantages and disadvantages of natural glues and synthetic resins have also been discussed.

S. Phookan, S. C. Das, *Synthesis of Some Natural and Synthetic Adhesives and Their Application Horizon*, (Ed. Book), 199, 223 (2023), copyright : Tinsukia College Publishing, ISBN: 978-81-19322-07-7

1. Introduction

Adhesives may be found naturally or produced synthetically. Any substance applied to one surface or both surfaces, of two separate items that binds them together and resists their separation is termed as an adhesive. Adhesives are also known as glue, cements, mucilage or paste. Prior to discovery of synthetic resins, all wood joints including plywood were done with adhesives from materials of natural origin- both plants and animals. Vegetable glue based on starch or soyabean glue based on soya-protein is obtained from plants. Glues like animal, casein or blood albumin have their origin from animals⁽¹⁾.

The principal constituent of vegetable glue is starch which may be extracted commercially from corn, potato, rice, sago, tapioca, cassava and many other plant sources. The protein in soya seed meal from which oil has been removed forms the basis of soya bean glue. Casein glue is prepared from casein curd which is precipitated milk either by natural souring or by addition of acids. Animal glue obtained from skin, bones and blood albumin glue obtained from animal blood are two very important adhesives used in wood industries before the introduction of synthetic adhesives⁽¹⁾.

The use of adhesives offers many advantages over binding techniques such as sewing, mechanical fastening, thermal bonding, etc. ⁽¹⁾.

Disadvantages of adhesive use include decreased stability at high temperatures, relative weakness in binding large objects with a small bonding surface area and greater difficulty in separating objects during testing. Adhesives are typically organized by the method of adhesion⁽¹⁾.

Adhesives are substances and can be defined as a mixture in a liquid or semi-liquid state, capable to join permanently to surfaces, by an adhesive process. The word 'adhesive' can be defined as substances which tend to adhere or stick to other substances. Adhesion refers to the interaction of the adhesive surface with the substrate surface, and it involves two dissimilar bodies being held in intimate contact such that mechanical force or work can be transferred across the interface⁽²⁾.

Several theories that explain the adhesion process have been postulated, the forces involved in the process being van der Waals forces, chemical bonding, or electrostatic attraction. On the other hand, the mechanical strength of the system depends not only on the interfacial forces but also on the mechanical properties of the interfacial area, as well as the two bulk faces⁽²⁾.

The earliest recorded use of adhesives was before 4000 BC when mud and animal manure were used. The first references of adhesives in literature first appeared in 2000 BC. The Greeks and Romans made great contributions to the development of adhesives. The Egyptians used bituminous cements and resins with great effect in mummification and other applications. Somewhere between 1500 BC and 1300 BC, casein, starch, sugar, fish and animal glues came into existence⁽¹⁾.

Among all the glues of natural origin, animal and vegetable proteins were the only new options from ancient times down through the industrial revolution and nearly to the present.

About 100 AD, egg white and lime combinations were reported. By 1674, patent concerning fish glue has been issued in England. During 1690, the first commercial glue factory was established in Holland which produced animal glues. In 1800, the first commercialization of casein occurred in both Switzerland and Germany. The first commercial use of natural rubber in adhesives was in 1823. By 1865, plywood patents were beginning to appear, and Mayo in the US received the first patent for these materials. Water proof casein adhesives appeared in 1892. In 1897, Goldschmidt obtained patent for urea formaldehyde (UF) resin. The first announcement of phenolic resin was noted in England in 1901 (phenol formaldehyde resins- Bakelite). In 1912, early patent appeared for both polyvinyl acetate and acrylic resins. In 1919, paper and cloth began to be used as adhesive carriers in film form. This opened the way for laminated plastic materials. In 1920, urea formaldehyde (UF) resin became commercial. In 1921, phenol furfural resins were developed. In 1922, melamine resins were developed. Although the formaldehyde based thermosetting resins are used most frequently, in present time other thermosetting resins are also used such as epoxies, urethanes, polyamides, poly sulphides, block polymers and many others⁽¹⁾.

Resins are any natural or synthetic organic compound consisting of a non-crystalline or viscous liquid substance⁽³⁾. It is a viscous hydrocarbon secretion of many plants, particularly coniferous trees, produced by injury and breakdown of cell wall. Industrially used adhesives are also called resins⁽¹⁾. Natural resins are typically fusible and flammable organic substances that are transparent or translucent and are yellowish to brown in colour. They are formed in plant secretions and are soluble in various organic liquids but not in water. Most natural resins are exuded from trees, especially pines and firs. Synthetic resins comprise a large class of synthetic products that have some of the physical properties of natural resins but are different chemically. In modern industry natural resins have been almost entirely replaced by synthetic resins, which are divided into two classes, thermoplastic resins, which remain plastic after heat treatment, and thermosetting resins, which become insoluble and infusible on heating⁽³⁾.

Rosin, also known as Colophony or Colophonium are translucent, brittle and friable resin which is used for varnishing and in manufacturing different variety of products⁽⁴⁾. It is a solid-form of resin that is obtained from liquid resin by vaporizing its volatile components⁽¹⁾. It becomes sticky when it is warm and has a faint pine-like odor. Gum rosin consists of the residue that is obtained upon distillation of the oleoresin (a natural fluid) from pine trees (the volatile component is the spirit of turpentine). Wood rosin which is obtained by solvent extraction of the stumps is usually of a darker colour. Rosin and its chemical derivatives are used chiefly to make soaps, varnishes, printing inks, sealing wax, sizes for paper, adhesives, soldering fluxes, driers, gloss oils for paints, pitch for casks and binders. Rosin is also used on bows of guitars and other stringed instruments, on the shoes of dancers, and on floors of studios and stages to prevent slipping⁽⁴⁾.

Gum-resins are a mixture of both gums and resins and possess the properties of both the groups. They contain traces of essential oils⁽⁵⁾. Gums are formed upon bacterial and fungal

infections in plants resulting in breakdown of cell-walls (extracellular)⁽¹⁾. These are usually derived from the plant growing in dry and arid regions. Some of the commonly used gum-resins are asafetida, myrrh, salai, guggul, etc.⁽⁵⁾.

The Mucilage is a water-soluble edible adhesive material that constitutes carbohydrates and uranic acids units present in different parts of plants including the mucous epidermis of the outer layer of seeds, bark, leaves, and buds⁽⁶⁾. Mucilage are normal physiological product of metabolism formed within the cell (intracellular)⁽¹⁾. The majority of plants produce mucilage from the seed coat and this process of producing mucilage is termed Myxospermy and some plant species produce it from the fruit epicarp which is known as Myxocarpy. Plants producing mucilage from seed coat belong to the family *Plantaginaceae*, *Acanthaceae*, *Linaceae*, and *Brassicaceae*, while Myxocarpy (fruit mucilage) is commonly found in families like *Poaceae*, *Asteraceae*, and *Lamiaceae*. The presence of mucilage on the seed coat prevents the plant from early seedling development and drought stress during the germination. Depending upon its origin it is characterized into many groups including hair secretion, intracellular mucilage, and cell membrane mucilage. The mucilage obtained from the seed coat is classified into three classes that are endosperm non-starch polysaccharide (galactomannans), cell wall material of the endosperm (soybean hemicelluloses and xyloglucans), and mucilaginous constituents of the seed coat (flaxseed, Chia seed, and yellow mustard). Mucilage develops a jelly-like structure around fruit and maintains moisture and prevents seeds from completely drying out and therefore acts as a hydrating agent and also acts as an energy reservoir. Mucilage also plays a significant role in the control of germination, the promotion of dispersal, and soil adhesion, root mucilage is usually exhibited from the outer layers of the root cap, consisting of mostly root border cells and polysaccharides, which produce various chemical substances such as flavonoids, phenolics acids, amino acids, galactosidase, antibiotics, sugars, peroxidase, proteins, and anthocyanins. Moreover, root mucilage plays a very important role for plant growth, such as for the maintenance of root-to-soil touch, root tip lubrication; soil micro-aggregate stabilization, water storage ability, selective storage, and the absorption of ions (Na^+ , Cd^{2+} , Pb^{2+} , and Al^{3+}) through root cells. Furthermore, it is primarily secreted by the secretory vesicles of hyper-secretory root cap cells as a coagulated polysaccharide (poly-galacturonic acid) and is subsequently passed during root extension to older root areas, but epidermal cells are also effective in secreting mucilage. Mucilage is also produced in the leaves and buds of several plant species; it may allow the leaves to retain water capacity when soil water deficits emerge; therefore, it helps in the storage of food and water⁽⁶⁾.

Glue is a gelatin-like adhesive substance extracted from animal tissue, particularly hides and bones, or from fish, casein (milk solids), or vegetables. Glue was used as early as 3000 BCE in wooden furniture construction in Egypt. Synthetic resin adhesives such as the epoxies are replacing glue for many uses, but glue is still widely used as an adhesive in woodworking, in the manufacture of such abrasives as sandpaper, and as a colloid in industrial processes—e.g., the recovery of solid particles suspended in a liquid⁽⁷⁾.

Polymers are any of a class of natural or synthetic substances composed of very large molecules, called macromolecules that are multiples of simpler chemical units called monomers. Polymers make up many of the materials in living organisms, for example, proteins, cellulose, and nucleic acids. Moreover, they constitute the basis of minerals such as diamond, quartz, and feldspar and some man-made materials such as concrete, glass, paper, plastics, and rubbers. The word *polymer* designates an unspecified number of monomer units. When the number of monomers is very large, the compound is sometimes called a high polymer. Polymers are not restricted to monomers of the same chemical composition or molecular weight and structure. Some natural polymers are composed of one kind of monomer. Most natural and synthetic polymers, however, are made up of two or more different types of monomers; such polymers are known as copolymers⁽¹⁰⁾.

Polymers can be classified into different types depending on various factors. Depending on the temperature, polymers can be divided into two types—*Thermoplastics* and *Thermosettings*.

Thermoplastics soften when heated and harden when cooled. This is totally reversible and repeatable. Most linear polymers and branched structure polymers with flexible chains are thermoplastics. They undergo purely physical changes on heating. Examples may be given of Polythene, Poly Vinyl Chloride (PVC), etc.^(8,9).

The settings is in contrast to thermoplastic polymers, which do not soften when heated due to strong covalent crosslink. Thermosetting polymers are generally harder and stronger than thermoplastics and have better dimensional stability. They undergo chemical change on heating and the change is irreversible. Examples can be cited of Urea Formaldehyde Resins (UF Resins), Melamine Formaldehyde Resins (MF Resins), etc.^(8,9).

Depending on the type of monomer units present and its pattern of repetition, polymers can be divided into two types—*Homopolymers* and *Copolymers*.

When a single type of monomer undergoes polymerization to form a macromolecule, it is referred to as a homopolymer. A homopolymer basically has repeating units of a single type of monomer. For example, polystyrene is a homopolymer with repeating units of styrene monomers. Homopolymers are formed through a polymerization technique that is referred to as 'addition polymerization'. In this process, monomers must have either a single or double bond. In most cases, the 'poly' in homopolymers is used as a prefix which is then followed by the chemical name of the repeating unit. E.g. repeating units of 'vinyl chloride' monomers forms polyvinyl chloride polymer. Homopolymer thermoplastics are widely used in injection molding. Some of the most common examples include—Nylon 6, Nylon 11, Polypropylene, Polycarbonate, Polyester (PET), PBT, LDPE, HDPE, Polyethylene, Polytetrafluoroethylene, Poly methyl methacrylate (PMA), Polyvinyl chloride (PVC), Polyacrylonitrile (PAN), etc. The homopolymers can be applied to various industries. They are also used in different processing technologies, such as injection molding, blow molding, sheet extrusion, and thermoforming⁽¹¹⁾.

A polymer that is formed through the polymerization of more than one type of monomer is referred to as a copolymer. Thus, a copolymer always has two or more types of repeating

units. Usually, most copolymers are formed through a process called condensation polymerization. Some examples of Copolymers are—Ethylene-vinyl acetate, Rubber-styrene graft copolymers (high-impact polystyrene), Acrylonitrile-butadiene-styrene graft copolymer (ABS), Vinyl acetate-maleic anhydride-copolymers, Styrene-butadiene rubber, Styrene-acrylonitrile rubber, Styrene-butadiene di-block copolymers, Styrene-butadiene-styrene tri-block copolymers, Polyurethane multi-block copolymers (elastomeric yarns), etc. Some of the applications of Copolymers are given here—Hygroscopic and hydrophilic copolymers are applied to manufacture self-polishing paint applications. Acrylic copolymers are applied to manufacture various cosmetic products such as nail polish, eyeliners, lipsticks, etc. Graft polymers are used to manufacture certain types of membranes that can separate gas and liquids. Spandex used as synthetic fiber is a block copolymer. Block polymers such as polyester and polyamide are often applied to manufacture sports goods and automotive components. Nylon -6,6- is frequently used in textiles, and carpeting, and as an engineering material in bearings and gears due to its self-lubricating properties⁽¹¹⁾.

On the basis of the source from where they are obtained, the polymers can be classified into three types- i) Natural Polymers, ii) Synthetic Polymers and iii) Semi-synthetic Polymers. Natural polymers are found in nature, i.e. plants and animals. They are of two types—Organic and Inorganic polymers. Organic polymers play a crucial role in living things, providing basic structural materials and participating in vital life processes. For example, the solid parts of all plants are made up of polymers. These include cellulose, lignin, and various resins. Cellulose is a polysaccharide, a polymer that is composed of sugar molecules. Lignin consists of a complicated three-dimensional network of polymers. Wood resins are polymers of a simple hydrocarbon, isoprene. Another familiar isoprene polymer is rubber. Other important natural polymers include the proteins, which are polymers of amino acids, and the nucleic acids, which are polymers of nucleotides—complex molecules composed of nitrogen-containing bases, sugars, and phosphoric acid. The nucleic acids carry genetic information in the cell. Starches, important sources of food energy derived from plants, are natural polymers composed of glucose. Many inorganic polymers also are found in nature, including diamond and graphite. Both are composed of carbon. In diamond, carbon atoms are linked in a three-dimensional network that gives the material its hardness. In graphite, used as a lubricant and in pencil “leads,” the carbon atoms link in planes that can slide across one another⁽¹⁰⁾.

Synthetic polymers are produced in different types of reactions. Many simple hydrocarbons, such as ethylene and propylene, can be transformed into polymers by adding one monomer after another to the growing chain. Polyethylene, composed of repeating ethylene monomers, is an addition polymer. It may have as many as 10,000 monomers joined in long coiled chains.

Polyethylene is crystalline, translucent, and thermoplastic—i.e., it softens when heated. It is used for coatings, packaging, molded parts, and the manufacture of bottles and containers. Polypropylene is also crystalline and thermoplastic but is harder than polyethylene. Its molecules may consist of from 50,000 to 200,000 monomers. This compound is used in

the textile industry and to make molded objects. Other addition polymers include polybutadiene, polyisoprene, and polychloroprene, which are all important in the manufacture of synthetic rubbers.

Some polymers, such as polystyrene, are glassy and transparent at room temperature, as well as being thermoplastic. Polystyrene can be colored any shade and is used in the manufacture of toys and other plastic objects. If one hydrogen atom in ethylene is replaced by a chlorine atom, vinyl chloride is produced. This polymerizes to polyvinyl chloride (PVC), a colourless, hard, tough, thermoplastic material that can be manufactured in a number of forms, including foams, films, and fibers.

Vinyl acetate, produced by the reaction of ethylene and acetic acid, polymerizes to amorphous, soft resins used as coatings and adhesives. It copolymerizes with vinyl chloride to produce a large family of thermoplastic materials. Many important polymers have oxygen or nitrogen atoms, along with those of carbon, in the backbone chain. Among such macromolecular materials with oxygen atoms are polyacetals. The simplest polyacetal is polyformaldehyde. It has a high melting point and is crystalline and resistant to abrasion and the action of solvents. Acetal resins are more like metal than are any other plastics and are used in the manufacture of machine parts such as gears and bearings.

A linear polymer characterized by a repetition of ester groups along the backbone chain is called a polyester. Open-chain polyesters are colourless, crystalline, thermoplastic materials. Those with high molecular weights (10,000 to 15,000 molecules) are employed in the manufacture of films, molded objects, and fibers such as Dacron. The polyamides include the naturally occurring proteins casein, found in milk, and zein, found in corn (maize), from which plastics, fibers, adhesives, and coatings are made. Among the synthetic polyamides are the urea-formaldehyde resins, which are thermosetting. They are used to produce molded objects and as adhesives and coatings for textiles and paper. Also important are the polyamide resins known as nylons. They are strong, resistant to heat and abrasion, noncombustible, and nontoxic, and they can be colored. Their best-known use is as textile fibers, but they have many other applications.

Another important family of synthetic organic polymers is formed of linear repetitions of the urethane group. Polyurethanes are employed in making elastomeric fibers known as spandex and in the production of coating bases and soft and rigid foams.

A different class of polymers is the mixed organic-inorganic compounds. The most important representatives of this polymer family are the silicones. Their backbone consists of alternating silicon and oxygen atoms with organic groups attached to each of the silicon atoms. Silicones with low molecular weight are oils and greases. Higher-molecular-weight species are versatile elastic materials that remain soft and rubbery at very low temperatures. They are also relatively stable at high temperatures.

Fluoro-carbon-containing polymers, known as fluoro-polymers, are made up of carbon-fluorine bonds, which are highly stable and render the compound resistant to solvents. The nature of carbon-fluorine bonding further imparts a nonstick quality to fluoro-polymers; this is most widely evident in the polytetrafluoroethylene (PTFE) Teflon⁽¹⁰⁾.

Semi-synthetic polymers are generally derived from naturally occurring polymers by chemical modifications. Examples can be cited of chemically modified natural polymers, vulcanized rubber, etc. ⁽⁸⁾.

Polymerization is a process in which relatively small molecules, called monomers, combine chemically to produce a very large chainlike or network molecule, called a polymer. The monomer molecules may be all alike, or they may represent two, three, or more different compounds. Usually at least 100 monomer molecules must be combined to make a product that has certain unique physical properties—such as elasticity, high tensile strength, or the ability to form fibers—that differentiate polymers from substances composed of smaller and simpler molecules; often, many thousands of monomer units are incorporated in a single molecule of a polymer. The formation of stable covalent chemical bonds between the monomers sets polymerization apart from other processes, such as crystallization, in which large numbers of molecules aggregate under the influence of weak intermolecular forces ^(8, 12).

Two classes of polymerization usually are distinguished. In condensation polymerization, each step of the process is accompanied by the formation of a molecule of some simple compound, often water. In addition polymerization, monomers react to form a polymer without the formation of by-products. Addition polymerizations usually are carried out in the presence of catalysts, which in certain cases exert control over structural details that have important effects on the properties of the polymer.

Linear polymers, which are composed of chainlike molecules, may be viscous liquids or solids with varying degrees of crystallinity; a number of them can be dissolved in certain liquids, and they soften or melt upon heating. Cross-linked polymers, in which the molecular structure is a network, are thermosetting resins (i.e., they form under the influence of heat but, once formed, do not melt or soften upon reheating) that do not dissolve in solvents. Both linear and cross-linked polymers can be made by either addition or condensation polymerization ^(8, 12).

Table 1: Classification of Adhesives into broad categories ⁽¹⁾.

ADHESIVES				
NATURAL GLUES			SYNTHETIC RESINS	
Animal Glue	Protein Glue	Starch Glue	Synthetic Wood Adhesives	Other Synthetic Wood Adhesives
• Hide glue • Bone glue	• Blood glue • Casein glue • Soyabean glue	Derived from starch	(further classification is given below)	(further classification is given below)

Table 2: Classification of Synthetic Wood Adhesives ⁽¹⁾.

SYNTHETIC WOOD ADHESIVES			
Synthetic Phenolic	Natural Phenolic	Amino plastic	Epoxy Resins
• Phenol Formaldehyde (PF) • Resorcinol Formaldehyde (RF) • Copolymer, i.e. Resorcinol Phenol Formaldehyde (RPF)	• Tannin Formaldehyde (TF) • Copolymer, i.e. Tannin Resorcinol Formaldehyde (TRF) • Copolymer, i.e. Lignin Phenol Formaldehyde (PLF) • Copolymer, i.e. Cardanol Phenol Formaldehyde (CPF)	• Urea Formaldehyde (UF) • Melamine Formaldehyde (MF) • Copolymer, i.e. Melamine Urea Formaldehyde (MUF)	

Table 3: Classification of Some Other Synthetic Wood Adhesives ⁽¹⁾.

SOME OTHER SYNTHETIC WOOD ADHESIVES		
Poly Vinyl Acetate (PVA)	Rubber based	Di-isocyanates or Poly Urethane Derivatives.
➤ Highly cross-linked ➤ Normal PVA	➤ Structural elastomeric ➤ Contact adhesives ➤ Poly Vinyl Chloride (PVC)	

2. Objectives

To synthesize different kinds of adhesives - both natural and synthetic and to study the use and application of these adhesives in different items.

2.1 Resins Prepared:

A total of six different types of adhesives were prepared. Out of these, three were natural adhesives and the other three were synthetic adhesives.

2.1.1 Blood Albumin Glue (BAG):

Natural Adhesives prepared were
The protein present in serum albumin, globulin and red cell hemoglobin is utilized as adhesive. Animal blood obtained from slaughter houses is the major sources for blood glue. Fresh blood was used in pre-historic days as protein in blood gets denatured on storing. Technique for storing dried blood without denaturing protein content was developed during the beginning of 19th century. Blood is spray dried into powder which can be stored for a longer period of time ⁽¹⁾.

2.1.2 Casein Glue (CG): Casein precipitated from milk has been recognized as adhesive since ancient time. Casein protein is recovered in two ways—

i. Either from skim milk by acid precipitation using either organic acids such as vinegar or mineral acids such as sulphuric acid.

Skim Milk + Acids $\rightarrow$ Casein

ii. The milk may be cultured with bacteria that convert lactose (milk sugar) to lactic acid which in turn precipitate the casein.

Milk (lactose) $\rightarrow$ lactic acid $\rightarrow$ casein (ppt.)

The precipitate is washed with water to remove acid, dried and ground. Adhesive capability of casein glue depends on its dispersion during glue precipitation from powdered casein. Two factors influence the dispersion of casein powder in aqueous solution, viz. particle size and alkali content in dispersion⁽¹⁾.

2.1.3 Soyabean Glue (SG): Soyabean is an important agricultural crop in India and many other countries in temperate region. Soyabean is a good source of triglyceride oil and edible protein. It contains substantial quantity of protein ranging from 35-55% in the oil free soya meal. For preparation of wood adhesive from soyabean, the beans are de-hulled and oil is removed by crushing at high pressure or by solvent extraction. It is then heated and crushed into powder and then treated with strong alkali like sodium hydroxide (NaOH) or trisodium phosphate (Na_3PO_4) to break the internal hydrogen bond of the coiled protein molecules. Strong alkali dispersed soya meal when used for wood bonding, gives reddish-brown stain on wood surface as they cure. To get colorless glue line, water dispersed soya meal may be hydrolyzed by mild alkali like hydrated lime or ammonia. However, the bond obtained with strong alkali dispersed soya meal is always stronger than that obtainable from adhesive when soya meal is dispersed with mild alkali⁽¹⁾.

Synthetic Adhesives prepared were—

2.1.4 Urea Formaldehyde (UF) resins: The major constituents of the resin are urea and formalin (-37% formaldehyde). Formalin (F) and urea (U) react in varying ratios of parts by weight from 2.0:1 -2.8:1. Resins with low F: U ratios (2.0:1 -2.4:1) have low formaldehyde emissions, but have higher viscosity, lower water tolerance, lower strength and stiffness compared to resins made with higher F: U ratios (2.5:1 -2.8:1).

The polymerization reaction of urea formaldehyde resin takes place in two stages. The first stage is an addition polymerization in which urea and formaldehyde undergo addition reaction giving rise to mono, di- and trimethylolureas.

Formation of tetramethylol urea has not been reported due to steric hindrance of the already existing methylol groups. The reaction takes place in neutral or slightly alkaline condition at pH 7.5-8.0 at elevated temperatures 90-100°C. The reaction depends on concentration of the ingredients and their ratio, pH and reaction temperature. In the first stage the pre-polymers those are formed have no adhesive properties. Methylol ureas formed in the first stage undergoes condensation polymerization at pH 5.0-6.0 and temperature at 90-95°C to form long chain and cross-linked polymers through methylene linkages. The reaction is allowed to proceed to the desired viscosity and the reaction is arrested midway by raising pH of the resin solution to 7.5-8.0. If the reaction is allowed to continue, the chain and cross-linking will continue and will form a gel which becomes solid on cooling. The most important factors determining the

properties of the reaction products are—

- The relative molar ratio of urea and formaldehyde,
 - Reaction temperature,
 - pH values at which reaction is carried out.
- These factors influence the rate of increase of the molecular weight, solubility, viscosity, water tolerance and rate of curing of resin.

Water tolerance is a measure of the number of times by weight of water which can be mixed with resin before incipient precipitate is formed⁽¹⁾.

2.1.5 Melamine Formaldehyde (MF) resins: The structure of melamine is shown in Fig 2. Melamine resins are chemically related to urea resins, but are superior in performance, having greater water resistance and stability at high temperatures. The major constituents of this type of resins are formalin (-37% formaldehyde) and melamine in the ratio of 2:1 by parts weight. Melamine is a white crystalline powder with melting point 250°C. Melamine molecule has three reactive amino groups with 6 replaceable hydrogen atoms attached to nitrogen atoms. With analogy to urea formaldehyde its possible that hexamethylol compound can be formed by reaction of formaldehyde with melamine. But in fact, under normal conditions, trimethylol compound is formed as the entry of second methylol group to a nitrogen atom is hindered by the already existing first methylol group.

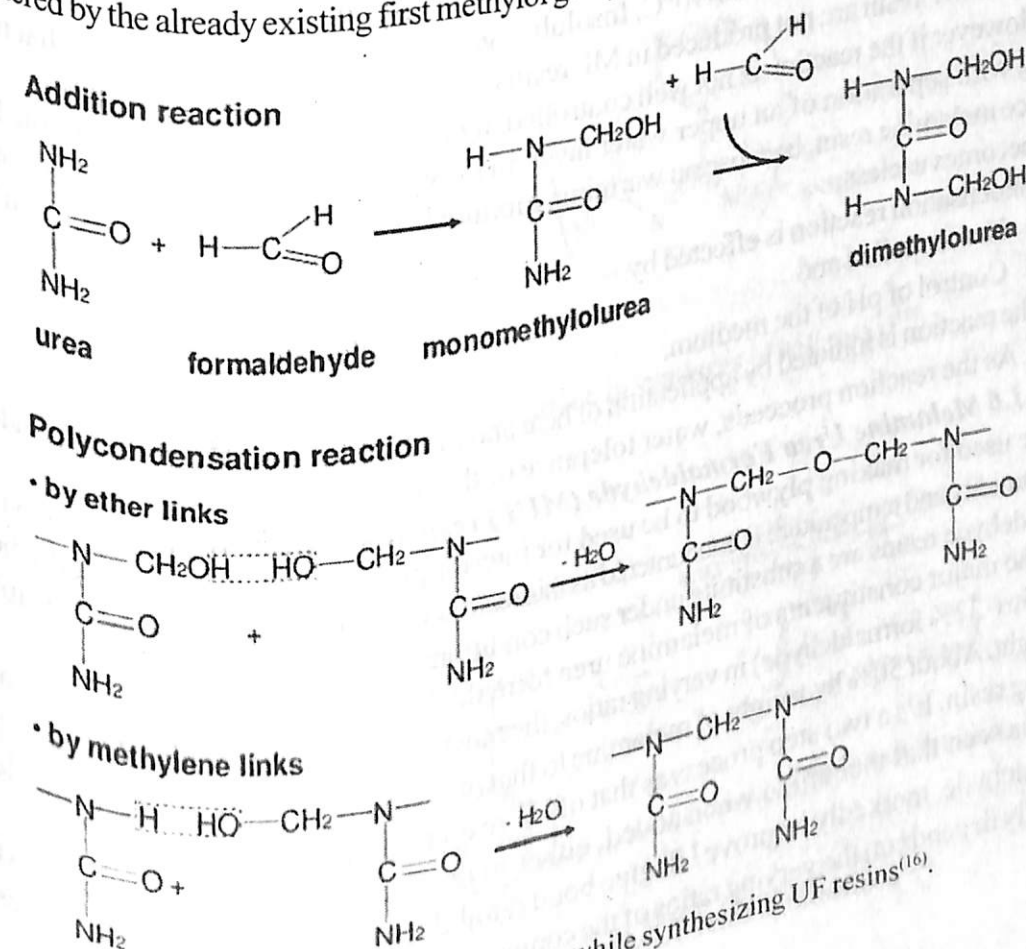


Figure 1: Reactions taking place while synthesizing UF resins⁽¹⁶⁾.

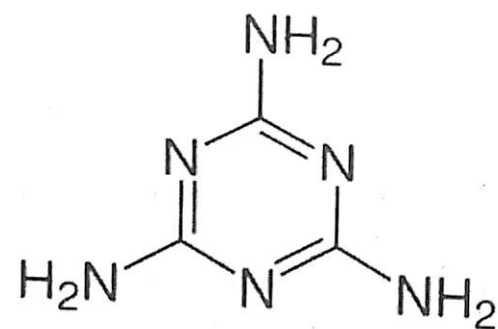


Figure 2: Chemical structure of Melamine ⁽¹⁷⁾.

An important difference between MF and UF is that the MF condensations occur not only under acidic conditions but also under slightly alkaline conditions. With melamine and formaldehyde, the optimum pH for methylation is in the range 8.5-9.0. Within this range, methylation and further condensation to medium and even high molecular weight resin can be safely carried out at about 80-90°C. Insoluble, granular methylol compounds that may occur with UF resin are not produced in MF resins.

However if the reaction is not well controlled, it may lead to viscous, hydrophobic MF resins, with separation of an upper water layer. pH lower than 7.0 should not be used to produce melamine resin, because on warming, mixture becomes lump very rapidly and the resin becomes useless.

The condensation reaction is effected by—

- i) Heat supplied and
- ii) Control of pH of the medium.

The reaction is initiated by application of heat and rate of reaction is primarily dependent on pH. As the reaction proceeds, water tolerance of the resin decreases⁽¹⁾.

2.1.6 Melamine Urea Formaldehyde (MUF) resins: Urea formaldehyde resin can only be used for making plywood to be used for interior purposes. It cannot be used where high humidity and temperature is encountered as this resin is very sensitive to moisture. Melamine formaldehyde resins are a substitute under such conditions, but again melamine is costly.

The major constituents of melamine urea formaldehyde resin are urea, melamine and formalin (-37% formaldehyde) in varying ratios, the most common being 1:1:2 or 3:2:10 parts by weight. About 50% by weight of melamine to that of urea is sufficient to give outstanding bonding resin. It's a two step process as that of UF resin formation.

It is seen that melamine when added, either in monomeric form or co-reacted with formaldehyde, markedly improve UF glue bond durability. The adhesiveness of this resin primarily depends on the varying ratios of the compositions used⁽¹⁾.

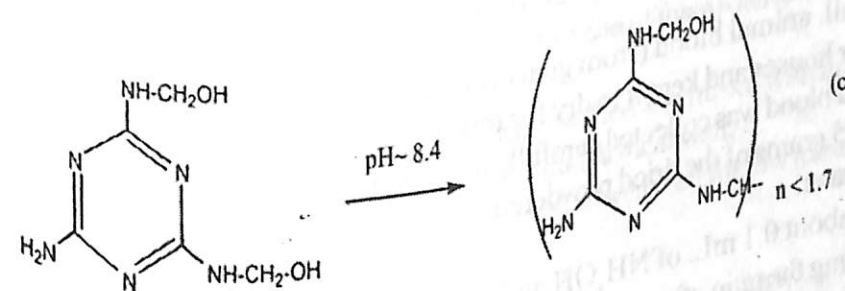
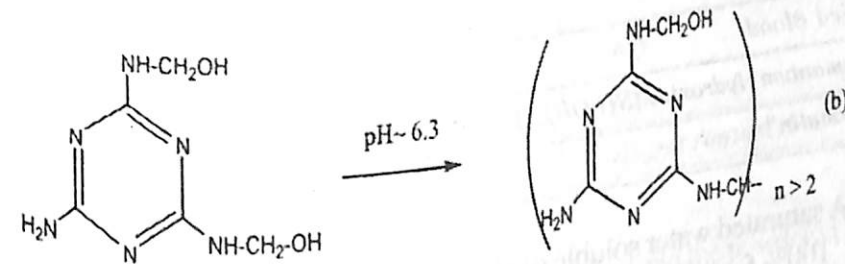
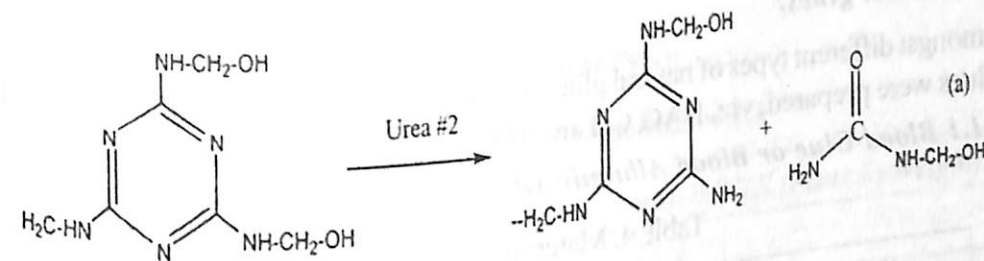


Figure 3: Possible reaction scheme in synthesis of MF resin ⁽¹⁸⁾.

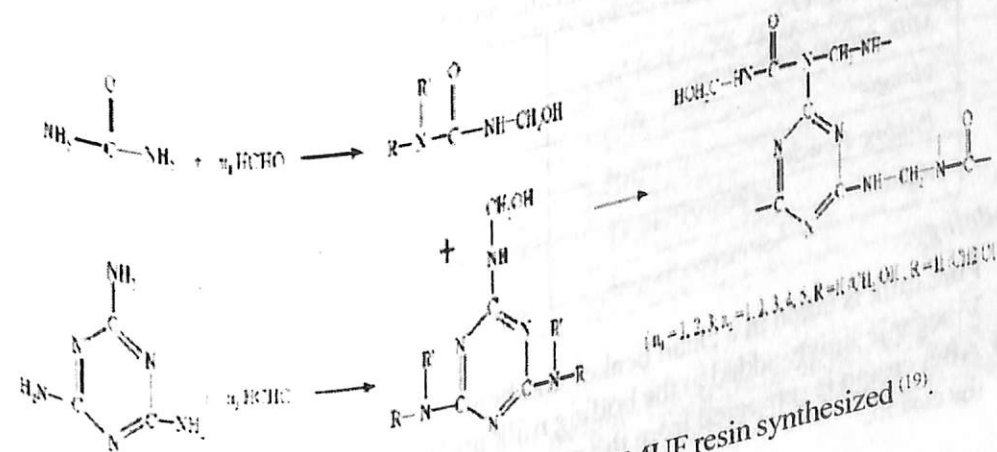


Figure 4: Possible reaction scheme of MUF resin synthesized ⁽¹⁹⁾.

3. METHODOLOGY:

3.1 Natural glues:

Amongst different types of natural glues, the protein-based glues were prepared. Three such glues were prepared, viz. BAG, CG and SG.

3.1.1 Blood Glue or Blood Albumin Glue (BAG):

Table 4: Materials required for BAG.

INGREDIENTS	AMOUNTS (parts by weight)
Dried Blood	50.0
Ammonium Hydroxide (NH_4OH)	3.0
Formalin*	7.5

*Formalin: A saturated water soluble of about 40% formaldehyde by volume or 37% by mass, is called '100% formalin' (13).

Methodology:

- First of all, animal blood (from goat) was collected in a thin glass plate from the slaughter houses and kept it to dry for few days.
- The dried blood was collected carefully in a clean small container.
- About 1.5 grams of the dried powdered blood was taken and dissolved in about 15 mL. of water.
- Finally, about 0.1 mL. of NH_4OH and about 0.2 mL. of formalin was added in it maintaining the ratio of amounts (in parts by weight) as given in the above Table 4. These are then stirred well till a thick paste was obtained.

3.1.2 Milk Glue or Casein Glue (CG):

Table 5: Materials required for CG.

INGREDIENTS	AMOUNTS
Milk (about 90% pure)	100 mL.
Vinegar	25 mL.
Baking Powder	5 grams
Water	As per required

Methodology:

- Pure milk is taken in a clean beaker and heated to boil.
- Vinegar is slowly added to the boiling milk and constantly kept on stirring
- After casein is separated from the milk, water is added slowly for the dispersion of the casein.

- Finally, baking powder is added and is continuously stirred for few minutes until a thick dispersion is attained.

3.1.1 Soyabean Glue or Soya-protein Glue (SG):

Table 6: Materials required for SG

INGREDIENTS	AMOUNTS (Parts by weight)
Soyabean flour	55.5
Defoamer (kerosene oil)	1.7
Calcium hydroxide [$\text{Ca}(\text{OH})_2$]	6.5
50% Sodium hydroxide (NaOH)	8.0
Water	As per required

Methodology:

- Packeted soyabean (100% fat free) is ground and the powdered soya meal is soaked in little amount of water to break any lumps.
- Defoamer (one which prevents the formation of foam)- kerosene oil is added followed by addition of calcium hydroxide.
- Once the flour particles are uniformly wetted and mixed with the above ingredients, further dilution is done to disperse the soya meal well.
- Finally, sodium hydroxide is added and stirred well for few minutes to get uniformly dispersed glue.

3.2 Synthetic resins:

Amongst the various types of synthetic resins, the amino-plastic resins were synthesized as part of this project. Three such resins prepared were UF resins, MF resins and MUF resins.

3.2.1 Urea Formaldehyde (UF) resins:

Table 7: Materials required for UF resin.

INGREDIENTS	AMOUNTS (Parts by weight)
Formalin	250
Urea (NH_2CONH_2)	100
50% Sodium hydroxide (NaOH)	As per required
50% Acetic acid (CH_3COOH)	As per required

Methodology:

- Urea and formalin (ratio by parts weight=1:2.5) are taken in a round bottom flask and mixed well. Here, instead of formaldehyde (HCHO), formalin is taken, as it is

diluted solution of HCHO containing 37% HCHO by mass (to reduce emission of HCHO during reactions).

- II. In the above mixture, NaOH is added drop-wise to raise the pH to about 7.5-8.0 (this pH is maintained throughout the reaction) and stirred well.
- III. This flask containing the mixture is then fitted with a reflux condenser and heated to boil for about 2 hours at 90-95 fC.
- IV. The solution obtained is a clear solution which is then cooled to about 50 fC and acetic acid is added to lower the pH to 5.5-6.0.
- V. This is again kept to boil at an elevated temperature of about 90-95 fC and the formation of resin is checked constantly and also the tolerance is maintained as required.
- VI. Finally, the reaction is arrested by raising the pH to 7.5-8.0 adding NaOH drop-wise and is cooled down gradually.

3.2.2 Melamine Formaldehyde (MF) resins

Table 8: Materials required for MF resin.

INGREDIENTS	AMOUNTS (Parts by weight)
Melamine	100
Formalin	200
50% Sodium hydroxide (NaOH)	As per required

Methodology:

- I. Melamine and formalin are taken in a clean dry round bottom flask in the given ratio 1:2 by parts weight and mixed well.
- II. Sodium hydroxide is added in that mixture drop wise to raise the pH to about 8.0-9.0 and stirred well.
- III. This mixture is then kept over a water bath fitting a reflux condenser to the flask and is heated to boil for about an hour at about 90-95 fC. pH is maintained at 8.0-9.0 throughout the proceeding of the reaction.
- IV. The formation of resin is checked by dropping few drops of the resin into water. Formation of white dispersion indicates the starting of formation of the resin.
- V. Tolerance is checked and maintained at the required level.
- VI. Finally the resin is allowed to cool down.

3.2.1 Melamine Urea Formaldehyde (MUF) resins:

Table 9: Materials required for MUF resins.

INGREDIENTS	AMOUNTS (Parts by weight)
Urea	60
Melamine	40
Formalin	200
30-40% Sodium hydroxide (NaOH)	As per required
50% Acetic acid (CH_3COOH)	As per required

Methodology:

- I. Urea, melamine and formalin are taken in a clean dried round bottom flask (ratio by parts weights= 60:40:200) stirred vigorously.
- II. The pH of this mixture is raised to 8.0-9.0 by adding NaOH drop-wise.
- III. A reflux condenser is fitted at the mouth of the flask and the mixture is set on a water-bath and is heated to boil at about 90-95 fC for about an hour for polymerization reactions to occur. pH is checked time to time.
- IV. This mixture which has formed a clear solution is allowed to cool to about 50-55 fC. At this level, 50% acetic acid solution is added to lower the pH to about 5.5-6.0.
- V. This mixture is again kept for boiling at about 90-95 fC and continuously the pH is maintained. The formation of resin is checked.
- VI. Finally, the tolerance is checked. And as soon as the required tolerance is reached, the reaction is arrested by increasing the pH to 7.5-8.0 adding NaOH solution.

4. Results

The following results were obtained after the synthesis of the adhesives:

4.1 Natural Glues:

Table 10: Product yield and color of Natural glues prepared.

GLUES	PRODUCT YIELD (in mL.)	COLOUR
BAG	~18	Dark-reddish brown
CG	~65	Dirty white
SG	~165	Brownish-yellow

4.2 Synthetic Resins:

Table 11: Product yield and color of Synthetic resins synthesized.

RESINS	PRODUCT YIELD (in mL.)	pH of Resin formed	COLOUR AND NATURE	Storage Life in Liquid Form (in days)	TOLERANCE OF FORMED RESIN
UF	60	8	Beige dense liquid	30	1:3
MF	59	8.5	Dense yellowish- brown liquid	1	1:9
MUF	57	8	Light-brown dense liquid	60	1:6

5. Applications of the Adhesives:

The adhesives (both natural glues and synthetic resins) prepared were tested for their adhesive properties in five different types of substances, viz.— cloth, glass, paper, plastic, wood.

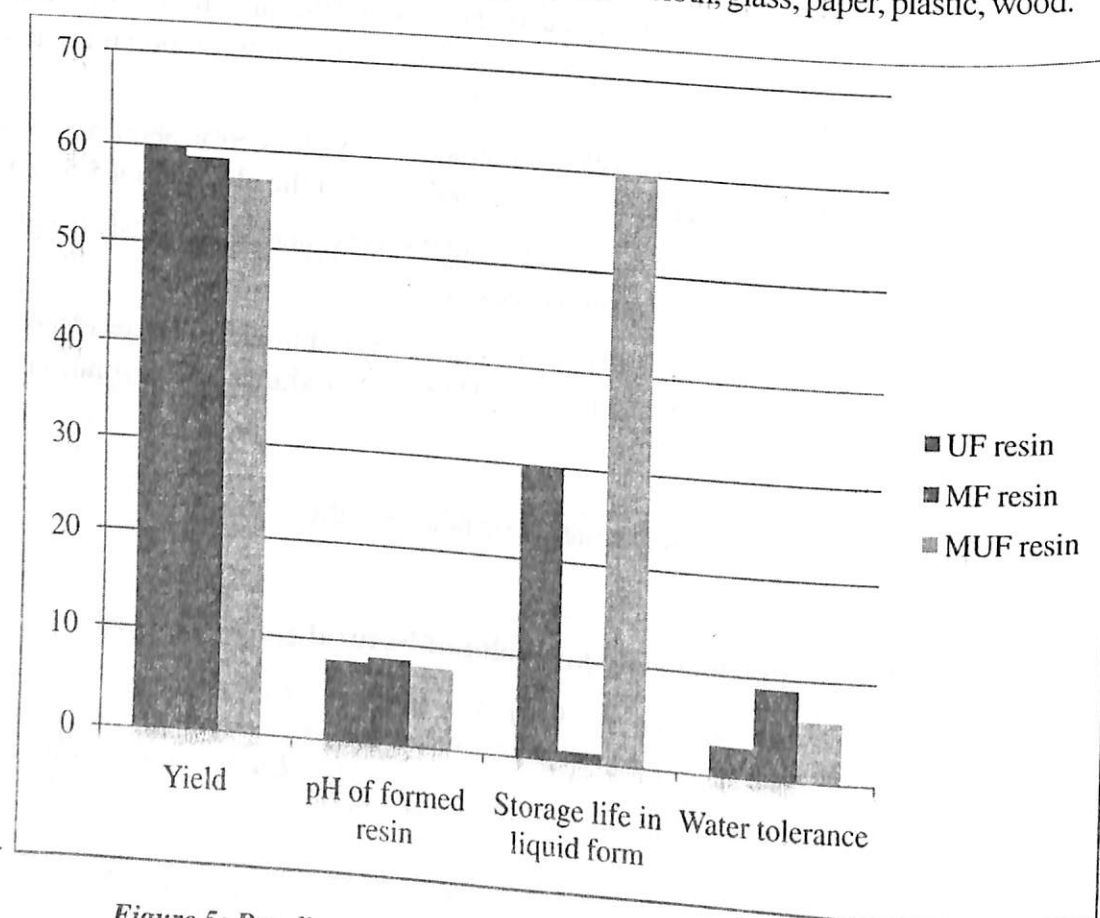


Figure 5: Bar diagram showing different properties of Synthetic Resins.

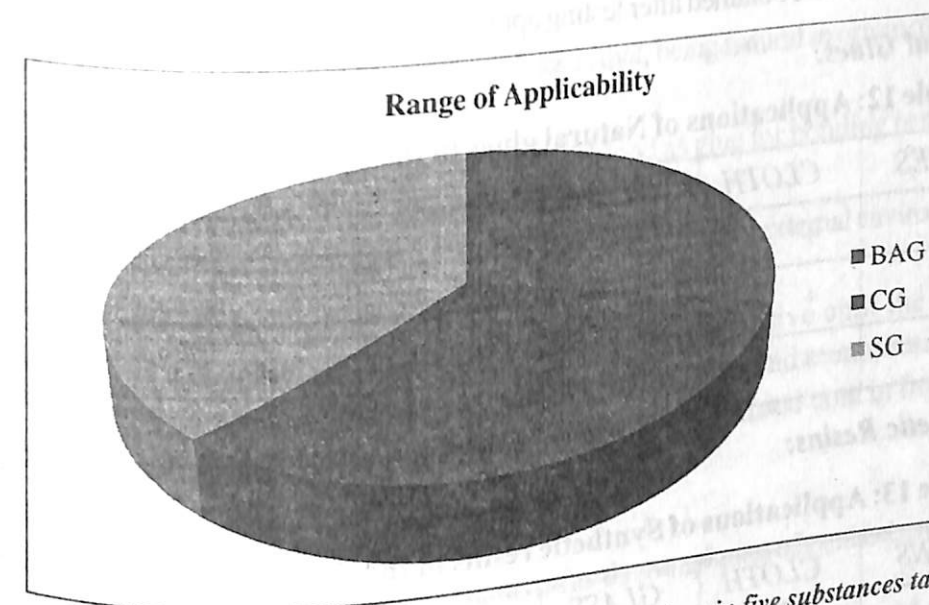


Figure 6: Pie chart showing applicability of natural glues in five substances taken.

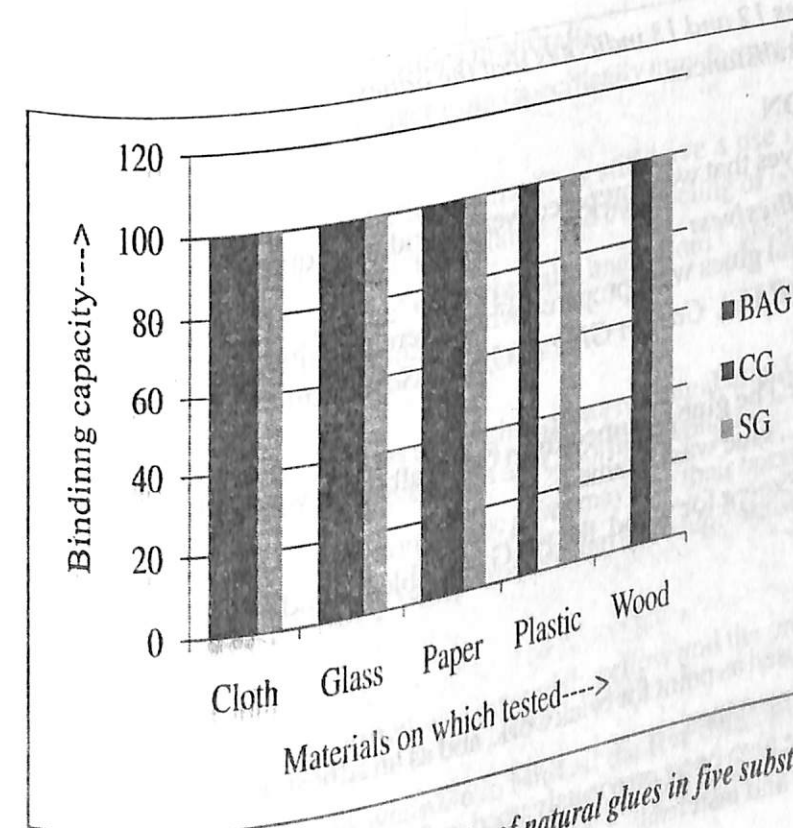


Figure 7: Bar diagram showing binding capacity of natural glues in five substances taken.

Following results were obtained after testing applicability of synthetic resins:

5.1 Natural Glues:

Table 12: Applications of Natural glues in the five selected substances.

GLUES	CLOTH	GLASS	PAPER	PLASTIC	WOOD
BAG	+	+	+	+	-
CG	+	+	+	-	+
SG	+	+	+	+	+

5.2 Synthetic Resins:

Table 13: Applications of Synthetic resins in the five selected substances.

RESINS	CLOTH	GLASS	PAPER	PLASTIC	WOOD
UF	+	+	+	+	+
MF	-	+	-	-	+
MUF	-	+	-	-	-

'+' in Tables 12 and 13 indicates that the adhesives were able to bind the substances with the same substances.

6. DISCUSSION

The adhesives that were prepared were found to be quite applicable.

6.1 Natural Adhesives:

Three natural glues were prepared which were protein-based glues. They were: Blood Albumin Glue (BAG), Casein Glue (CG) and Soyabean Glue (SG).

6.1.1 BAG:

Properties: The glue obtained from blood albumin was dark-reddish brown in colour and about 18mL. glue was obtained.

Applicability: Except for wood, the BAG was able to bind clothes, glass pieces, paper and plastic materials.

Advantages⁽¹⁾:

- They are used as paint for brickwork, and as an adhesive for filling joints between brick and building stones.
- They have also been previously used in food packaging, as the bonds are odourless, non-toxic and tasteless.

Drawbacks⁽¹⁾:

- One of the major drawbacks of blood glues is that, being natural in origin, they cannot be stored for longer period of time.
- As blood clots very quickly, so use of fresh blood as glue for bonding is not an easy task.
- Under high temperatures and long exposure to moisture and external environment, the bonds gradually become weaker.

Remarks: BAG represents the most water resistant adhesive until the advent of phenol formaldehyde resins. These glues yield the most water and mould resistant glue bonds. An important property of such glues is that they can be used both in fresh as well as dried forms ⁽¹⁾.

6.1.2 CG:

Properties: About 65mL. CG was obtained after the completion of synthesizing of casein glue which was dirty white in colour.

Applicability: CG was able to bind cloth pieces, glass, paper and wood materials. And exception was found in case of plastic as the adhesive did not work for plastic materials.

Advantages⁽¹⁾:

- These glues once also found applications in aircrafts.
- It is also used in transformer manufacturing (specifically transformer board) due to its oil permeability.
- While largely replaced by synthetic resins, CGs still have a use in certain niche applications, such as laminating fire-proof doors and the labeling of bottles.
- The popular 'Elmer's School Glue' was originally made from Casein because it was non-toxic and would wash out of clothing.

Drawbacks⁽¹⁾:

- Being natural in origin, these glues are highly prone to fungal attacks which can be prevented by addition of preservatives such as 'Chlorinated phenols'.
- Under harsh environmental conditions, these glues may lose their bonding strengths.

Remarks: These glues are highly water resistant and have high durability ⁽¹⁾.

6.1.3 SG:

Properties: The colour of glue formed was brownish-yellow and the product yield was almost 165mL.

Applicability: The glue obtained was able to bind all the five different materials that were selected for testing the glue, i.e. SG was able to bind cloth materials, glass, paper pieces, plastics and wooden pieces as well.

Advantages⁽¹⁾:

- i. They have great potentials of application for wood composites due to their eco-friendly and proper adhesion properties, fast and simple preparation methods as well as cheaper prices.
- ii. Novel and environment-friendly soy-based adhesives without any formaldehyde emission become perfect choice to replace synthetic resins with high human cancer risk.
- iii. It can also be used to make interior grade plywood.

Drawbacks⁽¹⁾:

- i. Soya meal glues being natural in origin cannot be stored longer due to fungal growth, especially in high humidity conditions for which preservatives such as 'Chlorinated phenol' or 'Copper naphthanate' are widely used.
- ii. The bonds gradually get weaker for long-term exposure to external atmosphere.

Remarks: These glues are good moisture resistant and the bonds formed from them are highly durable⁽¹⁾.

6.2 Synthetic Adhesives:

Three synthetic resins were prepared which were amino-based resins. They were—*Urea Formaldehyde (UF) resins*, *Melamine Formaldehyde (MF) resins* and *Melamine Urea Formaldehyde (MUF) resins*.

6.2.1 UF resins:

Properties: The product yield of UF resins was 60mL with a pH of 8. The formed UF resin was beige dense liquid having a storage life of 30 days in liquid form and with a tolerance of 1:3

Applicability: UF resin was found to show its binding capacity in all the materials that was taken for testing, i.e. cloth, glass pieces, paper pieces, plastic and wooden materials under room temperature and normal pressure conditions.

Advantages⁽¹⁾:

- i. It is used in decorative laminates, textiles, paper foundry, sand moulds, wrinkle resistant fabrics, cotton blends, rayon, etc.
- ii. These adhesives are also used to glue wood together and are mostly used in preparing interior grade ply-woods, doors, etc.
- iii. UF foam was commonly used when producing electrical appliances casing (e.g. desk lamps) due to good insulating properties.
- iv. It is also used in agriculture as a controlled release source of Nitrogen fertilizer.
- v. Foams of UF have been used as artificial snow in movies.

Drawbacks⁽¹⁾:

- i. Emission of formaldehyde is a common problem, which can be lowered by using molar ratio towards the lower end of the range. Emission of formaldehyde causes watery eyes, nose irritations, wheezing and coughing, fatigue, skin rash, severe allergic reactions, burning sensations in the eyes and throat, nausea, and difficulty in breathing in some humans (usually >1.0 ppm).
- ii. The durability of such bonding decreases on exposure to external uses, thus interior grades ply wood are mostly joined with such resins.

Remarks: They are highly durable for binding interior objects⁽¹⁾.

6.2.2 MF resins:

Properties: The obtained MF resin was a dense yellowish-brown liquid with a product yield of 59mL. The pH of the resin was 8.5 with a storage life of 1 day in liquid form and having a tolerance of 1:9.

Applicability: Only glass pieces and wooden materials were able to bind using MF resins under room temperature and normal pressure conditions. It did not show any positive result for binding cloth pieces, paper and plastic materials.

Advantages⁽¹⁾:

- i. The principle use of melamine resin is as the main constituent of high pressure laminates, such as Formica, and of laminate flooring.
- ii. Melamine resin tile-wall panels can also be used as white boards.
- iii. It is also used in plastic laminates.
- iv. It is often used in kitchen utensils and plates (such as Melmac) but such utensils and bowls are not microwave safe.
- v. They are also used in furniture and cabinet making.

Drawbacks⁽¹⁾:

- i. One of the major problem of such resins is that melamine is very costly as compared to urea.
- ii. Bonding at room temperatures is not obtained easily but under high temperature and high pressure, these adhesives give good bonding.
- iii. Their storage time in liquid form is very less, i.e. just 24 hours.
- iv. Although formaldehyde emissions are there but compared to UF resins, formaldehyde emissions are less in MF resins as formaldehyde is tightly bound in MF than in UF, hence reducing their emissions.

Remarks: The bonds obtained from these are highly durable⁽¹⁾.

6.2.3 MUF resins:

Properties: MUF resin obtained was a light brown dense liquid having a pH of 8. The product yield was 57mL. with a storage life of 60 days. The tolerance of the resin was 1:6.

Applicability: Only glass pieces were able to bind using MUF resins under normal pressure and room temperature. MUF resins were not found useful to bind the rest, i.e. it showed negative results for cloth, paper, plastic and wooden pieces.

Advantages⁽¹⁾:

- Although they are similar to UF resins, yet they are higher moisture resistance than UF resins and they also show good performance as compared to MF resins.
- These resins are mainly used to produce panels with higher performance, such as for kitchen, floor or some structural composites.

Drawbacks⁽¹⁾:

- One of the major drawbacks of MUF resins is that they require very high temperatures and pressures for good bonding or it may loosen at no time.
- These resins also liberate formaldehyde but to a very lesser amount.

Remarks: They have high durability and can be used for external applications unlike UF and MF resins ⁽¹⁾.

Following are some figures showing consumption of amino resins:

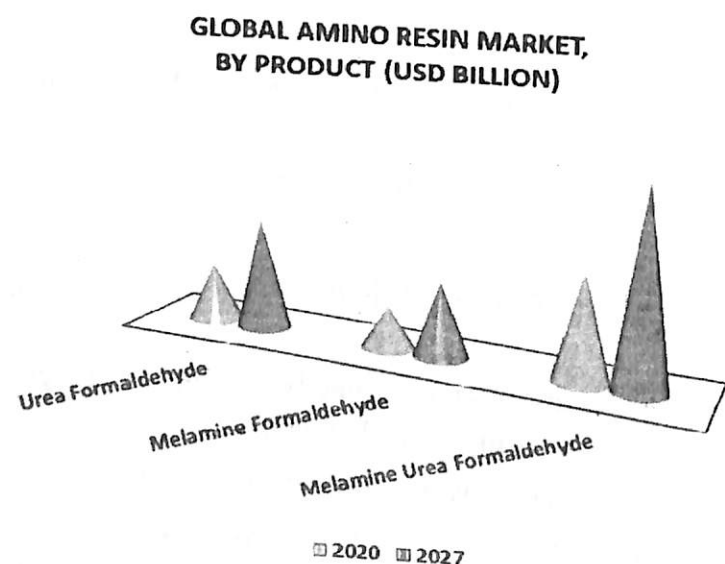


Figure 6: Global amino resin market showing variation from 2020 to 2027 (estimated) ⁽¹⁴⁾.

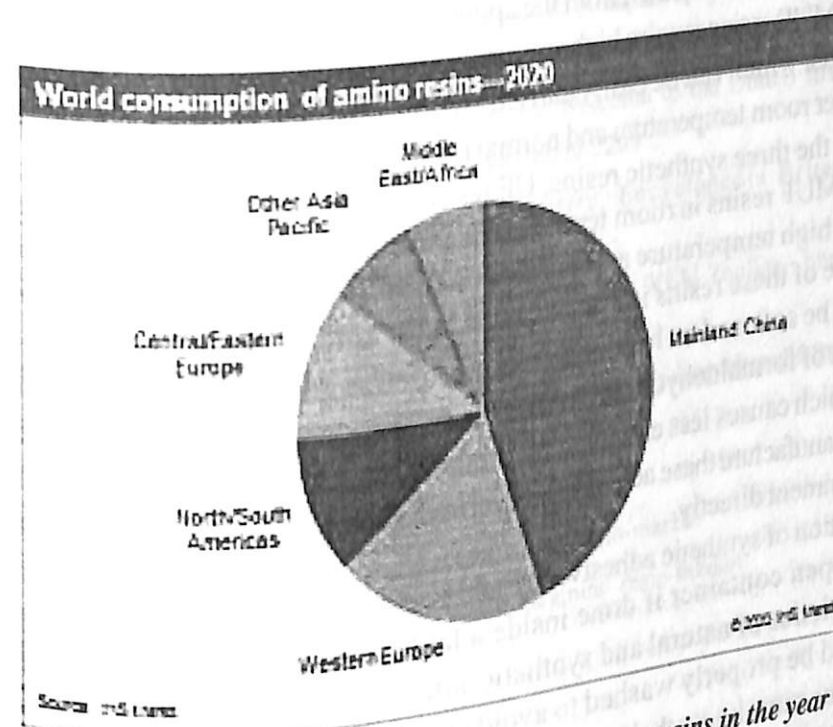


Figure 7: Pie chart showing World consumption of amino-resins in the year 2020⁽¹⁵⁾.

From the above discussion, a brief idea about different types of adhesives can be obtained which are commonly used for industrial as well as those that were used during the ancient and medieval period. Adhesives are frequently used by all of us in our day-to-day life but we rarely get to know about the chemistry that lies underneath. A brief knowledge about the adhesive chemistry was obtained. Although industrially the synthetic resins are prepared in large resin kettles situated in resin plants, it was an initiative to manufacture such resins in the laboratory. Adhesives are of utmost importance in our daily life in binding various substances although their composition, chemistry, etc. are of rare importance to think of.

From the results of the study, it can be analyzed and discussed that the natural glues—BAG, CG and SG have wide range of applications in our everyday used materials such as cloth, paper, wood, etc. amongst the three natural glues synthesized, SG showed a wide range of applications. But a common problem about these glues is about their storage as these glues, being of natural origin is prone to fungal attack. However CGs and SGs can be stored for a longer period of time by using preservatives such as 'chlorinated phenols' or 'copper naphthanate'.

On the other hand, the synthetic resins—UF, MF and MUF resins are widely used in industries such as in ply wood industries as glues, thermosetting materials manufacturing industries such as utensils, in cabinet manufacturing, in furniture manufacturing industries, etc.

Also, a conclusion can be drawn from the applications carried out in everyday used materials that MF and MUF resins require high temperatures and pressures for binding wooden frames and ply boards, which can be easily carried out in industries, whereas UF resins give good bonding under room temperature and normal pressure conditions.

Among the three synthetic resins, UF resins showed versatile use than the other two, i.e. MF and MUF resins in room temperature and under atmospheric pressure conditions, while under high temperature and pressure, MF and MUF resins give good result. But a disadvantage of these resins is that being thermosetting in nature; once they form lumps these cannot be softened on heating to be again used as glues. Overall, UF resins also emit high amount of formaldehyde, as a replacement of which melamine adhesives came into existence which causes less emission of formaldehyde. Along with that, large resin kettles are used to manufacture these adhesives, which reduce the emission of harmful formaldehyde to the environment directly.

Preparation of synthetic adhesives should be done cautiously only in reflux condenser, not in any open container if done inside a laboratory to prevent any kinds of health hazards. Varieties of natural and synthetic adhesives are available. After their use, the hands should be properly washed to avoid any skin problems. And before using these adhesives, one must know their proper applications and also handle these adhesives with utmost care. □

References:

1. Nath, S. K. (2009). *Plywood Manufacturing Practices in India* (pp.153-182). DVS Publication Pvt. Ltd. New Delhi.
2. Dinte, E., & Sylvester, B. (2018). Adhesives: Applications and Recent Advances. In *Applied Adhesive Bonding in Science and Technology* (1st ed., pp. 119-132). InTech. <http://dx.doi.org/10.5772/intechopen.71854>
3. Tikkanen, A. (2023, May 6). *Resin: Chemical Compound*. Encyclopedia Britannica. <https://www.britannica.com/science/resin>
4. Tikkanen, A. (2023, May 28). *Rosin: Chemistry*. Encyclopedia Britannica. <https://www.britannica.com/science/rosin>
5. <https://vikaspedia.in/agriculture/post-harvest-technologies/natural-resins-and-gums-of-commercial-importance/introduction>
6. Tosif, M. M., Najda, A., Bains, A., Kaushik, R., Dhull, S. B., Chawla, P., & Walasek-Janusz, M. (2021). A comprehensive review on plant-derived mucilage: Characterization, functional properties, applications, and its utilization for Nano-carrier fabrication. *Polymers*, 13(7), 1066. <https://doi.org/10.3390/polym13071066>
7. Augustyn, A. *Glue: Adhesive*. Encyclopedia Britannica. <https://www.britannica.com/science/glue>

8. Mukherjee, S. M., Singh, S. P., Kapoor, R. P., & Dass, R. (Chapter-8, Synthetic Polymers, pp. 337-374). *Organic Chemistry, Part-III* (2nd ed.). New Age International Publishers.
9. *Thermoplastic and thermosetting polymers*. (n. d.). Welcome to the Dutton Institute | Dutton Institute. <https://www.e-education.psu.edu/matse81/node/2209>
10. Gregersen, E. (2023, May 15). *Polymer: Chemistry*. Encyclopedia Britannica. <https://www.britannica.com/science/polymer>
11. Max.zm. (2022, November 29). *Homopolymer vs. copolymer*. VEM Tooling. <https://www.vem-tooling.com/homopolymer-vs-copolymer/>
12. Rogers, K. (n.d.). *Polymerization | Definition, classes, & examples*. Encyclopedia Britannica. <https://www.britannica.com/science/polymerization>
13. <https://en.m.wikipedia.org/wiki/Formaldehyde>
14. <https://www.verifiedmarketresearch.com/product/amino-resin-market/>
15. <https://www.mordorintelligence.com/industry-reports/amino-resin-market>

A Review of *Six Sigma* as a Quality Improvement Program

Porinita Dutta

Department of Statistics
Tinsukia College, Tinsukia, Assam
Email: dutta.porinita1@gmail.com

Abstract

Six Sigma is a smarter tool or methodology to manage business or departments. *Six Sigma* is used to improve business processes by reducing defects and errors, minimizing variation, and increasing quality and efficiency. The goal of *Six Sigma* is to achieve a level of quality that is nearly perfect, with only 3.4 defects per million opportunities. Defects free product can be in any organization by implementing *Six Sigma*. Ultimate objectives of the methodology to solve problems, improve the quality, profitability and customers satisfaction. *Six Sigma* is the tool through which we can improve the Quality and profitability by removing the cause of defects and variability in manufacturing and business processes. *Six Sigma* is perhaps the most widely-accepted initiative by all a broad range of organizations or business sectors. In this article, we reviewed the process and some aspects of *Six Sigma* such as objectives, methodologies, strategies, tools, techniques, principles, successful implementation, application etc.

1. Introduction

Six Sigma is a business methodology for quality improvement that measures how many defects there are in a current process and seeks to systematically eliminate them. The main objective of any business sector or organization is to make profit. For increasing the profit, the selling price should increase and/or the manufacturing cost should come down. Since the price is decided by the competition in the market, hence the only way to increase the profit is to cut down the manufacturing cost which can be achieved only through continuous

P. Dutta A review of six sigma as a quality improvement program Horizon (Ed. Book), 226-250(2023), copyright: Tinsukia College Publishing. ISBN: 978-81-19322-07-7

improvement in the industry's operation. *Six sigma* provide an overall framework for continuous improvement in the process of any organization or business sector. *Six Sigma* was developed by Motorola in 1986. Sir Bill Smith, "Father of *Six Sigma*" introduced this quality improvement methodology to Motorola. The invention was motivated by the high cost of poor quality discovered at Motorola. It has since been adopted by many other companies around the world, including General Electric, Toyota, and Amazon. It is used in industries such as manufacturing, healthcare, finance, and service industries to improve customer satisfaction, reduce costs, and increase profits. Now *Six Sigma* is an enormous 'brand' in the world of corporate development. Statistically, *Six Sigma* refers to a process in which the range between the mean of a process quality measurement and the nearest specification limit is at least six times the standard deviation of the process. [1]

The traditional quality management approaches, including Statistical Quality Control (SQC), Zero Defects and Total Quality Management (TQM), have been key players for many years, while *Six Sigma* is one of the more recent quality improvement initiatives to gain popularity and acceptance in many industries across the globe. *Six Sigma* differs from other quality programs in its top-down drive in its rigorous methodology that demands detailed analysis, fact-based decisions, and a control plan to ensure on going quality control of a process. Many companies including GE, Honeywell, Sony, Caterpillar, and Johnson Controls have adopted *Six Sigma* and obtained substantial benefits. [13]

Six Sigma is a long-term commitment. It won't work well without full commitment from upper management. *Six Sigma* changes the way a company thinks by teaching fact-based decision making to all levels. [2]

Six Sigma is a rigorous, focused, and highly effective implementation of proven quality principles and techniques. Incorporating elements from the work of many quality pioneers, *Six Sigma* aims for virtually error-free business performance. *Six Sigma* proponents claim its business strategy benefits include up to 50% process cost reduction, cycle-time improvement, less waste of materials, a better understanding of customer requirements, increased customer satisfaction and value stream, and more reliable products and services. Motorola holds the federal trademark for *Six Sigma*, and it is generally acknowledged that *Six Sigma* can be costly to implement and can take several years before a company begins to see bottom-line results. In 1995, General Electric CEO Jack Welch's very public endorsements of *Six Sigma* helped businesses outside of manufacturing understand how *Six Sigma* methodologies can be used to improve customer satisfaction in any industry.

2. History

Since the 1920's the word "Sigma"(s) has been used by mathematician & engineers as a symbol for a unit of Measurement in product quality variation. In the mid 1980's engineers in Motorola in the USA used "six sigma" (s) an informal name for an in-home initiative for reducing defects in production processes, because it represented a suitably high level of quality.

In the late 1980's *Motorola* extended the six sigma methods to its critical business processes, and significantly six sigma became an improvement methodology, i.e; beyond purely 'defect reduction'. In 1990, *Motorola* certified its first "Black Belt" six sigma experts, which indicates the beginning of the formalization of the accredited training of six sigma methods.

In 1995, six sigma became well known after *Mr. Jack Welch* made it a central focus of his business strategy at *General Electric*, and today it is used in different sectors of Industry. (General Electric is an American Multinational conglomerate corporation incorporated in New York) By the year 2000, six sigma was effectively established as an Industry in its own right, involving the training, consultancy & implementation of six sigma methodology.

3. What is Six Sigma

According to American Society for quality, Six Sigma is a method that provides organizations tools to improve the capability of their business processes. This increase in performance and decrease in process variation helps lead to defect reduction and improvement in profits, employee morale, and quality of products or services.

Six sigma is a customer oriented, structured, systematic, proactive and quantitative companywide approach for continuous improvement of manufacturing, services, engineering, suppliers and other business process. It is a statistical measure of the performance of a process or a product. It measures the degree to which the process deviates from the goals and then takes efforts to improve the process to achieve total customer satisfaction [13]. Six Sigma seeks to improve the quality of process outputs by identifying and removing the cause of defects. Six sigma approaches is a collection of managerial and statistical concept and techniques that focuses on reducing variation in process and preventing deficiencies in product. The concept of variation state "No two items will be perfectly identical." In a process that has achieved six sigma capabilities; the variation is small compared to the range of specification limit. A six-sigma process is one in which 99.999966% of the products manufactured are statistically expected to be free of defects (3.4 defects per million). Six sigma is a very clever way of branding and packaging many aspects of Total quality Management (TQM). TQM is a management approaches to long-term successes through customer satisfaction. Manufacturing methods of six sigma are used in Batch production, Job production & Mass production.

The very first definition of Six Sigma is that it is a defect rate metric, specifically; it means 3.4 Defects per Million Opportunities (DPMO).

The term "Six Sigma" is derived from the bell curve in statistics, in which sigma represents the standard deviation from the centre. Six Sigma is simply six standard deviations. What it truly means is that a process is highly capable that customer specifications are actually six standard deviations away from the process centre (see Fig. 2). Hence, a process with six sigma will achieve an extremely low defect rate. The failure of a business process or product is regarded as a defect. When a process produces less than 3.4 defects for one million chances, it is considered efficient.

4. Six-Sigma Metrics – Measuring Defect Rate

- i) Defects per unit (DPU) =
(Number of defects discovered / number of units produced)
- ii) Defects per million opportunities (DPMO) =
(Number of defects discovered / opportunities for error) $\times 1,000,000$

Since a product will only be considered defective if it is produced outside of customer specifications, a process with such a high capability will almost produce no defect. The mathematical calculation of 3.4 DPMO is based on two assumptions: the process output follows a normal distribution, and the process mean may shift up to 1.5 standard deviations in the long term [13]. Six Sigma equates 3.4 defects for every million parts made or process transactions carried out. This quality equates to 99.99966% defect free products or transactions. High quality standards do make sense but the cost required to pursue such high standards have to be balanced with benefits gained. The six sigma processes expose the root causes and then focuses on the improvements to achieve the highest level of quality at acceptable cost. This is essential to achieve and maintain a competitive advantage and high levels of customer satisfaction and loyalty.

Three key characteristics separate six sigma from quality programs of the past:

- i) Six Sigma is a customer focused.
- ii) Six sigma projects produce major returns on investments.
- iii) Six sigma changes how management operates.

5. Objectives of Six Sigma

The main goal or objective of any six sigma implementations is quality improvement. Six Sigma is used to identify and reduce errors and increase the efficiency of business processes. The primary objective of Six Sigma is customer satisfaction, and to achieve the objective, various methods are followed to improve the performance of a product or business process.

6. Principles of Six Sigma

There are five main principles of Six Sigma:

i. Customer focus

This is based on the popular belief that the "customer is the king." The main objective is to maximize the benefits for customers. Hence, a business must understand the needs of their customers, and what drives sales. It requires establishing quality standards according to the market or customer demands.

ii. Assess the value chain and find the problem

Outline the steps of a process to find out unwanted areas and gather related data. Define goals for data collection, purposes for data gathering, and expected insights. Verify that the data is assisting in achieving the objectives, whether more information is needed to be collected, or if data cleansing is required. Find out the problem and its root cause.

iii. Eliminate defects and outliers

After the identification of the problem, make appropriate modifications in the process to eliminate defects. Eliminate any activity in the given process that does not contribute to the customer value. If the value chain is unable to reveal the problem area, various tools are used to find out the problem areas and outliers. Eliminating the outliers and defects removes the bottlenecks in a given process.

iv. Involve stakeholders

A structured process should be adopted where all stakeholders collaborate and contribute to finding solutions to complex issues. The team needs to achieve proficiency in the methodologies and principles applied. Hence, specialized knowledge and training are required to lower project failure risks and ensure optimal performance of the processes.

v. Flexible and responsive system

Whenever an inefficient or faulty process is eliminated, the employee approach and work practices need to be changed. A flexible and responsive environment to the changes in processes can lead to the efficient implementation of the projects. The departments involved should be capable of adapting easily to the change. Companies that periodically examine the data and make appropriate changes to their processes may achieve a competitive advantage.

One of the main benefits of Six Sigma is that it provides a structured and systematic approach to process improvement. By following the five phases of *Define, Measure, Analyse, Improve, and Control (DMAIC)*, Six Sigma practitioners can identify the root causes of problems, implement solutions, and monitor the results. Six Sigma also uses statistical tools and techniques to quantify the performance and capability of processes, and to set clear and measurable goals and targets. This helps to ensure that the improvement efforts are aligned with the customer needs and expectations, and that the outcomes are validated and verified.

7. Challenges of Six Sigma

Six Sigma is not a one-size-fits-all solution to all problems or processes. It comes with its own set of challenges and limitations that must be addressed. To begin, it requires a considerable investment of time, resources, and training to implement and sustain. Six Sigma projects can be complex and lengthy, requiring the support and commitment of the top management, the project team, and the process owners. Additionally, skilled and certified professionals are needed in order to apply the tools and techniques correctly. Furthermore, Six Sigma may not be suitable or applicable for every process or situation; it is designed for stable, predictable, and repeatable processes but may not work well for processes that involve human factors or emotions. In addition, there may be resistance from stakeholders or employees who perceive Six Sigma as a threat to their autonomy or authority. Lastly, some may question the validity of data or statistics used in Six Sigma or the feasibility of proposed solutions.

Despite the challenges, Six Sigma can still be a powerful tool for process improvement if used correctly and strategically. To maximize effectiveness, it is important to ensure that the projects are aligned with the organizational vision and address key issues affecting business performance and customer satisfaction. Involving stakeholders and employees in the projects,

communicating benefits, and fostering a culture of collaboration and improvement are also essential. Additionally, it is important to customize the Six Sigma tools to suit the specific needs of the process, use a combination of quantitative and qualitative methods to collect data, and incorporate creativity and innovation into improvement solutions.

8. Methodology of Six Sigma

Six Sigma has been defined as the statistical unit of measurement, a sigma that measures the capability of the process to achieve a defect free performance. Six Sigma has the ability to produce products and services with only 3.4 defects per million, which is a world class performance. Six Sigma has also been described as a high-performance data driven approach in analysing the root causes of business problems and solving them [2].

There are two major methodologies used within Six Sigma, both of which are composed of five sections, according to the 2005 book "*JURAN Institute Six Sigma Breakthrough and*

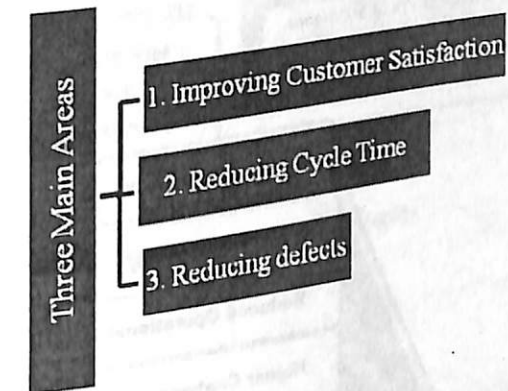


Fig.1. Six Sigma efforts target three main areas

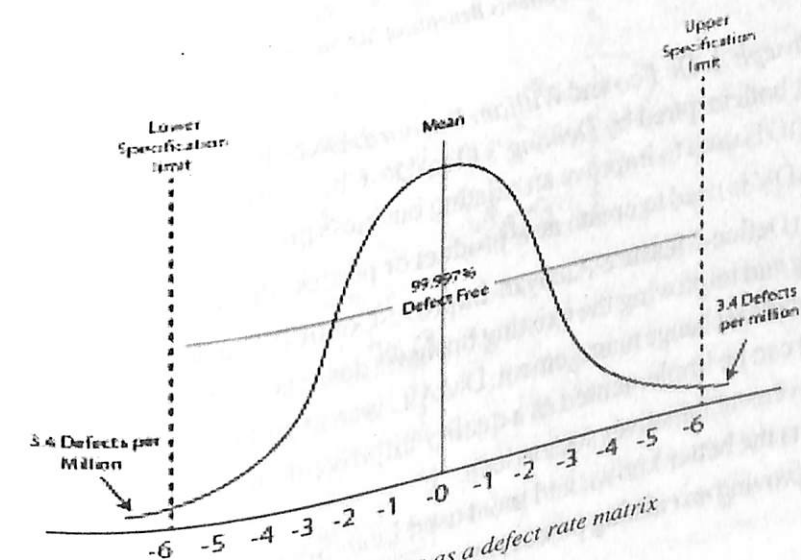


Fig 2: Six Sigma as a defect rate matrix



Fig 3: Objectives of Six Sigma



Fig 4: Key Principles of Six Sigma



Fig 5: Important Benefits of Six Sigma

Beyond" by Joseph A. De Feo and William Barnard. DMAIC Methodology and DMADV Methodology, both inspired by Deming's Plan-Do-Check-Act Cycle.

- i) DMAIC is used to improve an existing business process.
- ii) DMADV is used to create new product or process design.

DMAIC (Define-Measure-Analyze-Improve-Control) is a data-driven approach used for optimizing and improving the existing business designs and processes. It is an effective method of controlled change management. DMAIC is an integral part of a Six Sigma initiative, but in general can be implemented as a quality improvement procedure or as part of other process improvement initiatives such as lean.

DMAIC is the better known and most used Lean Six sigma project method. DMAIC focuses on improving an existing process by incorporating the following five phases.

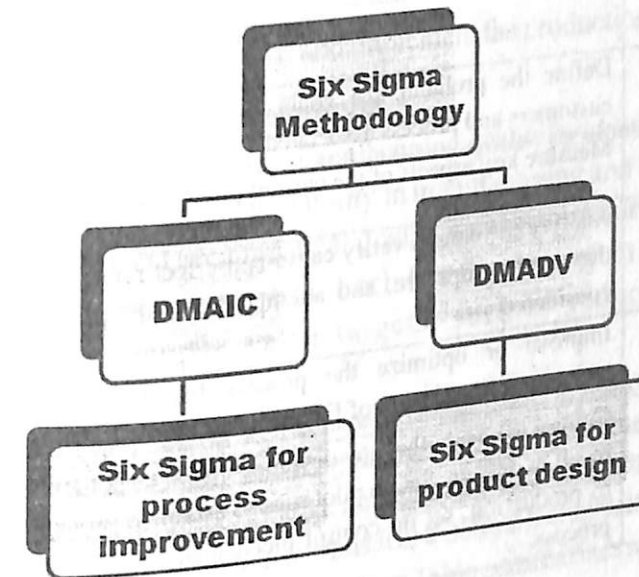
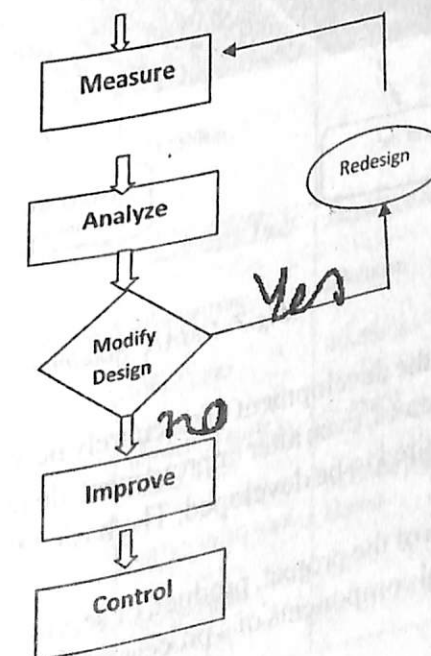


Fig 6: Six Sigma methodologies



DMAIC is the better known and most used Lean Six sigma project method. DMAIC focuses on improving an existing process by incorporating the following five phases.

PHASE	DESCRIPTION
<i>DEFINE</i>	Define the problem, the goals of the project, output to be improved, customers and process associated with the problem.
<i>MEASURE</i>	Measure key aspects of the current process and collect relevant data
<i>ANALYZE</i>	Analyze the data to verify cause-and-effect relationships. Determine what the relationships are, and attempt to ensure that all factors have been considered.
<i>IMPROVE</i>	Improve or optimize the process based upon data analysis using techniques like Design of Experiments
<i>CONTROL</i>	Control to ensure that any deviations from target are corrected before they result in defects. Set up pilot runs to establish process capability, move on to production, set up control mechanisms and continuously monitor the process.

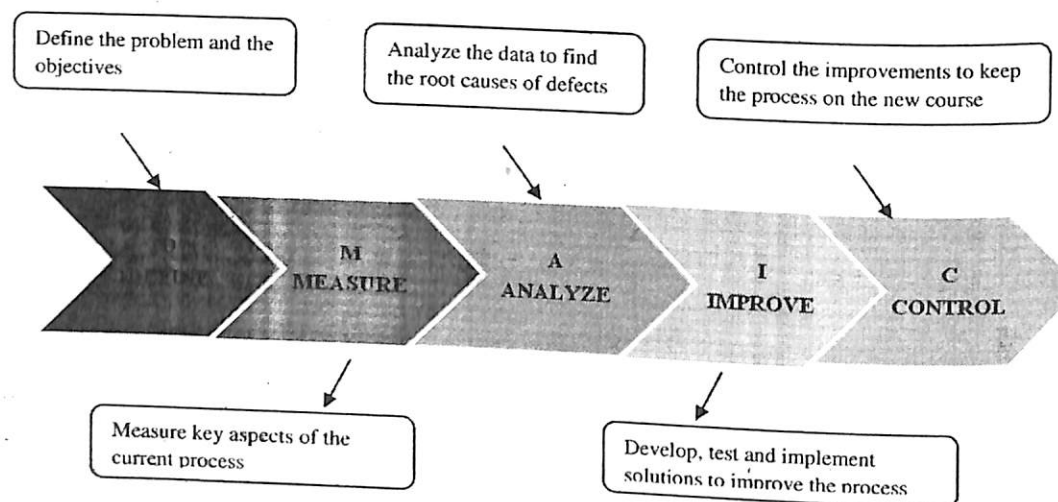


Fig.7: DMAIC Roadmap

DMADV focuses on the development of an entirely new process, product, or service. It is used when existing processes, even after improvement, do not satisfy the customer's needs, and new methods are required to be developed. The basic DMADV methodology consists of the following five steps:

1. Define the purpose of the project, product, or service.
2. Measure the crucial components of a process and product capabilities.

3. Analyze data and develop design alternatives, ultimately selecting the best design.
4. Design the selected best alternative and test details of the process. This phase may require simulations.
5. Verify the design, set up pilot runs, and implement the production process. [5]

9. Six Sigma strategies, tools and techniques

Six Sigma strategies seek to improve manufacturing quality by identifying and removing the causes of defects and minimizing variability in manufacturing and business processes. This is done by using empirical and statistical quality management methods and by hiring people who serve as Six Sigma experts. Each Six Sigma project follows a defined methodology and has specific value targets, such as reducing pollution or increasing customer satisfaction.

Six sigma is a systematic, data-driven approach using define, measure, analysis, improve, and control (DMAIC) process and utilizing design for six sigma method (DFSS). The fundamental principle of six sigma is to 'take an organization to an improved level of sigma capability through the rigorous application of statistical tools and techniques. It generally applies to problems common to production. The following Table summarizes six sigma strategies, tools, techniques, and principles. [6]

TABLE I: KEY TOOLS OF DMAIC METHODOLOGY

Objective:	Objective:	Objective:	Objective:	Objective:
DEFINE the Opportunity.	Measure Current Performance	Analyze the root causes of problems	Improve the process to eliminate root causes.	Control the process to sustain the gains
Key Define Tools	Key Measure Tools	Key Analyze Tools	Key Improve Tools	Key Control tools
i) Cost of Poor Quality(COPQ)	i)Critical to Quality Requirements (CTQs)	i)Histograms	i)Solution	i)Control Charts
ii)Voice of the Stakeholder (VOS)	ii)Sample Plan	ii)Box plots	ii)Selection	ii)Contingency and/or Action Plan(s)
iii)Project Charter	iii)Capability Analysis	iii)Multi-Variate Charts, etc.	iii)Matrix	
iv)As-Is Process Map(s)	iv)Failure modes and Effect analysis (FMEA)	iv)Hypothesis tests.	iv)To-Be process Map(s)	
v)Primary metric (Y)		v)Regression analysis.		

TABLE II: KEY TOOLS OF DMADV METHODOLOGY

Objective:	Objective:	Objective:	Objective:	Objective:
Define the Project goals	Measure the customer needs & specifications	Analyze the process options to meet the customer needs	Design the process to meet the customer needs	Verify the design performance
Key define Tools	Key Measure Tools	Key Analyze Tools	Key Design Tools	Key Verify Tools
i)Project Management Tools.	i)Brain storming	i)FMEA	i)Work design	i)Reliability test
ii)Project Selection.	ii)FMEA screening	ii)Reliability	ii)Machine design	ii)FMEA
iii)Bench marking	iii)Bench marking transfer	iii)Pareto Analysis	iii)Engineering design	iii)Simulation
iv)Survey	iv)Process Map	iv)Risk Analysis	iv)Specification design	iv)Control plan
		v)Gap Analysis		

10. Six Sigma strategies, tools and techniques

Six Sigma strategies seek to improve manufacturing quality by identifying and removing the causes of defects and minimizing variability in manufacturing and business processes. This is done by using empirical and statistical quality management methods and by hiring people who serve as Six Sigma experts. Each Six Sigma project follows a defined methodology and has specific value targets, such as reducing pollution or increasing customer satisfaction.

Six sigma is a systematic, data-driven approach using define, measure, analysis, improve, and control (DMAIC) process and utilizing design for six sigma method (DFSS). The fundamental principle of six sigma is to 'take an organization to an improved level of sigma capability through the rigorous application of statistical tools and techniques. It generally applies to problems common to production. The following Table summarizes six sigma strategies, tools, techniques, and principles. [6]

TABLE III: STRATEGIES, TOOLS, TECHNIQUES & PRINCIPLES

Strategies and Principles	Tools and Techniques
i) Project management	i)Statistical Process control
ii)Data-based decision making	ii)Process capability analysis
iii)Knowledge discovery	iii)Measurement system analysis
iv)Process control planning	iv)Design of experiments
v)Variability reduction	v)Quality function deployment
vi)Beit System	vi)Failure mode and effects analysis
vi)DMAIC process	vii)Regression analysis
vii)Change management tools	viii)Analysis of means and variances
	Hypothesis testing
	Root cause analysis
	Process mapping

11. Design for Six Sigma(DFSS)

Design for Six Sigma (DFSS) is a systematic methodology of improvement that helps businesses create new products or services at a high level of quality. The technique aims to meet the needs of customers and utilize the company's capability as much as possible during the original development of a process.

According to Lin Man-shan, Cai Chun-xing, Li Jin-hong, DFSS is a systematic methodology utilizing tools, training and measurements to enable the organization to design products and processes that meet customer expectations and can be produced at Six Sigma quality levels. The goal of DFSS is to achieve minimum defect rates, six sigma level, and maximum possible impact during the development stage of the products [7].

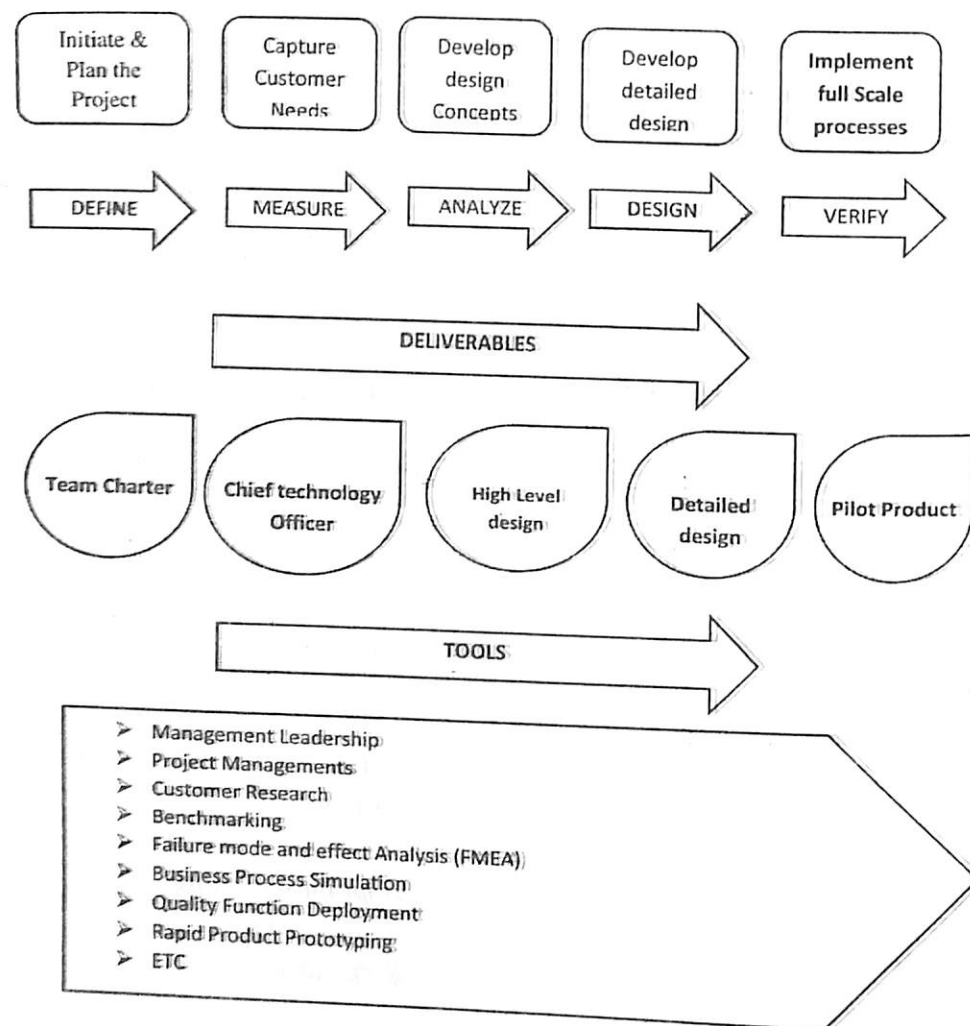


Fig.8: Five Steps of DFSS process

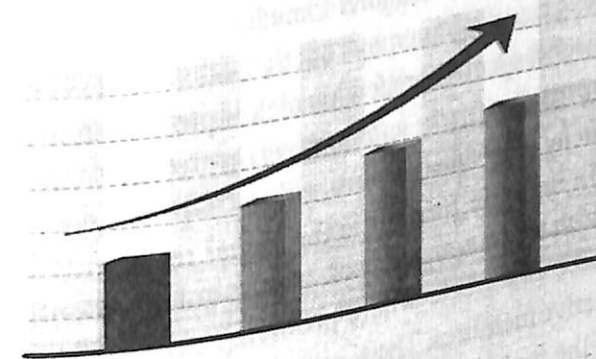
12. Six Sigma tools

Six Sigma tools are defined as the problem-solving tools used to support Six Sigma and other process improvement efforts. The Six Sigma expert uses qualitative and quantitative techniques to drive process improvement. Although the tools themselves are not unique, the way they are applied and integrated as part of a system is. Some of the statistical and graphical tools commonly used in improvement projects are:

i) Charts and Graphs: Usually, the first and best way to analysis measures of a process is to create a picture of the data. Charts and graphs are really nothing more than visual pictures

of data. For most of us, looking at a pie chart or a line graph is a lot more meaningful and convenient than reading tables of numbers. Charts and graphs are of various types, each offering a bit different picture of the data. A Black Belt will usually use at least a couple of these on a project.

ii) Pareto Chart: Pareto Chart is a graphic representation of the *Pareto Principle*: 20 percent of input produces 80 percent of output in any given situation. The Pareto Principle is also called the 80-20 rule. The chart combines a vertical bar graph and a line graph. The bar graph represents the metrics of various business process components, ordered from the largest to the smallest one. The line graph represents the cumulative total of these metrics. *Pareto Chart* is a tool that visualizes what part of the process influences output the most. To create such chart, one can first figure out the components of the process and how to measure them and then put this data into a *Pareto Chart*. This will help us to see how big of an influence on the outcome each component has. In addition to that, it will give a clear idea of what requires immediate attention.



i) Control Chart: A control chart is simply a run chart to which two horizontal lines called control limits are added, the upper control limit and the lower control limit. Control limits are chosen statistically so that they are highly probability (greater than 0.99) that points will fall between these limits if the process is in control. Control charts only give the signal when the process trends to go out of control and cannot determine the source of the problem. As problem-solving tools, control charts allow operators to identify quality problems in processes as they occur.

ii) Cause-and-Effect or Fishbone Diagram: A cause-and-effect analysis diagram is also called fishbone diagram because it resembles a skeleton of a fish. It's one of the most famous six sigma tools, as it allows you to brainstorm various causes of a problem. This is a diagram which gives the relationship between quality characteristic and its factors. It is a pictorial presentation in which all possible causes and their effects are displayed. A problem can be due to a variety of reasons or causes. The solution to the problem becomes similar and easier if only true causes for the problem can be identified. "The cause-and-effect diagram was introduced by

Kaoru Ishikawa of Japan." It is the simple, graphical method for presenting a chain of causes and effects and for sorting out causes and organizing relationships between variables.

The first step in performing the analysis is to identify the problem people want to solve and have to write down who works on the process and when and where the process occurs. Next step is to write the problem in a box on the left-hand side of the paper. Then draw a horizontal line extending to the other end of the paper. From there, draw vertical lines extending off of the "spine". On these lines, write the major reasons behind the problem and think through any possible causes. One can further break these causes down into sub-causes. Once complete the diagram with all possible causes of the problem, one can analyze the results. The results may need further testing and analysis.

iii) Statistical Process Control (SPC): SPC is the application of statistical techniques to determine whether the output of a process conforms to the product or service design. In SPC, control charts are used primarily to detect production of defective products or services or to indicate that the production process has changed and that products or services will deviate from their design specifications unless something is done to correct the situation. [13]

iv) Tree Diagram: A tree diagram maps out the paths and tasks necessary to complete a specific project or reach a specific goal. This technique is used to seek answers to such questions as "What sequence of tasks will address this issue?" A structure tree is used to show the links or hierarchy of the ideas brainstormed. [13]

v) Histogram: In its basic form, it is the visual representation of the frequency distribution of data. With this six-sigma tool, we can quickly identify the variations in a graph. Subsequently, we can locate where problems exist in processes so that we can rectify them with corrective measures. The significant thing in applying corrective measures is that we need to find the root causes accurately and then implement the measures. Only this approach will produce better results.

viii) Process Mapping: Business Process Mapping is a way to visualize the business process and better understand how it works. A typical map outlines the roles, responsibilities, and standards involved in the process. It presents this data in a structured way that shows the steps of the process along with that is responsible, what the inputs and outputs are, and other information relevant to the process.

Business process mapping is a great aid in problem-solving. It visualizes the entire process, making it easier to see what's wrong and get straight to the root cause. It also helps to visualize the roles of the people within the process and ensure everyone knows what to do. In addition to that, business process mapping is great for finding the potential risks the process creates. Constructing a map makes you rethink each step of the process and see if there are any liabilities hidden within.

Types of Process Maps:

i) **Flowchart** is the most common type of a process map. While this process map is less flexible, one can easily draw it by hand or in software like MS Office. Flowcharts are often

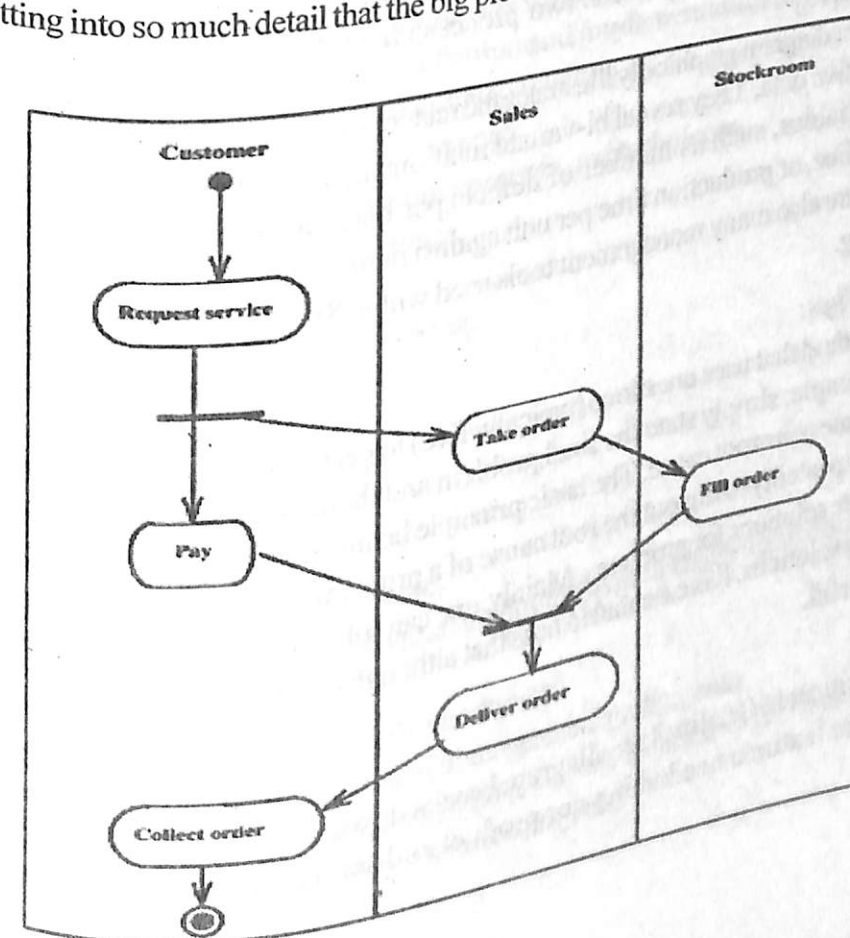
used for the creation of workflow diagrams. A flowchart is used to show details of a process, including tasks and procedures, alternative paths, decision points, and rework loops. A flowchart can be depicted as an "as-is" map showing a process as it currently works or as a "should-be" map showing how it ought to work. The level of detail will vary, depending on the objective. Many Black Belts now use software to draw their flowcharts but often start with a bunch of stickles on a wall. [13]

ii) **Swimlane diagram** is similar to generic flowcharts, except it is better structured in terms of who does what.

iii) **Value stream map** is a more in-depth alternative to the flowchart. It is common for lean six sigma operations, and find it harder to analyze at a glance.

iv) **SIPOC (Supplier Inputs Processes Outputs Customer)** is the most visually simplified map. It focuses on the essentials of the process and the people involved. By stripping all extra information, it defines a complex project better in terms of its basic elements.

SIPOC is used in the define phase of DMAIC and is often a preferred method for diagramming major business processes and identifying possible measures. The SIPOC diagram is used to show major activities, or sub processes, in a business process, along with the framework of the process, represented by the Suppliers, Inputs, Outputs, and Customers. A SIPOC diagram is used to help define the boundaries and critical elements of a process without getting into so much detail that the big picture is lost. [13]



ix) Value Stream Mapping: This is a six-sigma tool that uses a flowchart to outline the information as well as materials required for process improvement. Every flowchart has three core elements: *process map*, *information*, and *timeline*. With this flowchart, we can clearly visualize information such as time, delays, error rate, etc. Overall, we can streamline process flow with the necessary information required by using this six-sigma tool.

x) FMEA: FMEA is the short form for *Failure Modes and Effects Analysis*. Proactiveness is the basic principle behind this six-sigma tool. According to this method, anyone needs to find potential failures in processes and implement corrective measures before any failure can happen. Besides, it is one of the cost-effective tools since it allows making corrections in processes before implementing them in a production environment.

xi) Project Charter: Project Charter is a document that defines the purpose and the scope of the project. It works both as the blueprint for the business process and as legal authorization of the project. Project charter usually includes the project overview and its scope, details about the team and the resources, and the timeline. It gives you all the basic information about the project and clarifies the main points about it. The main benefit of a project charter is that it keeps things less chaotic. Once a team dives into the project, it is easy to lose track of who is responsible for what, which deadlines the team has to meet first, etc. If the company doesn't have a clear managerial hierarchy, things will get even messier.

xii) Scatter Diagram: Scatter diagrams are used to determine whether relationship really exists between two process two process characteristics and the direction of the relationship. [13]

A scatter diagram graphically illustrates the relationship between variables, typically based on quantitative data. They reveal bi-variate relationships, which are relationships between pairs of variables, such as number of defects per batch against changes in the speed of production line, or production time per unit against hours of training. [9]

There are also many management tools used within Six Sigma. Some examples include the following.

Five Whys:

This is a method that uses questions (typically five) to get to the root cause of a problem. The method is simple: simply state the final problem and then ask the question "why," breaking down the issue to its root cause. The basic principle behind this six-sigma tool is asking 'why' five times repeatedly to dig out the root cause of a problem. This method helps to choose the most accurate solutions for problems. Mainly, one can solve complex problems efficiently and derive better solutions. It is essential to note that although this method looks simply, the results are so powerful.

CTQ Tree:

The Critical to Quality Tree diagram breaks down the components of a process that produces the features needed in your product and service if you wish to have satisfied customers.

Root Cause Analysis:

Much like the Five Whys, this is a process by which a business attempts to identify the root cause of a defect and then correct it, rather than simply correcting the surface "symptoms."

All the Six Sigma tools and methodologies serve one purpose: to streamline business processes to produce the best products and services possible with the smallest number of defects. Its adoption by corporations around the globe is an indicator of its remarkable success in today's business environment.

13. Why are Six Sigma tools so Important

There are plenty of reasons for that. At a glance, we will look at some of the reasons below:

- i) Six sigma tools drive transformation in business processes slowly and positively and enhance productivity to high levels.
- ii) With six sigma tools, you can minimize the frequency of maintenance in a production environment
- iii) They help to prevent IT threats. As a result, you can reduce downtime to a minimum.
- iv) Ultimately, six sigma tools provide cost-effectiveness, optimized use of resources, improved cycle time, reliable products, and many more.

14. Successful implementation of Six Sigma

Antony and Banuelas (2002) and Banuelas Coronado and Antony (2002) presented the key ingredients for the effective introduction and implementation of six sigma in UK manufacturing and services organizations as the following:

- i) Management commitment and involvement.
- ii) Understanding of six sigma methodology, tools, and techniques.
- iii) Linking six sigma to business strategy.
- iv) Linking six sigma to customers.
- v) Project selection, reviews and tracking.
- vi) Organizational infrastructure.
- vii) Cultural change.
- viii) Project management skills.
- ix) Linking six sigma to suppliers.
- x) Training.
- xi) Linking six sigma to human resource.

Johnson and Swisher (2003) provided useful implementation tips for successful six sigma applications as:

- i) Sustained and visible management commitment.
- ii) Continuing Education and training of managers and participants.
- iii) Setting clear expectations and selecting project leaders carefully for leadership skills.
- iv) Picking and selecting strategically important projects.

Starbird (2002) argued that six sigma processes are part of a management system to achieve business excellence in organizations and presented keys to six sigma successes as:

- i) Start process management: identify core processes, customer needs, and measures.
- ii) Drive performance through reporting: Leaders must maintain and report opportunity lists, status of active projects/resources, and results from finished projects.
- iii) Consistent and visible leadership involvement.
- iv) A measurement system to track progress, providing accountability for the initiative
- v) Integrate championing of active projects: Select and charter projects and require updates during existing staff meetings.
- vi) Internal and external benchmarking of the organization's products, services, and processes. You must find out where you really are.
- vii) Setting challenging stretch goals that focus your employees on changing the process, not just tweaking it.
- viii) Educating and informing every member of your organization about the Six Sigma methodology.
- ix) Developing the infrastructure to support change throughout your organization.
- x) Working to create a "cause," not just a business success.

Successful Six Sigma implementation requires an ideological shift, everyone in the organization from the top down will need to change their behaviours to address problems once and for all. Leadership has a crucial role in the success of Six Sigma implementation in an organization or company. True improvements come when organizational leaders display a willingness to change. The best way to successfully implement a Six Sigma program is to start with a pilot project. One can identify a company process that generates defects or has other problems, usually production. The process of identifying a pilot project has to involve the people carrying out the work and consider their input. Six Sigma only works when everyone is involved.

15. Application of Six Sigma

The Six Sigma method focuses on understanding customers' requirements better and eliminating defects and waste. These objectives are achieved through profound knowledge of statistics, engineering and project management, as well as the underlying processes and systems. Objective of Six Sigma projects are to improve the organization's products, services and processes across various disciplines, including manufacturing, production, marketing, finance, and administration.

a) Manufacturing Sector

Motorola was the first organization to use the term six sigma in the 1980s as part of its quality performance measurement and improvement program. Six sigma has since been successfully applied in other manufacturing organizations such as *General Electric*, *Boeing*, *DuPont*, *Toshiba*, *Seagate*, *Allied Signal*, *Kodak*, *Honeywell*, *Texas Instruments*, *Sony*,

etc. The reported benefits and savings are composed and presented from investigating various literatures in six sigma (*Weiner, 2004; de Feo and Bar-El, 2002; Antony and Banuelas, 2002; Buss and Ivey, 2001; McClusky, 2000*).

Six Sigma in the manufacturing sector or industry has one guiding premise- the elimination of variation within the product lifecycle, ultimately ensuring each run aligns with an ideal production outcome. This goal has distinct relevancy for those in manufacturing, including many of the following:

- i) Product superiority
- ii) Financial visibility
- iii) Customer delight
- iv) Business stability

b) Financial Sector

Defects occur in every industry, including the finance industry or sector. And Six Sigma is about reducing defects to 3.4 for every million opportunities. That may seem ambitious, but Six Sigma makes it possible. In recent years, finance and credit department are pressured to reduce cash collection cycle time and variation in collection performance to remain competitive. Typical six sigma projects in financial institutions include improving accuracy of allocation of cash to reduce bank charges, automatic payments, improving accuracy of reporting, reducing documentary credits defects, reducing check collection defects, and reducing variation in collector performance (*Doran, 2003*).

Bank of America (BOA) is one of the pioneers in adopting and implementing six sigma concepts to streamline operations, attract and retain customers, and create competitiveness over credit unions. It has hundreds of six sigma projects in areas of cross selling, deposits, and problem resolution. BOA reported a 10.4% increase in customer satisfaction and 24% decrease in customer problems after implementing six sigma (*Roberts, 2004*). *American Express* applied six sigma principles to improve external vendor processes, and eliminate non-received renewal credit cards. The result showed an improved sigma level of 0.3 in each case (*Bolt et al., 2000*).

c) Healthcare Sector

Six sigma principles and the healthcare sector are very well matched because of the healthcare nature of zero tolerance for mistakes and potential for reducing medical errors. [13]. As more and more healthcare practitioners become increasingly aware of issues around quality of care, Six Sigma in healthcare is becoming an increasingly popular concept. Every business seeks to improve its processes in order to meet and exceed customer satisfaction. Now a days, in healthcare, the primary focus is on improving patient safety and satisfaction. In addition, patients look out for quality as a requirement choosing a healthcare service. In addition, patients can make more informed decisions about their treatments based on their experiences and level of satisfaction. This is where Six Sigma in healthcare plays a vital role.

Some of the successfully implemented six sigma projects include improving timely and accurate claims reimbursement (*Lazarus and Butler, 2001*), streamlining the process of healthcare delivery (*Ettinger, 2001*), and reducing the inventory of surgical equipment and related costs (*Revere and Black, 2003*).

d) Engineering And Construction Sector

In 2002, *Bechtel Corporation*, one of the largest engineering and construction companies in the world, reported savings of \$200 million with an investment of \$30 million in its six-sigma program to identify and prevent rework and defects in everything from design to construction to on-time delivery of employee payroll (*Eckhouse 2004*). [13]. For example, six sigma was implemented to streamline the process of neutralizing chemical agents, and in a national telecommunications project to help optimize the management of cost and schedules (*Moreton, 2003*).

e) Research And Development Sector

The objectives of implementing six sigma in Research and development organizations are to reduce cost, increase speed to market, and improve Research & Development processes. To measure the effectiveness of six sigma, organizations need to focus on data driven reviews, improved project success rate, and integration of Research & Development into regular work processes. [13]

One survey noted that as of 2003 only 37% of the respondents had formally implemented six sigma principles in their Research & Development organization (*Johnson and Swisher, 2003*). *Rajagopalan et al.* (2004) reported that the development and manufacturing of the new prototype at W.R. Grace (Refining Industry) was cut to 8–9 months from 11–12 months by implementing the DFSS process.

f) Food Industry/ Sector

Six Sigma is widely used methodology in the food industry/ sector to ensure the production of high-quality products. Six Sigma is a data-driven approach that aims to reduce defects & improve efficiency. One of the key benefits of Six Sigma in the food industry is that it can help companies to identify and eliminate sources of food contamination. Companies in various industries, including healthcare, service and manufacturing have widely adopted it. In the food industry, Six Sigma can be used to improve the quality and safety of food products, increase efficiency and reduce costs. The food industry is a highly regulated sector, and food companies must comply with strict laws & regulations to ensure the safety and quality of their products. Six Sigma can help food companies to meet these regulations by identifying and eliminating sources of defects. Six Sigma can be applied in the food industry in various ways, depending on the specific needs & goals of the company. Some of the key ways in which Six Sigma is applied in the food industry include:

- i) Identifying & eliminating sources of food contamination.
- ii) Improving process efficiency.

- iii) Creating a culture of quality & safety.
- iv) Improving customer experience.
- v) Compliance with regulations.
- vi) Supply chain management.

McDonald's, Nestle, Coco-Cola etc. are example of the food companies that have successfully implemented Six Sigma.

16. In What Industries / Business sector is Six Sigma most Prevalent?

Now days, many fortune 500 companies report implementing Six Sigma management methodologies into their organizations. The following manufacturing subfields, in particular, have adopted Six Sigma over the years to successful results:

- i) Automotive
- ii) Construction
- iii) Consumer electronics
- iv) Energy
- v) Food & beverage
- vi) General consumer goods
- vii) Heavy machinery & industrial equipment
- viii) Mining
- ix) Transportation

17. Real world examples of Six Sigma

Today, countless organizations have seen success implementing Six Sigma policies. Some notable success stories as follows:

MICROSOFT: Six Sigma drove success at Microsoft by paving the way for improvements that benefited employees & consumer alike.

MOTOROLA: One of the first instance of Six Sigma's use by a major corporation Motorola implemented the strategy to improve products quality.

BOEING AIRLINES: Six Sigma Methodology empowered Boeing Airlines to fix an engine air fan problem.

GENERAL ELECTRIC: Integrating Six Sigma into all operations helped General Electric pinpoint previously unnoticed defects that slowed production.

LINDE GROUP and so on.

18. Future of Six Sigma

Six sigma is likely to remain as one of the key initiatives to improve the management process than just being remembered as one of the fads (*Johnson and Swisher, 2003*). The primary focus should be on improving overall management performance, not just pinpointing and counting defects. Researchers and practitioners are trying to integrate six sigma with other existing innovative management practices that have been around to make six sigma method even more attractive to different organizations that might have not started or fully

implemented the Six sigma method. Six sigma will be around as long as the projects yield measurable or quantifiable bottom-line results in monetary or financial terms. When six sigma projects stop yielding bottom-line results, it might disappear. While six sigma will evolve in the forthcoming years, there are some core elements or principles within six sigma that will be maintained, irrespective of the "next big thing". One of the real dangers of six sigma is to do with the capability of black belts (the so-called technical experts) who tackle challenging projects in organisations. We cannot simply assume that all black belts are equally good and their capabilities vary enormously across industries (manufacturing or service), depending a great deal on the certifying body. Another danger is the attitude of many senior managers in organisations that six sigma is "an instant pudding solving all their ever-lasting problems. However, the total package may change in the evolutionary process [13].

It is important to remember that six sigma has a better record than *total quality management (TQM)* and *business process re-engineering (BPR)*, since its inception in the mid-late 1980s. The ever-changing need to improve will no doubt create needs to improve the existing six sigma methodology and hence develop better products and provide better services in the future [13].

The following are some of the limitations of six sigma which create opportunities for future research:

- i) The challenge of having quality data available, especially in processes where no data is available to begin with (sometimes this task could take the largest proportion of the project time). [13]
- ii) In some cases, there is frustration as the solutions driven by the data are expensive and only a small part of the solution is implemented at the end. [13]
- iii) The right selection and prioritisation of projects is one of the critical success factors of a Six sigma program. The prioritisation of projects in many organisations is still based on pure subjective judgement. Very few powerful tools are available for prioritising projects and this should be major thrust for research in the future. [13]
- iv) The statistical definition of six sigma is 3.4 defects or failures per million opportunities. In service processes, a defect may be defined as anything which does not meet customer needs or expectations. It would be illogical to assume that all defects are equally good when we calculate the sigma capability level of a process. For instance, a defect in a hospital could be a wrong admission procedure, lack of training required by a staff member, misbehaviour of staff members, unwillingness to help patients when they have specific queries, etc. [13]
- v) There is an overselling of six sigma by too many consulting firms. Many of them claim expertise in six sigma when they barely understand the tools and techniques and the Six sigma roadmap. The relationship between *cost of poor quality (COPQ)* and process sigma quality level requires more justification. [13]
- vi) The linkage between six sigma and organisational culture and learning is not addressed properly in the existing literature [13].

One of the limitations of the fact-driven process of identifying a problem and working toward a solution is that it tends to leave out a key component: humans- and more importantly, how humans' impact and work through different obstacles. Six sigma also does not technically allow for the introduction of new tools or methods, even when they could be beneficial. Since Six sigma generally requires total dedication across all teams, it's difficult to use or experiment with other process methodologies for other areas of the organization.

19. Conclusion

Six Sigma is a data-driven methodology that provides tools and techniques for making business processes more effective and efficient. The term "sigma" is used to designate the distribution or the spread about the mean of any process. Sigma measures the capability of the process to perform defect-free work. A defect is anything that results in customer dissatisfaction. For a business process, the sigma value is a metric that indicates how well that process is performing. Higher sigma level indicates less likelihood of producing defects and hence better performance [13].

Six sigma is a performance standard to achieve operational excellence. With six sigma, the common measurement index is "defects-per-unit" where a unit can be virtually anything- a component, piece of material, administrative form etc. Conceptually, six sigma is defined as achieving a defect level of 3.4 ppm or better. Operationally, six sigma is defined as staying within half the expected range around the target.

The objective of this article has been to review the effect of six sigma as a tool to improve quality and profitability and its most important aspects. Six Sigma is generally described as a metric, a mindset & a methodology for strategic management and process improvement. Six Sigma provides an effective mechanism to focus on customer requirements, through improvement of process quality. The Six Sigma Approach is customer-driven.

Six sigma is a business methodology for quality improvement that measures how many defects there are in a current process and seeks to systematically eliminate them. The Six Sigma Approach is also data-driven. It focuses on reducing process variation, cantering the process and on optimizing the process. Various approaches to six sigma have been applied to increase the overall performance of different business sectors and organizations. □

References

- [1] James E. Brady, Theodore T. Allen, "Six Sigma Literature: A Review and Agenda for future Research," *Quality and Reliability Engineering International*, 2006.
- [2] Hongbo Wang, "A Review of Six Sigma Approach: Methodology, Implantation and Future Research," 2008.
- [3] Weiyong Zhang & Arthur V. Hill & Glenn H. Gilbreath, "Six Sigma-A Retrospective & Prospective study", 2009.
- [4] P. K. Bharti, M. I. Khan, Harbinder Singh, "Six Sigma approach for quality and performance excellence in plastic injection molding industry", 2011.

- [5] Young Hoon Kwak, Frank T. Anbari; "Benefits, obstacles, and future of six sigma Approach", 2006.
- [6] Miroslav RUSKO & Ruzena KRALIKOVA. "Application of six sigma method to EMS Design", 2011.
- [7] Lin Man-shan, Cai Chun-xing, Li Jin-hong. "Realization & Research of a Six-Sigma- based project management & analysis system". 2009
- [8] K. Shridhara Bhat. "Total Quality Management"; text & cases. Himalaya Publishing House, pp.294-295. 2011
- [9] Brue, G. "Six Sigma for small Business", *Entrepreneur Media*. Madison. 2006.
- [10] Brady, J.E. & Allen, T.T. "Six sigma literature: A review and agenda for future Research". *Qual. Reliab. Eng. Int.*, 22(3), 335-367, 2006.
- [11] Folaron, J. "The Evolution of Six Sigma". *Six Sigma Forum Magazine*, 2(4), 38-44, 2003.
- [12] George, M.L. "Lean Six Sigma". New York, NY: McGraw-Hill. 2002.
- [13] Gupta, N. "An Overview on Six Sigma: Quality Improvement Program." *International Journal of Technical Research and Applications*, 1(1), 29-39, 2013.
- [14] Godfrey, A.B. "In the beginning: Six Sigma"; *Forum Magazine*, 1(3), 46-49, 2002.
- [15] Hahn, G.J., Doganaksoy, N. & Hoerl, R.W. "The evolution of six sigma"; *Qual. Eng.*, 12(3), 317-326, 2000.
- [16] Hahn, G.J., Hill, W.J., Hoerl, R.W. & Zinkgraf, S.A. "The impact of six sigma Improvement", *A glimpse into the future of statistics. Amer. Statist.*, 53(3), 208-215, 1999.
- [17] Hoerl, R.W. "Six Sigma Black Belts: What do they need to know?" (With discussion). *J. Qual. Technol.*, 33(4), 391-435, 2001
- [18] Hoerl, R.W. "One perspective on the future of Six-Sigma". *Int. J. Six Sigma Competitive Advantage*, 1(1), 112-119, 2002.
- [19] Bowser, M.D. & Holcomb, D.R. "A university-based Six Sigma program". *Qual. Reliab. Eng. Int.*, 21(3), 243-248, 2005.
- [20] Nonthaleerak, P. & Hendry, L.C. "Six Sigma: Literature review and key future research Areas". *Int. J. Six Sigma Competitive Advantage*, 2(2), 105-161, 2006.
- [21] Watson, G. "Building on Six Sigma effectiveness". *ASQ Six Sigma Forum Magazine* 5 (4), 14-15, 2006
- [22] Rajagopalan, R., Francis, M., Suarez, W., "Developing novel catalysts with Six Sigma. Research—Technology Management" 47, 13-16, 2004.
- [23] V. Antony, J., Kumar, M., Madu, C.N., "Six Sigma in small and medium sized UK manufacturing enterprises: Some empirical observations". *International Journal of Quality & Reliability Management* 22 (8), 860-874, 2005.
- [24] Chakravorty, S.S., "Six Sigma Programs: An Implementation Model". *Int. J. Production Economics*, 1-16, 2005.
- [25] Chakravorty, S.S., Hales, D.N., "The evolution of manufacturing cells: an action research study." *European Journal of Operational Research* 188, 153-168, 2008.
- [26] Foster, S.T., "Does Six Sigma improve performance?"; *The Quality Management Journal* 14 (4), 7-20, 2007.
- [27] Mast, J.D., Bisgaard, S., "The science in Six Sigma. Quality Progress". 40 (1), 25-29, 2007.

From Certainty to Uncertainty- A paradigm shift

Deepika Bhattacharjee
Department of Mathematics
Tinsukia College, Tinsukia, Assam, India

Throughout the history of science, there has been a gradual transition from traditional views to the relatively modern views. This type of transition is usually referred to as a paradigm shift, which happens periodically. The concept of scientific paradigm was introduced by Thomas Kuhn in his famous book *The Structure of Scientific Revolutions* (University of Chicago Press, 1962). Scientific Paradigm is defined as a set of theories, standards, principles and methods that are taken for granted by the scientific community in a given field. In his book, Kuhn explains the notion of paradigm shift with the help of some well known examples which can be considered as scientific revolutions. Copernicus in the field of Astronomy, Newton in the field of Mechanics, Darwin in the field of Biology were some of the most popular scientists associated with remarkable paradigm shifts.

Although each paradigm shift has its own unique set of characteristics, they also share some common properties. Each paradigm shift came into existence when the scientific fraternity could not explain a problem with the existing paradigm. Initially each new paradigm when proposed are either rejected or ignored by most of the scientists in the respective field but gains its acceptance among the experts when it can be proved that the new paradigm is more successful than the existing paradigm in dealing with the critical problems.

In general, to deal with a problem models are used which are either found in nature or are some desirable man-made objects. The purpose of using a model is to understand the phenomenon in a better way, making predictions and also to learn how to control the phenomenon in a desirable way. While Mathematical theories are developed for "perfect"

objects, in practical experiments it can be observed that the properties demanded by the theory are only approximately true. For example, to understand the theory of spheres we can consider the shape of an orange but this is definitely an imperfect copy (or approximate) of an ideal sphere.

Among the numerous paradigmatic changes in the field of Science and Mathematics, the concept of Uncertainty is one of the most significant changes. Traditionally it is expected that Science should strive for certainty in all respects and what is uncertain is considered as unscientific. But uncertainty is unavoidable and hence the concept of uncertainty is essential to scientific discussions also.

The introduction of the concept of Fuzzy set marks significant paradigm shift in Mathematics. It emerged from the need to bridge the gap between the mathematical models and their empirical interpretations. This gap is very much disturbing in the areas of social sciences, biological sciences or in case of medicines.

The modern concept of uncertainty was introduced by Lotfi A. Zadeh in his seminal paper [1965b], "Fuzzy sets", Information and control, 8(3), pp.338-353. In this article we discuss the concept of a Fuzzy Set and will try to analyze how it is different from the classical (non fuzzy) set which is generally referred to as Crisp set.

We start with the concept of a crisp set which is most basic in Mathematics. In almost all Mathematical system we require certain collection of sets.

A (classical or crisp) set is defined as a collection of well defined objects. By the terms "well defined" we mean that what is there in the collection and what is not should be very clear to us. For example the collection of all vowels of English alphabet is a set because the collection contains the five members viz, a, e, i, o and u and nothing else. On the other hand the collection of all important rivers of India is not a set because the term important is not "well defined". For some person Mahanadi or Tapti River is important but there may be some person who may not consider these rivers as important rivers of India. The concept of "Important river" is very much relative and so this collection is not considered as a set. Sets are denoted by uppercase letters and their members by lowercase letters. A few examples of sets used in Mathematics are:

N (collection of Natural numbers), Z (set of Integers), Q (set of Rational numbers), R (set of Real numbers) etc. Thus 5 being a natural number belongs to N but 0.7 does not belong to N.

Representation of a set:

In general any set can be represented by one of the following methods :

1. The Tabular method or List method

In Tabular method all the elements of the set are listed, the elements being separated by commas and enclosed within braces { }. For example, if V is the set of vowels of English alphabets then we represent this set in Tabular method as $V = \{ a, e, i, o, u \}$.

2. The Rule method or Set builder method

In Rule method, we describe the elements of the set by using a symbol x (or any other symbols like y, z, u etc.) followed by a ":", and after the sign of colon, we write the common property which is satisfied by all the elements of the set and enclose the whole description within braces { }. Thus the above set V can be represented by using Rule method as $V = \{ x : x \text{ is a vowel of English alphabet} \}$.

3. Using Characteristic function

Let us now explain another method to represent a set where a set is defined by a function. Let U denotes the collection of all the possible members in a particular context that is, the universe of discourse. We call this set as the Universal set. Then we define a set by a function called Characteristic function or membership function that declares which member of U are the members of the set and which are not. Thus a set A is defined by using characteristic function as follows: For any element x of the universal set U,

$$A(x) = 1 \text{ when } x \text{ is a member of } A \text{ and}$$

$$A(x) = 0 \text{ when } x \text{ is not a member of } A.$$

We now explain this method for the above mentioned example of the set V of all vowels of English alphabet. Clearly in this example, the universal set U consists of all the alphabets of English and the set V is defined as $V(x) = 1$ or 0 according as x is a vowel or not where x represents any member of U. Thus we have $V(a) = V(e) = V(i) = V(o) = V(u) = 1$ and $V(x) = 0$ for all the remaining alphabets which are not vowels. This way of representing a set becomes very useful to handle the concept of uncertainty.

From the above discussion, it is clear that the definition of Crisp set dichotomizes the items into two groups: members (which certainly belong to the set) and non members (those that certainly not belong). However in many cases in our day to day life we come across the situations where this concept of classification does not work. When we talk about set of intelligent students, expensive cars, highly contagious diseases or set of very large numbers etc. we understand that these sets do not have precise boundaries and the concept of membership or non membership is meaningless in such cases. In such scenarios the concept of characteristic function becomes relevant. We have seen that the characteristic function of a crisp set assigns a value of either 1 or 0 to each individual in the universal set, thereby discriminating between members and nonmembers. This function can be generalized such that the values assigned to the elements of the universal set are within a specified range to indicate the degree or grade of membership. When the assigned value of the membership function is large for an element of U then this will indicate the higher grade of membership for that particular element and similarly the smaller values indicates lower degree of membership. Thus the problem of dichotomization in case of Classical sets into either membership or nonmember ship can be resolved. This set which is defined by using a membership function having different values in a given range is called a Fuzzy Set. The most commonly used range of values of membership function is the unit interval $[0, 1]$. In case of classical set, the membership

function takes the value either 0 or 1 whereas in a fuzzy set the membership can take any real number between 0 and one 1 depending on the degree of membership of a particular individual on the fuzzy set. Thus a fuzzy set allows us to represent vague concepts which we encounter in natural language. Let us now consider the example of collection important rivers of India which is not a classical or crisp set as explained earlier can now be considered as a fuzzy set. If A denotes this fuzzy set then we can assign different membership values to different rivers to indicate the degree of importance of the river. The larger the membership value, the greater the degree of importance of the river. If the rivers Ganga, Yamuna, Brahmaputra, Mahanadi, Narmada, Tapti are represented by G, Y, B, M, N and T respectively then we can describe the fuzzy set A as follows:

$$A = \{ (G, 0.9), (Y, 0.8), (B, 0.85), (M, 0.7), (N, 0.73), (T, 0.6), \dots \}.$$

Here the universal set U is the crisp set containing the rivers of India and A is the Fuzzy set of important rivers of India. The three dots at the end of the description of fuzzy set A indicate that some other members of U with their respective grade of importance can be included in the set.

Similarly in a fuzzy set representing the concept of sunny day might assign the degree of membership 1 if the cloud cover is 0%, 0.7 to a cloud cover of 30% and 0.1 to a cloud cover of 90% etc.

Thus the significance of fuzzy sets can be best established by the fact that they accommodate flexibility by facilitating the gradual transition between states and possess natural capability to express and deal with measurement of uncertainties.

A brief note on Pi

Ranjan kumar Gogoi
Department of Mathematics
Tinsukia College, Tinsukia, Assam

Pi is the world most popular mathematical constant. The Greek letter π is the symbol used in mathematics to represent a constant—the ratio between circumference and diameter of a circle which is approximately 3.14159. This constant is perhaps most popular irrational mathematical constant widely used in various branches of mathematics and physics. The symbol was first used in 1706 by British mathematician William Jones. This symbol was later developed by Swiss mathematician Leonhard Euler in the 18th century and it is now used universally to denote this mathematical constant.

The history of Pi is very ancient. Around 2000 BC Babylonian mathematicians first calculated the area of a circle by taking 3 times the square of its radius which gave a value of Pi was 3. Later they calculated the value 3.125 for Pi which was a little closer approximation. In 1650 BC Rhind Papyrus, one of the greatest mathematician of Egypt calculated the area of a circle by formula that gave approximate value 3.1605 for Pi.

The first rigorous approach to finding the value of Pi was based on geometrical approximation. Around 250 BC, Archimedes a Greek mathematician, known as father of mathematics calculated a more accurate value of Pi by using Pythagorean approach. Archimedes drew square both around the outside and within the interior of a circle. He then using hexagon and polygons with more and more sides and thus ultimately calculated three accurate digits 3.14 for Pi. In 150 AD, Greek Roman mathematician Ptolemy used this method to calculate the value 3.1416.

Many mathematician tried to calculate the values of Pi in more accurate as it had used in different formulae of mathematics and physics from the earliest. In 5th century Chinese mathematician Zu Chongzhi approximated Pi to seven digits. The first computational formulae for calculating the value of Pi based on infinite series were used by ancient mathematicians. In

1500 century, Indian mathematicians discovered that Pi could be represented by the infinite series as

$$\pi/4 = 1 - 1/3 + 1/5 - 1/7 + \dots$$

In 1630 Christoph Grienbeger calculated the value of Pi to 38 digits which was the most accurate approximation of Pi using polygon algorithm. In 1665 English mathematician Sir Isaac Newton used infinite series expansion of the arcsine function

$$\text{Arcsine}(x) = x + \frac{1}{2} \frac{x^3}{3} + \frac{1}{2 \cdot 4} \frac{x^5}{5} + \frac{1}{2 \cdot 4 \cdot 6} \frac{x^7}{7} + \dots$$

As arcsine(1/2) = $\pi/6$, so the series would be

$$\pi = 6 \left(\frac{1}{2} + \frac{1}{2 \cdot 3 \cdot 2^3} + \frac{1 \cdot 3}{2 \cdot 4 \cdot 5 \cdot 2^5} + \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6 \cdot 7 \cdot 2^7} + \dots \right)$$

using this series Newton calculated the value of Pi to 16 decimal places.

In 1748, Euler used the following infinite series to calculate the value of Pi

$$\frac{\pi^2}{6} = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \dots$$

$$\frac{\pi^2}{6} = \left(\frac{2^2}{2^2-1} \right) \left(\frac{3^2}{3^2-1} \right) \left(\frac{5^2}{5^2-1} \right) \left(\frac{7^2}{7^2-1} \right) \dots$$

One of the most well known and famous way to calculate Pi was to use the Gregory Leibniz series (1671-1674)

$$\pi/4 = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \dots$$

Another famous series which converges more quickly to get the more accurate approximate value of Pi was the Nilakantha series which was developed in the 15th century

$$\pi = 3 + \frac{4}{2 \cdot 3 \cdot 4} - \frac{4}{4 \cdot 5 \cdot 6} + \frac{4}{6 \cdot 7 \cdot 8} - \frac{4}{8 \cdot 9 \cdot 10} + \dots$$

The great Indian mathematician Ramanujan (1914) discovered the formulae for computing the value of Pi rapidly. His formulae were

1) Ramanujan 1 1914

$$\frac{1}{\pi} = \frac{\sqrt{8}}{99^2} \sum_{n=0}^{\infty} \frac{(4n)!}{(4^n n!)^4} \frac{1103+26390n}{99^{4n}}$$

2) Ramanujan 2 1914

$$\frac{4}{\pi} = \frac{1}{882} \sum_{n=0}^{\infty} \frac{(-1)^n (4n)!}{(4^n n!)^4} \frac{1123+21460n}{882^{2n}}$$

In 1987, Chudnovsky brothers discovered the Ramanujan-type formula that converges more rapidly. The formula was

$$\frac{1}{\pi} = 12 \sum_{n=0}^{\infty} \frac{(-1)^n (6n)!}{(3n)! (n!)^3} \frac{13591409+545140134n}{(640320^3)^{n+\frac{1}{2}}}$$

Efforts were made to calculate the more accurate values for Pi and it reached 71 digits in 1699, 100 digits in 1706 and 620 digits in 1956 which were the best approximation calculated without using calculator or computer. In the mid-20 century mathematicians use computer for calculating the value of Pi more accurate. Since computer programmes can add up more and more terms of a series and calculate the value of Pi to extra ordinary degree of accuracy. In 2014 the world record was that a computer has calculated Pi correct to 13,300,000,000,000 decimal places. But this record is not so long. In 2019, Google has recorded the value of Pi to 31.4 trillion decimal places.



A Study on the Method of Finding Rank of Elliptic Curves

Debakanta Buragohain

Dept. of Mathematics,
Tinsukia College, Assam.
E-mail: debakanta8@gmail.com

Abstract

A curve C over $\mathbb{Q}$ is an equation $f(x, y) = 0$, where f is a polynomial: $f \in \mathbb{Z}[x, y]$. We denote $\mathbb{C}(\mathbb{Q})$, the set of rational points for curves which consists of pairs $(x, y) \in \mathbb{Q} \times \mathbb{Q}$ satisfying $f(x, y) = 0$. For curves of degree 1 and 2 we know how to write $\mathbb{C}(\mathbb{Q})$ as a parametrized set, therefore enabling us to know all the rational points for curves of these degrees. For irreducible, smooth curves of degree 3 equipped with at least one rational point, which are called elliptic curves. $\mathbb{C}(\mathbb{Q})$ form a group called Mordell-Weil Group. This group is always finitely generated abelian group and hence due to fundamental theorem of finitely generated abelian group, $\mathbb{C}(\mathbb{Q})$ is isomorphic to $\mathbb{Z}^r \oplus \mathbb{C}(\mathbb{Q})_{tors}$. Here r is a non-negative integer and $\mathbb{C}(\mathbb{Q})_{tors}$ is the subgroup of all rational points of finite order in $\mathbb{C}(\mathbb{Q})$. The non-negative integer r is called the rank of the elliptic curve over $\mathbb{Q}$. In this paper we will describe the basic theory of Mordell-Weil Group of an elliptic curves and then focus on rank.

1 Introduction

The most general cubic equation over $\mathbb{Q}$ is given by

$$F(x, y) = ax^3 + bx^2 + cxy^2 + dy^3 + ex^2 + fxy + gy^2 + hx + iy + j = 0 \quad (1)$$

where $a, b, c, \dots, i, j \in \mathbb{Q}$.

A point (x_1, y_1) on a curve C is called singular if

$$\frac{\partial F}{\partial x}(x_1, y_1) = \frac{\partial F}{\partial y}(x_1, y_1) = 0$$

Definition 1. A curve without a singular point is called a non-singular curve. An elliptic curve over the field $\mathbb{Q}$ is a non-singular cubic curve in two variables with rational co-efficient. Elliptic curves are denoted by E instead of C .

By appropriate change of variables, a general elliptic curve over the field $\mathbb{Q}$ in (1) can be written in the form

$$y^2 = x^3 + ax + b, \quad a, b \in \mathbb{Q} \quad (2)$$

This particular form of an elliptic curve is known as Weierstrass Normal Form. Let us now prove the following two lemma. These two results can be used as alternative definitions of elliptic curves defined over $\mathbb{Q}$.

Lemma 1. A cubic curve defined by (2) is an elliptic curve (that is, nonsingular) if and only if $x^3 + ax + b = 0$ has no repeated roots.

Proof. Let $F(x, y) = y^2 - f(x)$, where $f(x) = x^3 + ax + b = 0$. If $f(x)$ has a double root at x_0 , then $f(x) = (x - x_0)^2(x - x_1)$. Now

$$\frac{\partial F}{\partial x} = -f'(x) = 2(x - x_0)(x - x_1) + (x - x_0)^2, \quad \frac{\partial F}{\partial y} = 2y$$

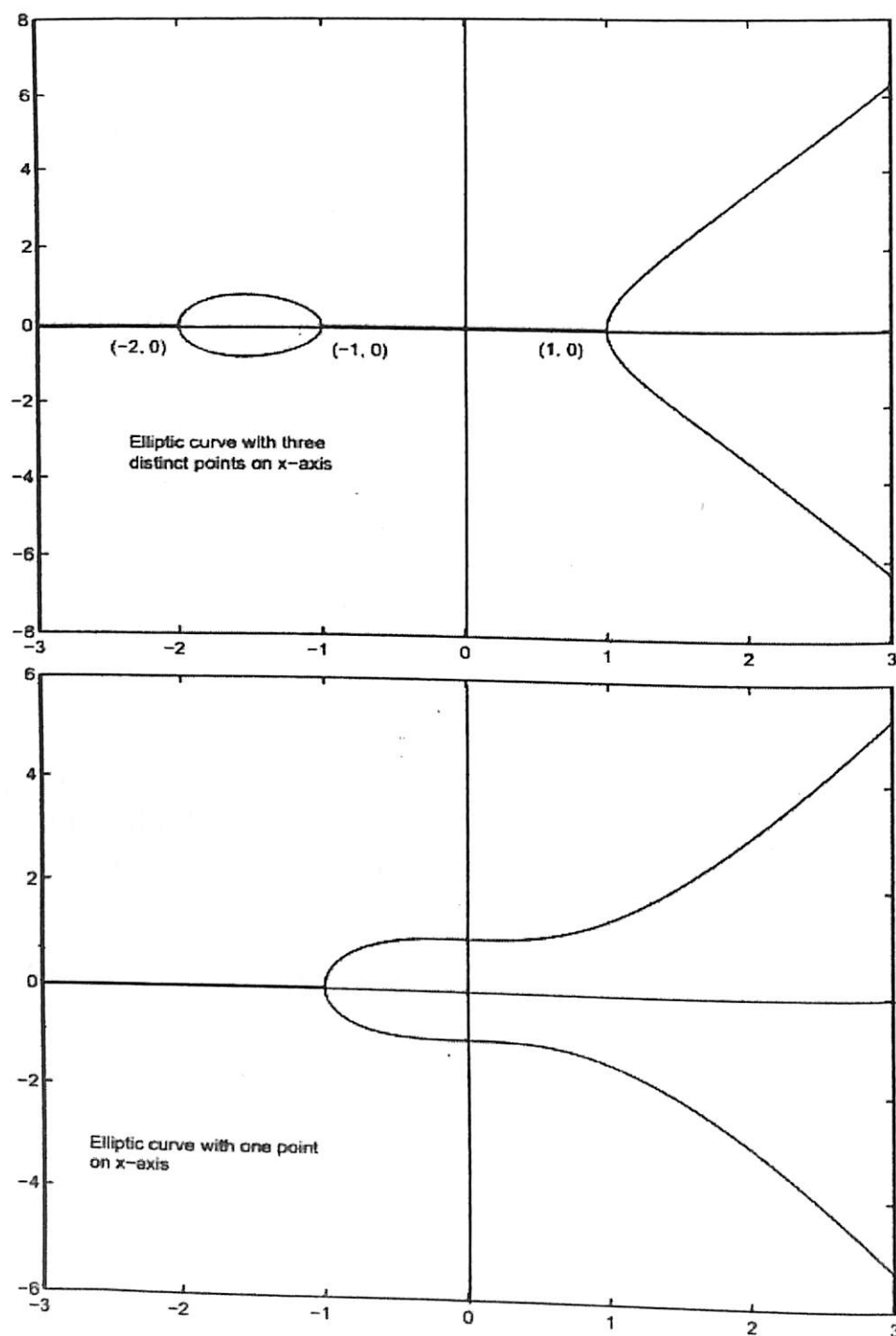
Hence $(x_0, 0)$ is a singular point on the curve. Again, if (x_0, y_0) is a singular point, then clearly $y_0 = 0$ and x_0 is a root of both $f(x) = 0$ and $f'(x) = 0$. This proves that x_0 is a repeated root of $f(x) = 0$. This completes the proof of the lemma. $\square$

Lemma 2. A cubic curve defined by (2) is an elliptic curve (i.e nonsingular) if and only if $4a^3 + 27b^2 \neq 0$.

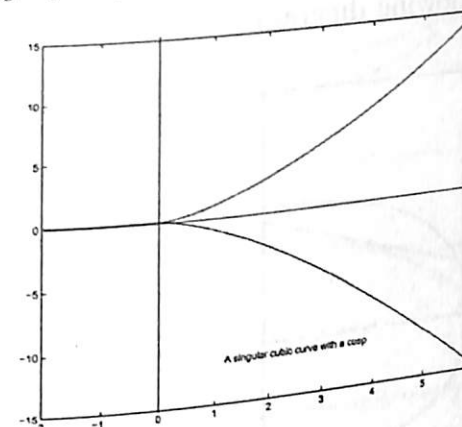
Proof. Clearly $f(x) = x^3 + ax + b = 0$ has repeated roots if and only if $f(x) = f'(x) = 0$. Solving these two equations, it is easy to deduce the required condition. $\square$

2 Examples of Singular Cubic Curves

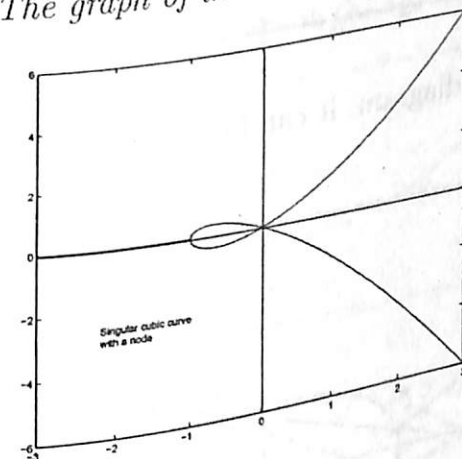
Let $E : y^2 = x^3 + ax + b$ be an elliptic curve over $\mathbb{Q}$. Clearly $x^3 + ax + b = 0$ will have one real root or all the three roots are real. Thus, we have two possible types of diagrams of elliptic curves. Below, diagram of two such elliptic curves are given.



Example 1. The cubic curve given by $y^2 = x^3$ is singular. Here $(0, 0)$ is a singular point. The graph of this singular curve is given below:



Example 2. The cubic curve given by $y^2 = x^2(x+1)$ is singular. Here $(0, 0)$ is a singular point. The graph of this singular curve is given below:



3 Mordell-Weil Group Law on Elliptic Curves

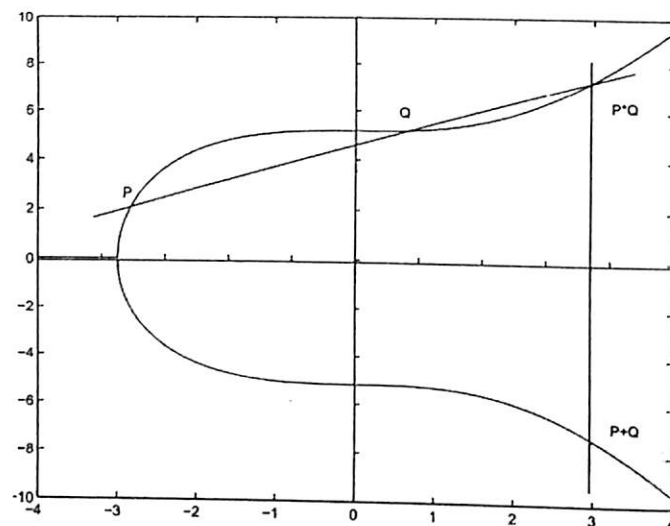
Let $E : y^2 = x^3 + ax + b$ be an elliptic curve over the field of rational numbers $\mathbb{Q}$. Let

$$E(\mathbb{Q}) = \{(x, y) \in \mathbb{Q}^2 : y^2 = x^3 + ax + b\} \cup \{O\} \quad (3)$$

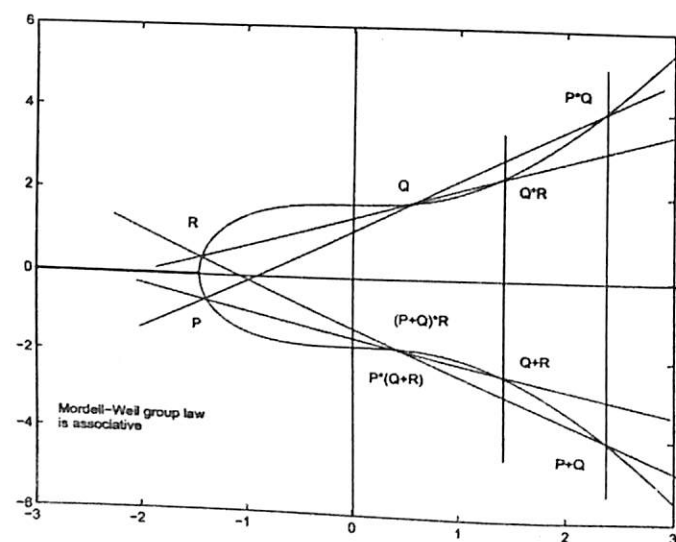
be the set of all rational points on the elliptic curve with an extra point O at infinity. One can imagine this point at infinity as a point lying on every line parallel to y -axis. Let P and Q be any two rational points on the elliptic curve E . We can uniquely determine the third point which is the intersection of the curve with the line through P and Q , and the third point is denoted by $P * Q$. We now define a binary operation $+$ on $E(\mathbb{Q})$ given by

$$P + Q = O * (P * Q),$$

which is the third point of intersection of the curve with the line parallel to y -axis and passing through $P*Q$. If $P*Q = (x, y)$, then clearly $P+Q = (x, -y)$. This is shown in the following diagram.

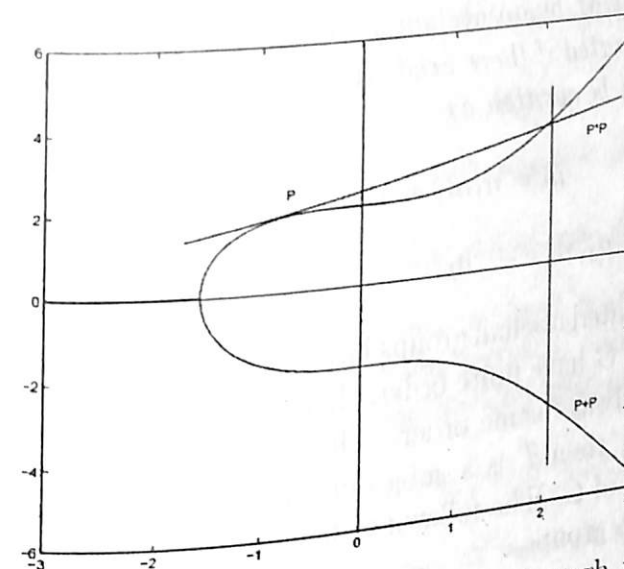


From the following diagram, it can be seen that the Mordell-Weil group law is associative.

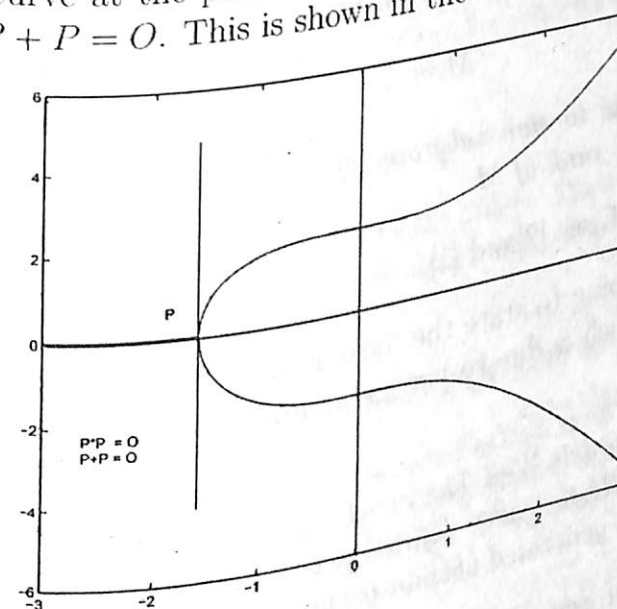


3.1 Special cases

To find $P + P$, we have to draw a tangent at the point P . This tangent will intersect the curve at another point, say $P * P$. Then we will get $P + P$ as below



If P is on the x -axis, then the line passing through the point P will intersect the curve at the point O at infinity. Hence, $P * P = O$ and this implies that $P + P = O$. This is shown in the following diagram.



4 Mordell-Weil Theorem

We have mentioned about a binary operation on the set of rational points $E(\mathbb{Q})$ of an elliptic curve defined over $\mathbb{Q}$. It can be proved that $E(\mathbb{Q})$ is an abelian group under the above binary operation with O as an identity element. This group is called the Mordell-Weil group for the elliptic curve E and the group law is called the Mordell-Weil group law. Although it is a very difficult group to study, it belongs to a special class of groups, namely, finitely generated abelian groups which are discussed below.

Definition 2. Let M be an abelian group. Then the abelian group M is said to be finitely generated if there exist elements $m_1, m_2, \dots, m_k$ in M such that every $m \in M$ can be written as

$$m = n_1 m_1 + n_2 m_2 + \dots + n_k m_k$$

for some integers $n_1, n_2, \dots, n_k$.

Finitely generated abelian groups have a very nice structure. In an abelian group G , if $a, b \in G$ have finite order, then $a + b$ also has finite order. Again, a and $-a$ have always same order. Thus, if T is the set of all elements of G of finite order, then T is a subgroup of G . This subgroup is called the torsion subgroup of G . The following theorem gives the structure of finitely generated abelian groups.

Theorem 1 (Fundamental Theorem of Finitely Generated Abelian Group). If M is an abelian group which is finitely generated then

$$M = \mathbb{Z}^r \oplus T(M),$$

where $T(M)$ is the torsion subgroup of M and r is a non-negative integer which is called the rank of M .

Proof. For a proof, see [2] and [1]. $\square$

We are now going to state the most important theorem in the theory of elliptic curves which is due to two great mathematicians L. J. Mordell and A. Weil.

Theorem 2. (Mordell-Weil Theorem)

Let E be an elliptic curve defined over $\mathbb{Q}$. Then the Mordell-Weil group $E(\mathbb{Q})$ is a finitely generated abelian group.

Proof. For a proof, see [4]. $\square$

To prove the Mordell-Weil theorem, one needs to prove the following theorem which is known as Weak Mordell-Weil Theorem. We need this theorem to discuss a method for finding rank of elliptic curve. The proof of this theorem can be found in [4].

Theorem 3. (Weak Mordell-Weil Theorem)

Let E be an elliptic curve defined over $\mathbb{Q}$. Then the subgroup $2E(\mathbb{Q})$ has finite index in $E(\mathbb{Q})$. That is, the quotient group $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite.

Definition 3. Let E be an elliptic curve defined over $\mathbb{Q}$. By Mordell-Weil theorem,

$$E(\mathbb{Q}) = \mathbb{Z}^r \oplus E_{\text{tors}}(\mathbb{Q}),$$

where r is a non-negative integer. The non-negative integer r is called the rank of the elliptic curve E over $\mathbb{Q}$.

The Mordell-Weil group $E(\mathbb{Q})$ is finite if and only if its rank is zero.

Definition 4. Let E be an elliptic curve defined over $\mathbb{Q}$ and $P \in E(\mathbb{Q})$. Then P is said to have order m if it has order m under the Mordell-Weil group law. That is, if $P = O$ then P has order 1 and otherwise

$$mP = P + \dots + P(m \text{ times}) = O,$$

but $nP \neq O$ for all integers $1 \leq n < m$. If such an m exists, then P has finite order; otherwise it has an infinite order.

Theorem 4 (Nagell-Lutz Theorem). Let $y^2 = f(x) = x^3 + ax^2 + bx + c$ be an elliptic curve with integer coefficients a, b, c and let D be the discriminant of the cubic polynomial $f(x)$,

$$D = -4a^3c + a^2b^2 + 18abc - 4b^3 - 27c^2,$$

Let $P = (x, y)$ be a rational point of finite order. Then x and y are integers, and either $y = 0$, in which case P has order two, or else y divides D .

5 A Method for finding rank of Elliptic Curve

We are now going to discuss a method for finding rank of elliptic curve with a rational point of order 2. We already know that if a point has order 2, then its y -co-ordinate is 0 (due to Nagell-Lutz Theorem). Thus, if the elliptic curve is given by the following normal form:

$$E_1 : y^2 = f(x) = x^3 + a_1x^2 + b_1x + c_1,$$

where a_1, b_1, c_1 are rational numbers, has rational point $T = (x_0, y_0)$ of order 2, then $y_0 = f(x_0) = 0$. Note that the discriminant of $f(x)$ is denoted as D and is defined as

$$D = -4a_1^3c_1 + a_1^2b_1^2 + 18a_1b_1c_1 - 4b_1^3 - 27c_1^2$$

In case if $a_1 = 0$ then $D = -4b_1^3 - 27c_1^2$ and if $c_1 = 0$, then $D = b_1^2(a_1^2 - 4b_1)$. Making a change of coordinates, we can move the point $(x_0, 0)$ to the origin.

This obviously does not effect the group $E_1(\mathbb{Q})$. The new equation again has integer coefficients: in the new coordinates the curve will have the form

$$E: y^2 = x^3 + ax^2 + bx$$

with $a, b \in \mathbb{Z}$. Then $T = (0, 0)$ is a rational point on E and satisfies $2T = O$.

Theorem 5. If $P = (x, y)$ is a rational point on $y^2 = f(x) = x^3 + ax^2 + bx + c$ then x, y have the form

$$x = \frac{m}{e^2}, \quad y = \frac{n}{e^3}$$

for some integers m, n, e with $e > 0$ and $\gcd(m, e) = \gcd(n, e) = 1$.

Proof. Since x and y are rational, so we can write $x = \frac{m}{M}$ and $y = \frac{n}{N}$ in lowest terms with $M > 0$ and $N > 0$. Substituting these values of x and y in the equation of the curve, we find that

$$\frac{n^2}{N^2} = \frac{m^3}{M^3} + a \frac{m^2}{M^2} + b \frac{m}{M} + c,$$

and then clearing denominators yields

$$M^3 n^2 = N^2 m^3 + a N^2 M m^2 + b N^2 M^2 m + c N^2 M^3 \quad (4)$$

Here we see that $N^2 \mid M^3 n^2$. But $\gcd(n, N) = 1$, so $N^2 \mid M^3$. Now we will prove that $M^3 \mid N^2$.

From equation (4) we immediately see that $M \mid N^2 m^3$. Also, since $\gcd(m, M) = 1$, we find that $M \mid N^2$. Using this fact back in (4), we get $M^2 \mid N^2 m^3$, so $M \mid N$. Finally, from the equation (4), we see that $M^3 \mid N^2 m^3$, so $M^3 \mid N^2$. Thus we have shown that $N^2 \mid M^3$ and $M^3 \mid N^2$, so $M^3 = N^2$. Also we have seen that $M \mid N$. So, if we let $e = \frac{N}{M}$, then we find that

$$e^2 = \frac{N^2}{M^2} = \frac{M^3}{M^2} = M \quad \text{and} \quad e^3 = \frac{N^3}{M^3} = \frac{N^3}{N^2} = N$$

Therefore x and y takes the desired form: $x = \frac{m}{e^2}$ and $y = \frac{n}{e^3}$. $\square$

Let E and E_1 be two elliptic curves given by the equation

$$E: y^2 = x^3 + ax^2 + bx \quad (5)$$

$$\bar{E}: y^2 = x^3 + \bar{a}x^2 + \bar{b}x \quad (6)$$

where $\bar{a} = -2a$ and $\bar{b} = a^2 - 4b$. Let $T = (0, 0)$ which is a rational point of order 2 on E . The following results are very crucial to understand the method for finding rank. The proof can be found in [4].

Theorem 6. There is a homomorphism $\Phi: E \rightarrow \bar{E}$ defined by

$$\Phi(P) = \begin{cases} \left(\frac{y^2}{x^2}, \frac{y(x^2-b)}{x^2} \right), & \text{if } P = (x, y) \neq O, T; \\ \bar{O}, & \text{if } P = O \text{ or } P = T. \end{cases} \quad (7)$$

The kernel of Φ is $\{O, T\}$.

Theorem 7. There is a homomorphism $\Psi: \bar{E} \rightarrow E$ defined by

$$\Psi(\bar{P}) = \begin{cases} \left(\frac{\bar{y}^2}{4\bar{x}^2}, \frac{\bar{y}(\bar{x}^2-\bar{b})}{8\bar{x}^2} \right), & \text{if } \bar{P} = (\bar{x}, \bar{y}) \neq \bar{O}, \bar{T}; \\ O, & \text{if } \bar{P} = \bar{O} \text{ or } \bar{P} = \bar{T}. \end{cases} \quad (8)$$

The kernel of Ψ is $\{\bar{O}, \bar{T}\}$.

The composition map $\Psi \circ \Phi: E \rightarrow E$ is the multiplication by 2. That is

$$(\Psi \circ \Phi)(P) = 2P, \quad \forall P \in E$$

We will now proceed to determine the image of Ψ .

Let $\mathbb{Q}^*$ be the multiplicative group of non-zero rational numbers, and let $\mathbb{Q}^{*2}$ denote the subgroup of squares of elements of $\mathbb{Q}^*$, i.e.,

$$\mathbb{Q}^{*2} = \{u^2 : u \in \mathbb{Q}^*\}$$

We introduce a map $\alpha: E(\mathbb{Q}) \rightarrow \frac{\mathbb{Q}^*}{\mathbb{Q}^{*2}}$ defined by

$$\begin{aligned} \alpha(O) &= 1 \pmod{\mathbb{Q}^{*2}}; \\ \alpha(T) &= b \pmod{\mathbb{Q}^{*2}}; \\ \alpha(x, y) &= x \pmod{\mathbb{Q}^{*2}} \text{ if } x \neq 0. \end{aligned}$$

Theorem 8. We have

(a) The mapping $\alpha: E(\mathbb{Q}) \rightarrow \frac{\mathbb{Q}^*}{\mathbb{Q}^{*2}}$ described above is a homomorphism.

(b) The kernel of α is the image $\Psi(\bar{E}(\mathbb{Q}))$. Hence α induces a one to one homomorphism

$$\frac{E(\mathbb{Q})}{\Psi(\bar{E}(\mathbb{Q}))} \rightarrow \frac{\mathbb{Q}^*}{\mathbb{Q}^{*2}}$$

(c) Let $p_1, p_2, \dots, p_t$ be the distinct primes dividing b . Then the image of α is contained in the subgroup of $\frac{\mathbb{Q}^*}{\mathbb{Q}^{*2}}$ consisting of the elements

$$\{\pm p_1^{\epsilon_1} p_2^{\epsilon_2} \dots p_t^{\epsilon_t} : \text{each } \epsilon_i \text{ equals } 0 \text{ or } 1\}$$

(d) The index $E(\mathbb{Q}) : \Psi(\bar{E}(\mathbb{Q}))$ is at most 2^{t+1} .

Hence,

$$(E(\mathbb{Q}) : \Psi(\overline{E}(\mathbb{Q}))) = \# \alpha(E(\mathbb{Q}))$$

Observation II: Using the analogous homomorphism $\bar{\alpha} : \overline{E}(\mathbb{Q}) \rightarrow \frac{\mathbb{Q}^*}{\mathbb{Q}^{*2}}$, we find that

$$(\overline{E}(\mathbb{Q}) : \Phi(E(\mathbb{Q}))) = \# \bar{\alpha}(\overline{E}(\mathbb{Q}))$$

This gives the following attractive formula for the rank of $E(\mathbb{Q})$:

$$2^r = \frac{\# \alpha(E(\mathbb{Q})) \# \bar{\alpha}(\overline{E}(\mathbb{Q}))}{4}$$

Using the above formula we can compute the rank of an elliptic curve. $\square$

References

- [1] D.S Dummit and R. M. Foote, Abstract Algebra, John Wiley and sons, Wiley Student Edition, 2006.
- [2] S. Lang, Algebra, Springer-Verlag, Revised Third Edition, 2008.
- [3] K. Rubin and A. Silverberg, Ranks of Elliptic curves, Bulletin of the American Mathematical Society, Vol. 39, No. 4, 455-474, 2002.
- [4] Joseph H. Silverman and John Tate, Rational Points on Elliptic Curves, UTM, Springer-Verlag, New York, 1992.
- [5] Derek F. Lawden, Elliptic Functions and Applications, Springer-Verlag, New York.
- [6] Joseph H. Silverman, Advanced Topics in the Arithmetic of Elliptic Curves, Springer-Verlag, New York.
- [7] Ranks of Elliptic Curves and Random Matrix Theory, London Mathematical Society Lecture Note Series: 341.

Investigation of Phytochemicals in Some Common Vegetables Sold in the Market of Tinsukia Town

P. Saikia

P. Prasad

Department of Chemistry
Tinsukia College, Tinsukia, Assam

Introduction

Vegetables are the major source of a number of nutrients. They are rich sources of minerals, fibres, vitamins, carbohydrates, proteins and amino acids, both essential and non-essential, which provides important nutrient to human health. Thousand of phytochemical compounds, with several different biological functions, have been identified in plant-based foods. Plant also produces a wide range of redox-active secondary metabolites such as ascorbic acids, carotenoids, polyphenols and enzymes with antioxidant activity, which protects the cell from oxidative damages^{1,2}. Therefore epidemiological studies have consistently shown that there is a significant positive correlation between the consumption of fruits and vegetables and reduction in the risk of heart diseases mortalities, common cancers, and other degenerative diseases as well as aging^{3,4}.

What are Phytochemicals ?

Phytochemicals are the substances found in plant foods that contain protective compounds that may help prevent diseases⁵. Phyto comes from the Greek word meaning plant. Phytochemicals are plant chemicals. Phytochemicals are certain organic components of plants which scientists have isolated as being beneficial to human health in a different way. Phytochemicals, also known as phytonutrients, are naturally occurring protective chemicals that are found in foods of plant origin. Numbers of phytochemicals have been identified; many more are being discovered. Phytochemicals are not like vitamins or minerals, so there is no recommended daily allowance (RDA) for them calories, so they don't provide energy. They

P. Saikia, P. Prasad "Investigation of Phytochemicals in Some Common Vegetables Sold in the Market of Tinsukia Town", Horizon (Ed. Book), 271-280(2023), copyright: Tinsukia College Publishing, ISBN: 978-81-19322-07-7

have subtle, drug-like effects which can influence your body's chemistry. Some may help inhibit the growth of many diseases like cancers or some degenerative diseases. Phytochemicals-rich food is generally used to treat or cure rheumatoid arthritis, but it is desirable to know which foods are high in content. There are numbers of phytochemicals which act as antioxidants and therefore that protects cells from cancer and cardiovascular disease, as well as from urinary tract infections, rheumatoid arthritis, etc. Scientific study has shown that people who consume a diet rich in fruits and vegetables, and therefore in phytochemicals, have a lower incidence of many disorders, including cardiovascular disease, diabetes, and certain types of cancer. Additionally, phytochemicals may enhance immune response and cell-to-cell communication, allowing for the body's built-in defenses to work more efficiently. Phytochemicals may even alter estrogen metabolism, cause cancer cells to die (apoptosis), repair DNA damage caused by smoking and other toxic exposure, and detoxify carcinogen. Wild vegetables and traditional medicinal plants contain a number of organic photochemical that have been linked to the promotion of good health.

There are thousands of phytochemicals found in foods of plant origin. Some of common phytochemicals are listed below-

Table1: Different types of phytochemicals present in vegetables and fruits⁶

Types	Subtypes	Examples
Terpenoids	Carotenoids	α - carotenes, β -carotenes, xanthophylls, etc.
	Triterpenoids	Saponins, oleanolic acid, etc.
	Monoterpenes	Limonene, perillyl alcohol, etc.
	Steroids	Sitosterol, etc.
Polyphenolic compounds	Flavonoids	flavones, flavonols, isoflavones, flavanones, etc.
	Phenolic acids	Caffeic acid, Vanillic acid, Ferulic acid, Rosmarinic acid, etc.

Phenolic compounds are secondary plant metabolites that are thought to be an integral part of human diet and naturally present in almost all plant including food products of plant origin. The majority of natural antioxidants are phenolic compounds, and the most important group of natural antioxidants are the tocopherols, flavonoids and phenolic acids. Flavonoids are a large group of naturally occurring plant phenolic compounds including flavones, flavonols, isoflavones, flavanones and chalcones which occur in all type of higher plant tissues. Flavones and flavonols are found in almost every plant, particularly in leaves and petals, with flavonols occurring more frequently than flavones. More than 8000 polyphenolics, including 4000 flavonoids, have been identified in different plant species and the number is still growing⁷.

How do phytochemicals work

Phytochemical compounds have different biological functions. Some possible actions of Phytochemical are⁸:

Antioxidant Activity: Antioxidants are the compounds that slow or inhibit the oxidation of lipids or other molecules by inhibiting the initiation and propagation of oxidizing chain reactions. Most phytochemicals have antioxidant activity and protect our cells against oxidative damage and reduce the risk of different chronic and degenerative diseases such as cancers, heart diseases, liver disease, inflammation, diabetes, atherosclerosis, Alzheimer's and Parkinson's disease⁷.

Hormonal action: Some specific phytochemical regulate hormonal action. For example, isoflavones, found in soy, imitate human estrogens and help to reduce menopausal symptoms and osteoporosis.

Stimulation of enzymes: Phytochemical are also involved to stimulate enzyme activity. For example, indoles, which are found in cabbages, stimulate enzymes that make the estrogen less effective and could reduce the risk for breast cancer.

Interference with DNA replication: Some phytochemical have specific role in the process of DNA replication. Saponins found in beans interfere with the replication of cell DNA, thereby preventing the multiplication of cancer cells. Capsaicin, found in hot peppers, protects DNA from carcinogens.

Anti-bacterial effect: Some phytochemical have also anti-bacterial effect. For example allacin from garlic has anti-bacterial properties.

Physical action: Some phytochemicals bind physically to cell walls thereby preventing the adhesion of pathogens to human cell walls. Proanthocyanidins are responsible for the anti-adhesion properties of cranberry. Consumption of cranberries will reduce the risk of urinary tract infections and will improve dental health.

Rich sources of phytochemicals

Foods containing phytochemicals are already part of our daily diet. In fact, most foods contain phytochemicals except for some refined foods such as sugar or alcohol. Plant based foods, such as whole grains, vegetables, beans, fruits and herbs contain many phytochemicals. The fruits blueberries, cranberries, cherries, apple, etc. and vegetables cauliflower, cabbage, carrots, broccoli, green leafy vegetables are the major source of phytochemicals of our daily requirement. Also some refined foods such as sugar or red wine also contains phytochemicals. It is recommended take daily at least 5 to 9 servings of fruits or vegetable. Fruits and vegetables are also rich in minerals, vitamins and fibre and low in saturated fat.

This study aimed as to examine presence of some phytochemical in some common vegetables.

Materials and Methods

Vegetable analysed

11 vegetables, five green leafy vegetables (GLV) and six different types of brinjal (eggplant) were analysed (table 2). All the vegetables were collected from weakly market of Tinsukia town, Assam. These vegetables are shown in Fig. 1 to Fig.11.

Table 1: List of examined vegetables

Sample label	Local name	Family	Scientific name	Part used
A	Bengena	Solanaceae	<i>Solanum melongena</i> var <i>esculentum</i>	Fruit part
B	Bengena	Solanaceae	<i>Solanum melongena</i> var <i>serpentium</i>	Fruit part
C	Bengena	Solanaceae	<i>Solanum melongena</i> var <i>depressum</i>	Fruit part
D	Bengena	Solanaceae	<i>Solanum melongena</i> var <i>esculentum</i>	Fruit part
E	Boga bengena	Solanaceae	<i>Solanum melongena</i> var <i>esculentum</i>	Fruit part
F	Bengena	Solanaceae	<i>Solanum melongena</i> var	Fruit part
G	Methi	Fabaceae	<i>Trigonella foenum-graecum</i>	Leaf, twigs
H	Tengasi	Oxalidaceae	<i>Oxalis corniculata</i> L	Leaf, whole plant
I	Kasu	Araceae	<i>Colocasia esculenta</i>	Leaf, stem
J	Jilmil (Bhatua)	Chenopodiaceae	<i>Chenopodium album</i>	Leaf, twigs
K	Khutura	Amaranthaceae	<i>Amaranthus viridis</i>	Whole plant

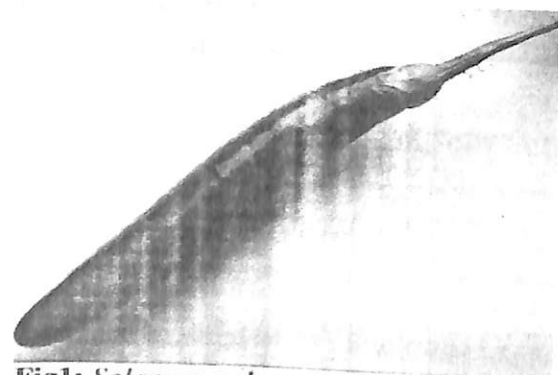


Fig1: *Solanum melongena* var *esculentum*

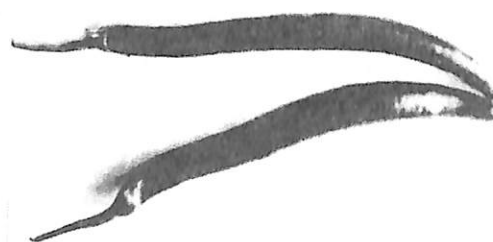


Fig2: *Solanum melongena* var *serpentium*



Fig3: *Solanum melongena* var *depressum*

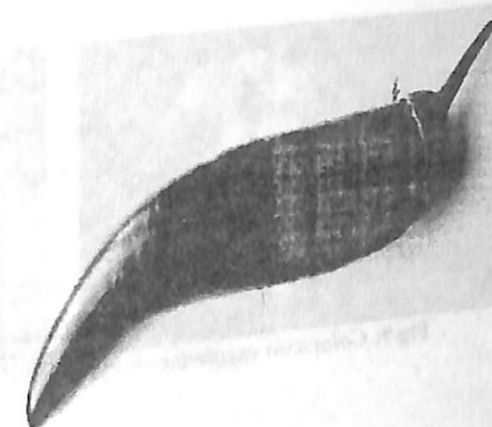


Fig4: *Solanum melongena* var *esculentum*

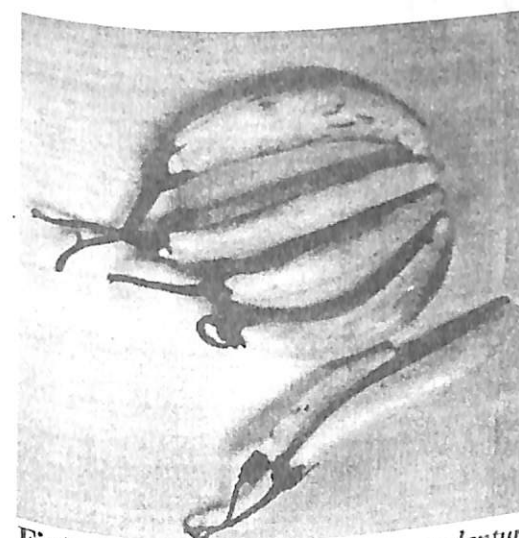


Fig5: *Solanum melongena* var *esculentum*

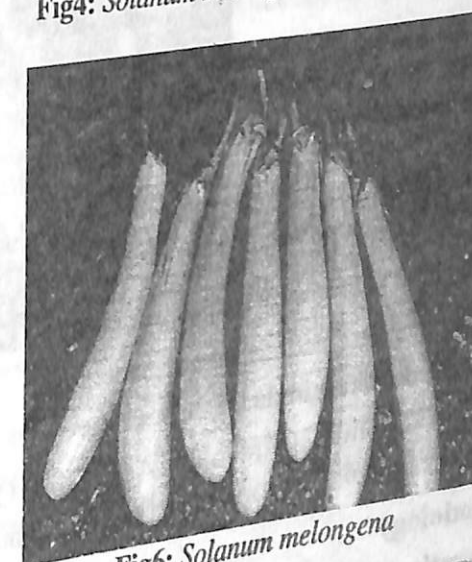


Fig6: *Solanum melongena*

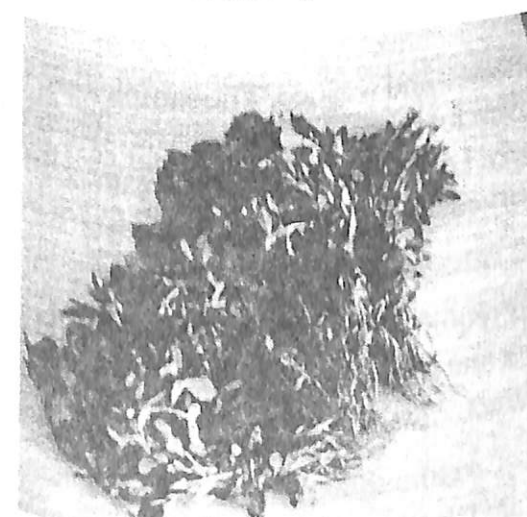


Fig7: *Trigonella foenum-graecum*



Fig 8: *Oxalis corniculata*

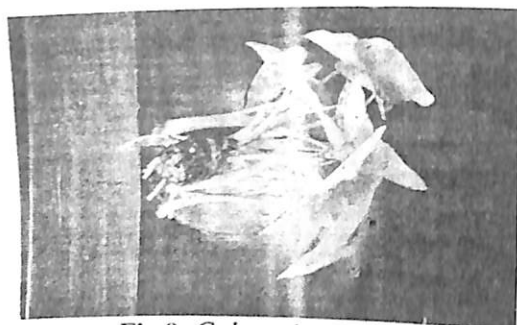


Fig 9: *Colocasia esculenta*



Fig 10: *Chenopodium album*

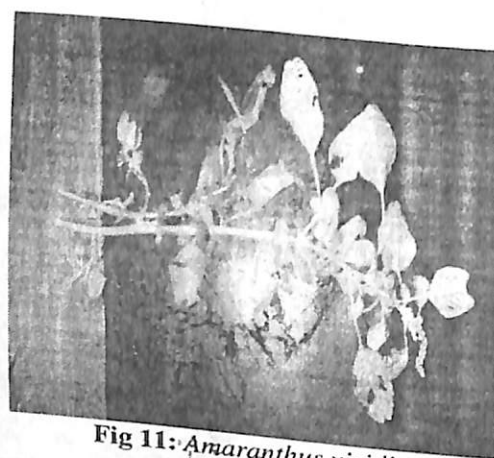


Fig 11: *Amaranthus viridis*

Reagents & Chemicals

Ferric Chloride, Hydrochloric Acid, Olive Oil, Ammonia Solution, Sulphuric Acid, Acetic Anhydride, Ethanol, Chloroform, Distilled Water, etc.

Methodology

Preparation of powdered samples

Initially the samples of different varieties are collected and washed. The samples are then cut into small pieces and weighed, then dried in an oven for constant weight. After drying the weight of the dried samples are again taken accurately. Lastly, the dried samples are grinded in a mixer grinder and powdered sample were preserved for further test.

Preparation of aqueous extract

About 1 g of each powdered samples (when required) is taken in a 100ml beaker containing distilled water and put on magnetic stirrer and stirred for 40 minutes. Filtered the mixture and the filtrate is considered as aqueous extract.

Preparation of ethanolic extract

About 1 g of each powdered samples (when required) is taken in a 100ml beaker containing ethanol (20 ml) and put on v and stirred for 30 minutes. Filtered and considered as ethanolic extract.

Estimation of Moisture content

Moisture content was determined by a gravimetric method by drying the samples to constant weight at 95-100 °C in an oven, cooled in a desiccator and weighted. The percentage loss in weight was expressed as percentage moisture content.

Test for tannins

About 0.5 g of the dried powdered samples was boiled in 20 ml of water in a test tube and then filtered. A few drops of 0.1% ferric chloride was added and observed for brownish green or a blue-black colouration indicates the presence of tannin.

Test for phlobatannins

Deposition of a red precipitate when an aqueous extract of each plant sample was boiled with 1% aqueous hydrochloric acid was taken as evidence for the presence of phlobatinins.

Test for saponins

About 2 g of the powdered sample was boiled in 20 ml of distilled water in a water bath and filtered. 10ml of the filtrate was mixed with 5 ml of distilled water and shaken vigorously for a stable persistent froth. The frothing was mixed with 3 drops of olive oil and shaken vigorously, then observed for the formation of emulsion.

Test for flavonoids

5 ml of dilute ammonia solution were added to a portion of the aqueous filtrate of each plant extract followed by addition of concentrated H_2SO_4 . A yellow coloration observed in each extract indicated the presence of flavonoids. The yellow coloration disappeared on standing.

Test for steroids

Two ml of acetic anhydride was added to 0.5 g ethanolic extract of each sample with 2 ml sulphuric acid. The colour changed from violet to blue or green in some samples indicating the presence of steroids.

Test for terpenoids (Salkowski test)

Five ml of each extract was mixed in 2 ml of chloroform, and concentrated H_2SO_4 (3 ml) was carefully added to form a layer. A reddish brown coloration of the inter face was formed to show positive results for the presence of terpenoid

Results and Discussion

Moisture content of the vegetables

Moisture content of the vegetables is shown in the table 2. The moisture content was calculated as the weight differences between fresh and dry sample.

$$\text{Moisture content (\%)} = \frac{A - B}{A} \times 100$$

A= weight of fresh sample, B= weight of dry sample

Table 2: Moisture content of the vegetables

Sample	Moisture content (%)
A	92.81 %
B	90.23 %
C	91.39 %
D	93.45 %
E	92.69 %
F	89.50 %
G	84.97 %
H	85.49 %
I	89.37 %
J	87.32 %
K	87.83 %

Moisture content of all the vegetables were shown in the **table 2**. Green leafy vegetables are believed to occupy a modest place as a source of trace elements due to their high water content. The moisture content of the vegetables ranged from 84.9% (*Trigonella foenum-graecum*, sample - G) to 93.45% (*Solanum melongena* var *esculentum*, sample-D).

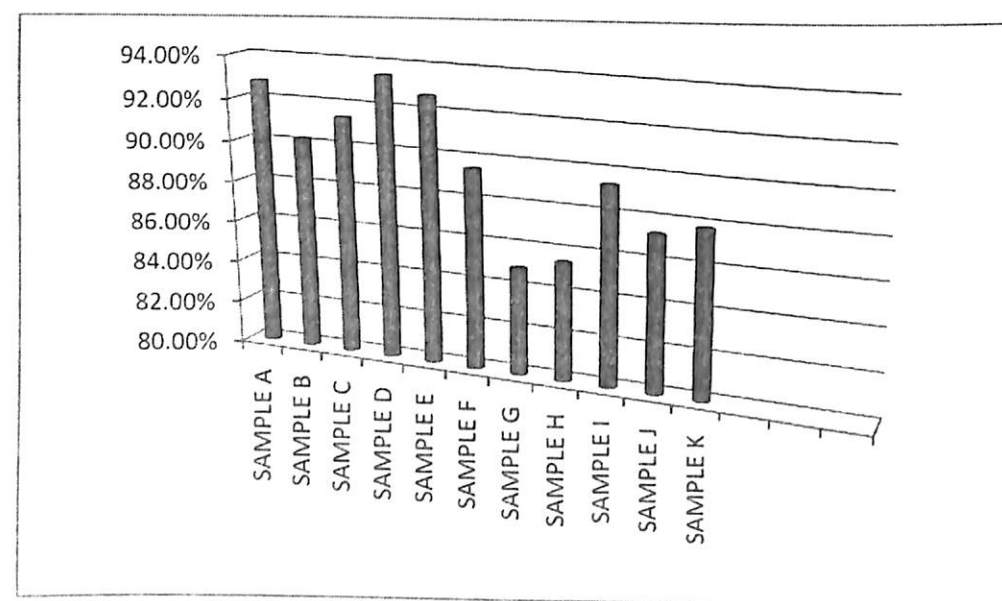


Figure 12: Comparison between moisture content of the vegetables

Phytochemicals in the vegetables

The present study carried out on the vegetables samples revealed the presence of medicinally active constituents. The phytochemical screenings of the 11 investigated vegetables are summarized in **Table 3**. Tannins and flavonoids are present in all the samples. Phlobatannins is present in most of the sample but absent in samples G, H, J and K. Steroid is found only in the sample K. Terpenoid is present in the sample C and F whereas saponis is present only in sample H and J.

Table 3: Phytochemicals identified in the vegetables

Samples	Tannins	Phlobatannins	Saponins	Flavonoids	Steroids	Terpenoids
A	+	+	-	+	-	-
B	+	+	-	+	-	-
C	+	+	-	+	-	+
D	+	+	-	+	-	-
E	+	+	-	+	-	-
F	+	+	-	+	-	+
G	+	-	-	+	-	-
H	+	-	+	+	-	-
I	+	+	-	+	-	-
J	+	-	+	+	-	-
K	+	-	-	+	+	-

+ Present; - Absent

Tannins, flavonoids, phlobatannins, steroid, alkaloid, saponin etc are secondary metabolites which showed some important biological activities. Tannins and flavonoids are good natural antioxidants and the presence of these phytochemicals proves that they have strong antioxidant activity. Flavonoids are a large group of naturally occurring plant phenolic compounds including flavones, flavonols, isoflavones, flavanones and chalcones which occur in all type of higher plant tissues⁹. Flavonoids and other phenolics have been suggested to play a protective role in the development of cardiovascular disease and cancer. Tannins also have antioxidant activity as well as antibacterial activity. Food containing tannin are beneficial to lower total cholesterol and stimulate the immunity. Terpenoid exhibit some pharmacological effects like anti-inflammatory

effect, anticancer, antibacterial activities. Phlobatannins also have antioxidant effect as well as anti-inflammatory properties. Steroidal compounds are of importance and interest due to their pharmacological effects.

Conclusion

Common vegetables containing phytochemicals can be very useful to our body systems. These are non-nutritive but very effective to our immune system. Foods containing phytochemicals are already part of our daily diet. Phytochemicals are naturally present in many foods but it is expected that through bioengineering new plants will be developed, which will contain higher levels. This would make it easier to incorporate enough phytochemicals with our food. Present study revealed that all vegetable contains important phytochemicals tannins and flavonoids. Most of them also contain saponin, phlobatannin. Therefore consumption of these vegetables are beneficial for prevent many diseases. However more detail study is required to establish the medicinal property of these vegetables.

References

1. Stengaland T., Remberg S. F. & Lye K. E. (2009). Total antioxidant activity in 35 Ugandan fruits and vegetables. *Food Chemistry*, 113, 85-91.
2. Hakkim F. L., Shankar C. G. & Girija S. (2007). Chemical composition and antioxidant property of Holy Basil (*Ocimum sanctum* L) leaves, stems and inflorescence and their in vitro callus cultures. *Journal of Agricultural and Food Chemistry*, 55, 9109-9117.
3. Oliveira A. P., Pereira J. A., Endrade P. B., Valentao P., Seabra R. M. & Silva B. M. (2007). Phenolic profile of *Cydonia oblonga* Miller Leaves. *Journal of Agricultural and Food Chemistry*, 55, 7926-7930.
4. Fattouch S., Caboni P., Coroneo V., Tuberoso C. I. G., Angioni A., Dessi S., Marzouki N. & Cabras P. (2007). Antimicrobial activity of Tunisian quince (*Cydonia oblonga* Miller) pulp and peel polyphenolic extract. *Journal of Agricultural and Food Chemistry*, 55, 963-969.
5. UC Cooperative Extension Center for Health and Nutrition Research, Department of Nutrition *University of California, Davis, CA 95616-8669, October 2008*.
6. www.phytochemicals.info
7. Gülçin Y. (2012). Antioxidant activity of food constituents: an overview. *Archive of Toxicology*, 86, 345-391.
8. Edeoga H.O.; Okwu D. E.; and Mbaebie O.B. (2005). Phytochemical constituents of some Nigerian medicinal plant. *African Journal of Biotechnology*, 4 (7), 685-688
9. White P. J. & Xing Y. (1997). Antioxidant from cereals and langmues. In: Shahidi F (ed) Natural antioxidants, chemistry, health effects, applications. AOAC Press, Champaign, pp 25-63.

NB: This study was a project work done in the department of chemistry. Tinsukia College for the partial fulfilment of B.Sc degree under Dibrugarh University.





Tinsukia College Publication
<https://tinsukiacollege.in>

ISBN 978-81-19322-07-7



9 788119 322077 >