ISBN: 978-93-88901-74-1

RESEARCH TRENDS IN SCIENCE AND TECHNOLOGY VOLUME III

EDITORS

DR. ALOK RANJAN SAHU

MR. BASANT DESHWAL

DR. NARAYAN GAONKAR

DR. SANJAY SINGH

Bhumi Publishing, India

First Edition: July 2023

Research Trends in Science and Technology Volume III

(ISBN: 978-93-88901-74-1)

Editors

Dr. Alok Ranjan Sahu

Department of Botany,

Vikash Degree College,

Bargarh, Odisha

Mr. Basant Deshwal

Division of Nematology,

Indian Agricultural Research Institute

(IARI), New Delhi

Dr. Narayan Gaonkar

Department of Physics,

University College of Science,

Tumkur University, Tumkur, Karnataka

Dr. Sanjay Singh

Department of Physics,

Chintamani College of Arts and Science,

Gondpipri, Maharashtra

Bhumi Publishing

July 2023

First Edition: July, 2023

ISBN: 978-93-88901-74-1

Published by:

Bhumi Publishing,

Nigave Khalasa, Kolhapur 416207, Maharashtra, India

Website: www.bhumipublishing.com

E-mail: bhumipublishing@gmail.com

Book Available online at:

https://www.bhumipublishing.com/book/

PREFACE

In the ever-evolving landscape of science and technology, knowledge is a beacon that illuminates the path to progress. The pursuit of understanding and innovation has been the driving force behind the remarkable advancements that have shaped the world we live in today. As we embark on a new era, it becomes increasingly crucial to navigate through the diverse and dynamic currents of research to discern the trends that will define our future.

"Research Trends in Science and Technology" represents a collective effort to explore and elucidate the cutting-edge developments that are shaping the fields of science and technology. This book is an assembly of insightful chapters contributed by leading experts, researchers, and visionaries, all of whom share a common passion for unraveling the mysteries of the universe and harnessing the power of technology for the betterment of humanity.

In this volume, we delve into a broad spectrum of disciplines, ranging from fundamental sciences such as physics, chemistry, and biology to the transformative fields of artificial intelligence, nanotechnology, and biotechnology. By curating a diverse selection of research trends, we aim to showcase the interdisciplinary nature of modern scientific inquiry and the interconnectedness of technological breakthroughs.

We believe that knowledge should be shared and disseminated freely, fostering a collaborative spirit that transcends geographical and disciplinary boundaries. As such, "Research Trends in Science and Technology" serves as a platform for disseminating the latest discoveries, ideas, and perspectives that shape the course of human progress.

We extend our heartfelt gratitude to all the contributors who have dedicated their expertise and passion to enrich this compilation. Their invaluable insights and visionary outlooks have made this endeavor possible.

We hope that this book will inspire readers, whether they are students, researchers, policymakers, or curious minds, to embrace the spirit of inquiry and embark on their own explorations. By staying attuned to the latest research trends and leveraging collective knowledge, we can collectively chart a course towards a more sustainable, equitable, and innovative future.

Editors

TABLE OF CONTENT

APPLICATIONS OF ARTIFICIAL INTELLIGENCE IN BIOTECHNOLOGY

Bindu Rajaguru*, Meenakshi Johri and Bhakti Gavhane

Department of Biotechnology,

Pillai College of Art Commerce & Science (Autonomous),

New Panvel, Navi Mumbai, Maharashtra

*Corresponding author E-mail: bindurajaguru@gmail.com

Abstract:

Biotechnology combines two different worlds: living organisms that are highly complex and can never be understood with technology, an artificial entity created by humans to bring convenience to their life, which has changed our lives even though a lot more is yet to happen. Artificial intelligence or AI refers to the ability of machines to mimic or simulate the intelligence of higher organisms. AI has the potential to solve global problems, such as food security, climate action, clean energy, responsible production, and consumption, and achieve sustainable development goals. It has paved the way into every possible field, including education, agriculture, healthcare, engineering, the pharmaceutical industry, astronomy, biology, gaming, automobiles, social media, and the list. Diverse applications of AI in various fields have been made feasible using machine learning and neural networks. Al has the potential to transform the field of biotechnology completely. orks. In this article, we discuss the influence of artificial intelligence on different fields of biotechnology. Al has numerous applications in the agricultural sector for the development of good quality crops, in the medical sector for drug discovery and development of precision medicine, determination of protein structure, and identification of novel drug targets, etc. Al has the potential to completely transform the field of biotechnology.

Keywords: Artificial intelligence, precision medicine, Machine learning, Neural network

Introduction:

Over the past 50 years, rapid advancements in information technology and biotechnology have occurred in parallel at an unmatched pace in any other industry. The extensive use of big data, produced by modern high-throughput instrumentation technologies and stored in thousands of public and private databases, is becoming increasingly important for the advancement of biotechnology. The field of biotechnology is concerned with both the discovery of new biological systems knowledge and the development of effective bioprocesses to apply this knowledge. This knowledge was then used to develop a new product. (Bhardwaj, Kishore, & Pandey, 2022) The development of new sensor technologies and big data methodologies has made it simpler to understand, predict, and improve bioprocesses. The fundamental concept involves creating bioprocess models that can be used to fine-tune the parameters that regulate these processes. A mathematical model has been developed and is widely used to assist in the control of reactors and manufacturing equipment in bioprocesses. Furthermore, artificial intelligence plays a crucial role in this case (Oliveira, 2019).

Artificial intelligence is a branch of computer science that focuses on building a smart machine that can perform tasks that typically require human intelligence. It imitates thought processes, learning abilities, and knowledge abilities, and finds a vast number of applications in various fields. Artificial intelligence is no less than magic, it has become more advance in the last 20 years and has tremendous applications from big data, medical research, and autonomous vehicle. Artificial intelligence has the power to improve quality of life. AI involves three basic concepts, namely deep learning, machine learning, and neural networks (Sarker, 2022).

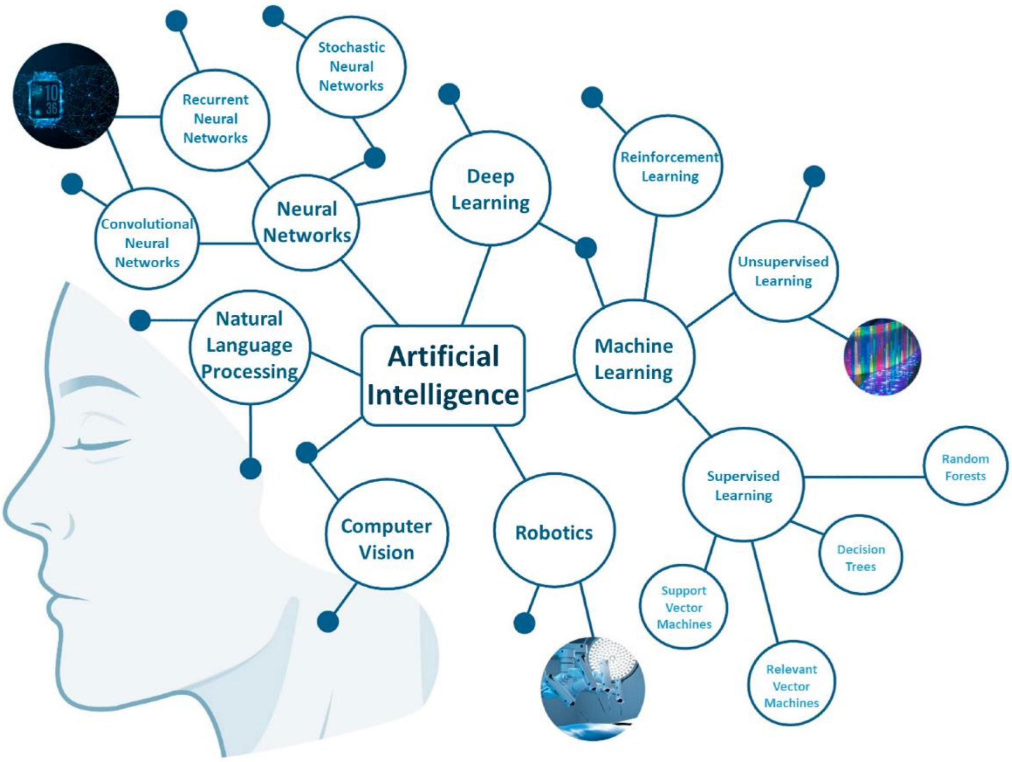


Figure 1: Fields of artificial intelligence (Athanasopoulou *et al.,* 2022)

In machine learning, the software allows a machine to learn tasks (Tran *et al.,* 2021). But this software cannot learn on its own, and they cannot think without their codes. While deep learning and neural networks work together, deep learning helps the machine to learn more than one task. Deep learning is possible because of the neural network, which is inspired by a biological neuron network. However, this neural network was built from codes. Millions of nodes help the machine learn new tasks (Oliveira, 2019).

Now, global society is faced with many challenges due to the increase in population, be it in the agriculture sector, healthcare sector, or education Sector, Meeting the demands of a growing population is a challenging task. But the advancement in technology and Research Fields can solve many problems AI is one such Technology (Holzinger *et al.,* 2023). In the medical field, a variety of biomarkers and genetic loci linked to various diseases and appropriate treatments are being identified by utilizing multi-omics and neuroimaging methods, thanks to the most recent advancements in artificial intelligence and machine learning techniques. Several studies have used artificial intelligence and machine-learning techniques to predict the diagnosis of specific diseases, including schizophrenia and Alzheimer's disease. The use of neuroimaging (such as structural and functional neuroimaging) in conjunction with artificial intelligence and machine learning methods is becoming increasingly popular for gaining new insights into diseases (Lin *et al.,* 2020).

To achieve an early/optimal diagnosis, monitoring, and intervention, it is essential to depict patient dynamics using artificial intelligence (AI). The essential components in this cutting-edge field are algorithms and predictive models. Artificial Intelligence (AI) is already used in biotechnology, for example, in various approaches such as agricultural development, drug discovery, Pharmacogenomics, and Medical Imaging. proteomics, genetic engineering, process development and optimization, quality control, bioinformatics, etc. (Holzinger *et al.,* 2023).

Artificial Intelligence in agriculture biotechnology

Agriculture occupies around 38% of the total land surface of the planet. It is a significant source of revenue for many countries, including India. It has been reported by the United Nations 'Food and Agricultural Organizations (FAO) that the global population will reach 9 billion by 2050. And this will put a strain on the agricultural system to meet the demands of a growing population. There are various challenges in the agricultural field like pest infestation, insufficient use of fertilizers or overuse of pesticides and insecticides, lack of irrigation system, improper storage, drought, etc., which are hard to overcome by the traditional or manual methods of agriculture (Bhardwaj *et al.,* 2022). Artificial intelligence (AI) has the potential to revolutionize the field of agriculture by enabling researchers and farmers to improve crop yield, reduce waste, get a good market value for their crops, and make help in making more informed decisions about farming practices. It can be beneficial for the rural economy as well as for the livelihood of villagers (Khan *et al.,* 2022).

Advancement in AI includes the use of drones for real-time monitoring and robots are used for the subsequent harvesting and processing of crops (Ranjha *et al.,* 2022). Many Biotech companies use machine learning and AI to design and train robots, capable of performing agricultural activities.

Enormous data from the crop field can be collected by drones, which are processed and evaluated by deep learning which can be further used for building a machine or robot. Data can be gathered by various technologies like image-based phenotyping platforms, unmanned aircraft system and satellite-based remote sensing, etc., and computer vision algorithm uses this information to advance our system understanding (Xu *et al.,* 2021). AI can be used in plant breeding by helping plant breeders to identify desirable genetic traits and develop new and efficient varieties of crops. The machine learning algorithms can be used to evaluate data on plant genetics, DNA sequences, gene expression, etc. which will help in the production of new varieties of crops (Khan *et al.,* 2022).

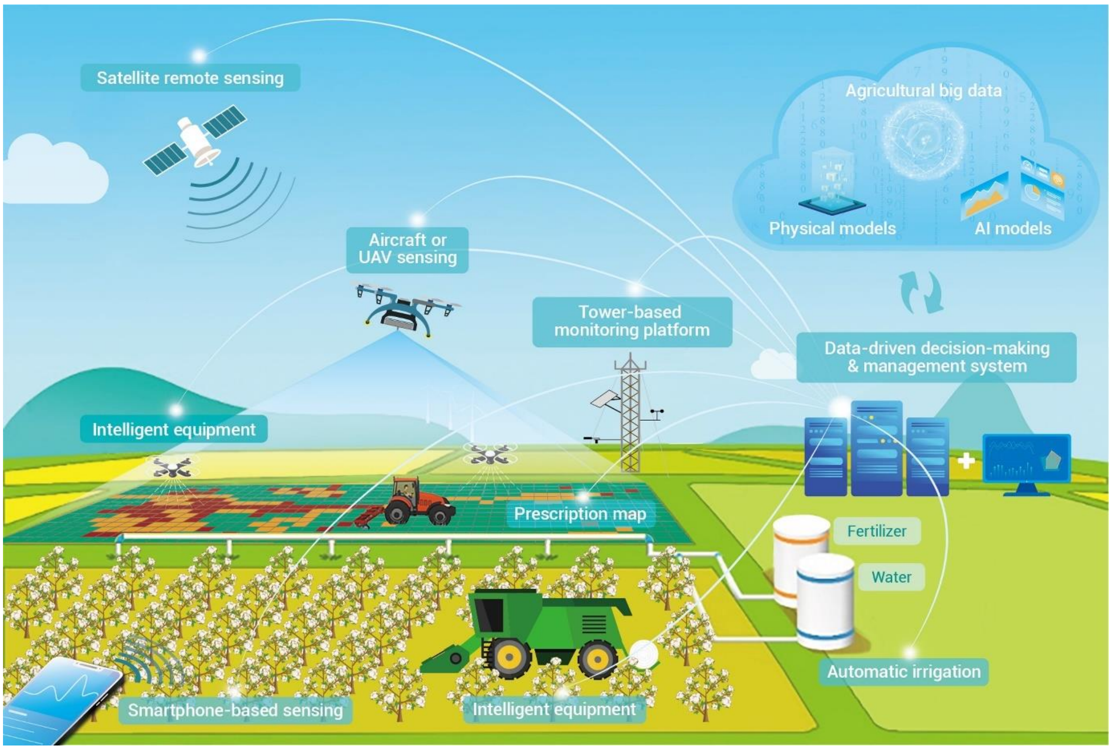


Figure 2: Use of AI in the agriculture (Xu *et al.*, 2021)

AI can be used to monitor crop growth rates and detect changes in plant health over time. This can help farmers identify potential problems early and take corrective actions before crop yields are impacted. AI can make use of satellite imagery and other data sources to detect signs of disease outbreaks in crops. Machine learning algorithms can be trained to identify visual patterns associated with diseased plants and alert farmers to potential problems. AI can be used to analyze large amounts of data on soil composition, weather patterns, and crop growth rates to identify optimal planting times and fertilizer applications. This can help farmers optimize their crop yields and reduce waste by avoiding over-fertilization and under-fertilization. In this way, AI can a vital role in agricultural biotechnology by enabling farmers and researchers to make an informed decisions about farming practices, disease-resistant crop production, crop monitoring, etc. (Prasad, 2020).

Artificial Intelligence in medical biotechnology

In medical biotechnology, living cells are used for the production of drugs and antibiotics. Advancements in medical science and biotechnology have opened new avenues for the development of medications and antibiotics. The healthcare system is one of the most vital sectors of a country, and it has been evolving continuously with the new advancements in technologies and new findings in the research for the diagnosis and treatment of different diseases (Brogi & Calderone, 2021). Around 5-11 % of the total GDP of the country is spent on the healthcare system all around the world, India spends almost 2.2% of its total GDP on healthcare. DNA or genetic manipulation is done in the cells to obtain desired and beneficial traits. Artificial intelligence (AI) has a wide range of applications in medical biotechnology (Lancellotti *et al.,* 2021).

AI plays a pivotal role in the drug discovery process, it can also be used for the designing and testing of new drugs. Machine learning algorithms are used to analyze the gigantic amount of data that be used to predict which compounds are most likely to be effective in the treatment of specific diseases. It also helps in understanding complex issues like quantitative and predictive epidemiology, precision-based medicines, or host-pathogen relationship (El Karoui *et al.,* 2019) AI is used to analyze medical images like X-rays, CT -scans, or MRI scans., etc, AI algorithm helps in understanding the abnormalities in the images like tumors or fractures with more accuracy and speed than humans. Machine-based algorithms analyze the patient's data like medical records or lab results or image scans to predict the patient's health (Hamamoto, 2021). AI can help researchers to analyze genomic data and to identify the genetic markers or biomarkers associated with a particular disease, which may help personalized treatments or therapies. For example, AI has a program, which is based on ten cellular algorithm models, which can accurately distinguish between benign and malignant tumors (Carpenter & Huang, 2018; Gómez-González *et al.,* 2020).

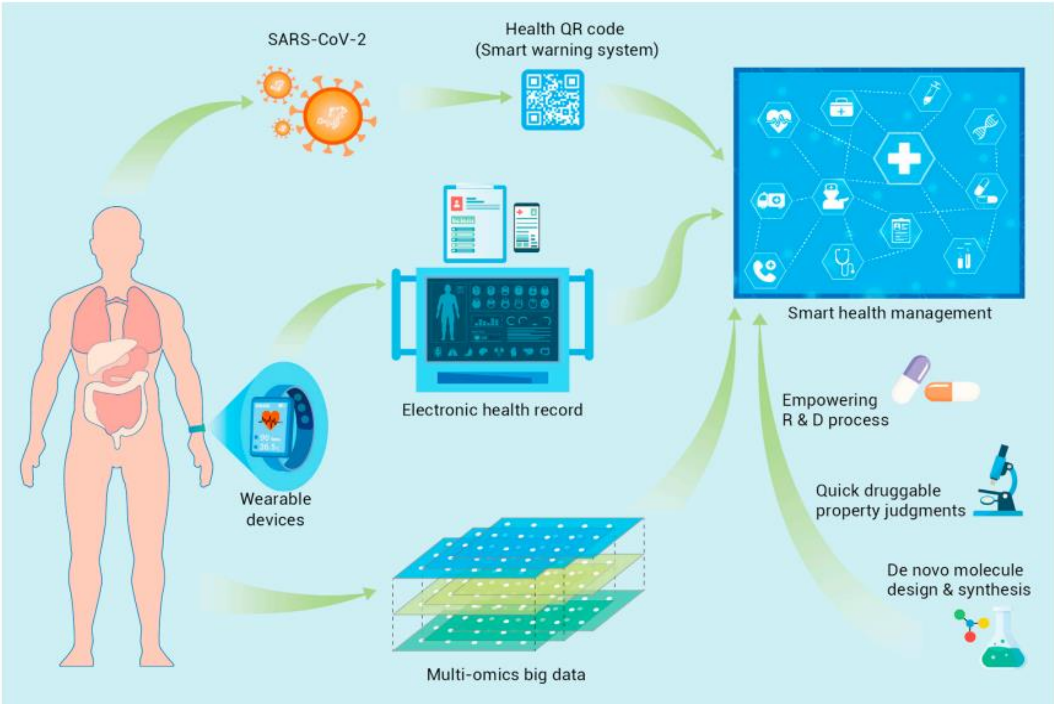


Figure 3: Application of AI in medical sciences (Xu *et al.,* 2021)

Personalized medicine is one of the novel applications of AI in the medical sciences. Precision medicine or personalized medicine takes into account individual differences in the genetic constitution, ecology, and lifestyle of patients, and how it can affect the person's body's response to drugs in a certain way (Hee Lee & Yoon, 2021). It's totally in contrast to the one umbrella approach where all patients are treated with the same drug (Lu, Goldstein, Angrist, & Cavalleri, 2014). AI can be used to make millions of predictions, just to identify the best therapeutic candidate molecule for the treatment of a particular disease. The purpose of personalized medicine is to maximize treatment benefits and minimize adverse effects. And it can also reduce the treatment cost, save time and improve patient care (Verma, 2012).

Table 1: AI in disease detection and prediction modeling (Bhardwaj *et al.*, 2022)

Disease studied	Algorithm	Findings
AMD	ML-based predictive model	The AI-based predictive model was able to predict the progression of AMD with high accuracy
Alzheimer's disease	RF, SHAP	AI-model was able to accurately detect and predict the progression of Alzheimer's disease with an accuracy of 93.95% in the first layer and 87.08% in the second layer
COVID-19	PA	An accuracy of 70–80% was achieved in predicting severe COVID-19 cases
Ovarian cancer	ANN	An accuracy of 93% was achieved in predicting the survival of ovarian cancer patients, and 77% accuracy was achieved in predicting the surgical outcome
pulmonary cancer	LCP-CNN, Brock model	LCP-CNN was able to predict the malignancy of pulmonary nodules with higher accuracy and lower false negative results than the Brock model
Influenza	IAT-BPNN	IAT-BPNN was able to predict influenza-like illness in large population size with a high accuracy

Artificial Intelligence in industrial biotechnology

Industrial biotechnology uses biotechnology-based tools to the traditional industrial processes for manufacturing bio-based products. It mainly uses living cells or microorganisms and their microbial products like enzymes. Industrial biotechnology is the blooming sector as it confluence consumer demand with attractive feedstock quantity, quality, and price. And with technological advances in the last few decades, it's been a boom in the market. AI has a vast number of applications in industrial biotechnology which involves the use of living organisms and bioprocesses for the production of chemicals, materials, and fuels (Holzinger *et al.,* 2023).

AI can be used for the optimization of bioprocess. It can analyze a large amount of data and machine learning can help in the recognition of and optimization of parameters like pH, temperature or nutrient concentration, etc. to increase the yield of products and reduce the cost of production. It can also be used for the quality control steps, as AI can be used to monitor the quality of bioprocess by evaluating the data from the sensors and other monitoring devices. Machine algorithms can detect fluctuations or deviations from normal operating conditions and predict when the maintenance of equipment or other interventions are needed to be done to maintain the quality of products or it can also predict when equipment failure is likely to occur. The machine learning algorithm can detect patterns in the data that indicate equipment degradation or impending failure, allowing for timely maintenance and repair, which can minimize downtime and reduce cost. AI can be used to engineer microbial strains for the production of desired chemicals or biomolecules. For this a large number are databases from proteomics, genomics, metabolomics, transcriptomics, etc for a respective strain is processed to develop an in silico model or simulation of the organism before actually proceeding to wet lab experiments. For example, the production of insulin from *E. coli* using recombinant DNA technology. where the gene for human insulin is inserted in the *E. coli* for the production of insulin.

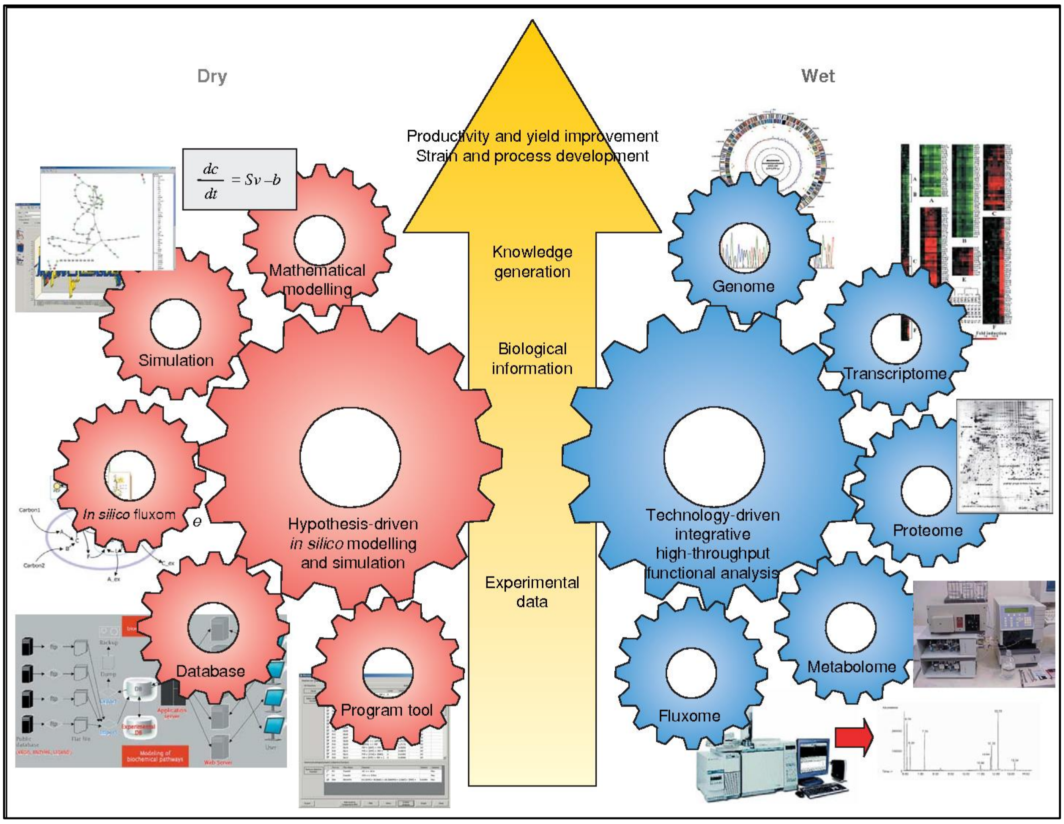


Figure 4: Process development and strain improvement (Lee *et al*, 2005)

AI or machine learning algorithms can use the genetic data available to search for the most promising strain for the given application and optimize its performance through genetic modifications. so, AI can be used to optimize the bioprocess, develop new products, or improve overall efficiency.

Artificial Intelligence in bioinformatics

Bioinformatics is a multidisciplinary subject that encompasses biology, computer science, information science, and statistics. The main applications of bioinformatics are DNA sequencing, modification, structural studies, and gene data analysis. Bioinformatics plays an important role in the research field, bioinformatics is used for designing in-silico experiments or models, and it is used in the drug discovery process. Bioinformatics play and vital role in multi-omics approaches, like genomics, metabolomics, proteomics, or transcriptomics. etc. advancement of precision medicine is because of the multi-omics, which had widen the knowledge (Zheng *et al.,* 2022).

Artificial intelligence has the potential to revolutionize the field of bioinformatics by enabling scientists or researchers to analyze a large amount of complex biological data more efficiently and accurately. DNA sequencing techniques are one of the most powerful contributions of machine learning in bioinformatics because sequencing the whole human genome took us almost a decade by traditional approaches can be one within a day now. Machine learning algorithms can be trained on large databases of known sequences to identify similarities and differences among sequences and to predict the function of unknown sequences.

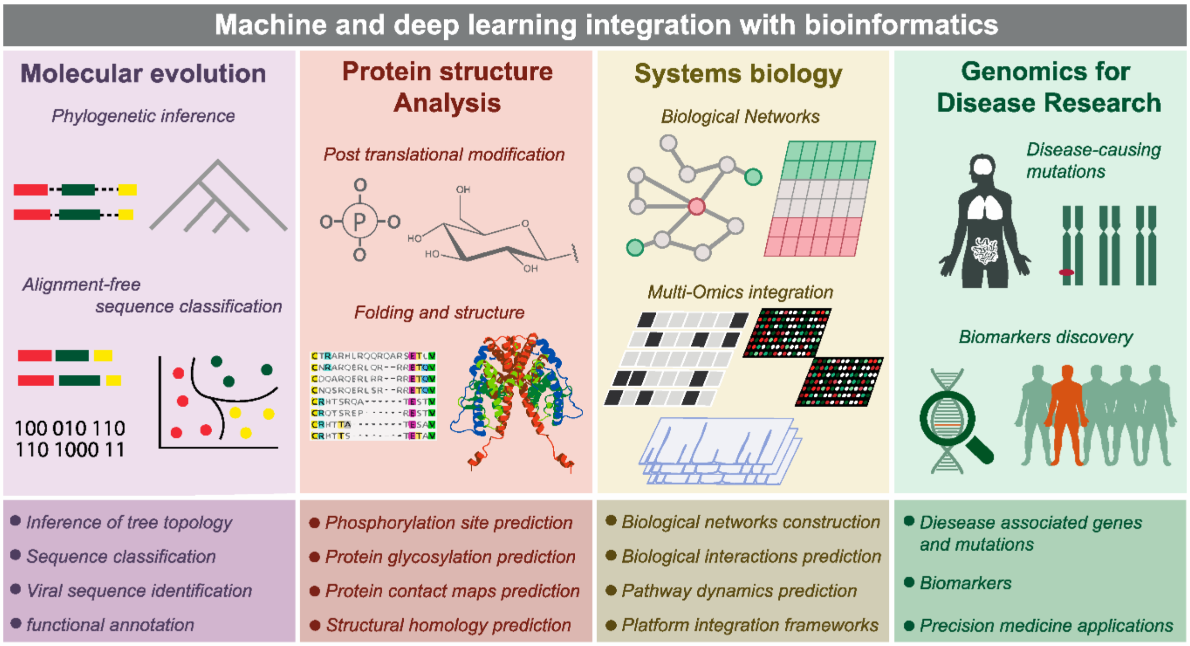


Figure 5: Artificial intelligence in bioinformatics (Auslander *et al.*, 2023)

DNA sequencing techniques are one of the most powerful contributions of machine learning in bioinformatics because sequencing the whole human genome took us almost a decade by traditional approaches can be one within a day now. Along with sequencing DNA, analyzing gene expression, finding new drug targets, defining protein functions, interpreting genetic variants, etc are some of the applications of AI in bioinformatics. For example, gene editing, where machine learning algorithms are used to search for the most feasible mutation, reduces the screening time for that mutation by around 95%.

The use of neural networks in the protein structure determination positions protein amino acid profile into three types- alpha helix, bets-sheet or random coils, etc., and its theoretical accuracy is around 84% (Zheng *et al.,* 2022). In conclusion, AI has the potential to revolutionize the field of bioinformatics by enabling more efficient sequence analysis, genomics, drug discovery, structural biology, systems biology, and precision medicine. As AI technology continues to advance, we will likely see even more innovative applications of this technology in bioinformatics (Chicco *et al.,* 2020).

Artificial Intelligence in animal biotechnology

The field of animal biotechnology has seen significant advancements over the years, and one area where artificial intelligence (AI) has the potential to make a significant impact is in improving animal breeding and genetic engineering. AI is the ability of machines to learn and make decisions based on data without being explicitly programmed, and this technology can be applied to animal biotechnology in several ways. One of the primary applications of AI in animal biotechnology is animal breeding. Traditional animal breeding involves selecting animals with desirable traits and breeding them to produce offspring with those traits. However, this process is quite expensive and requires a lot of time. AI can be used to analyze large amounts of data on animal genetics and select the best animals for breeding, resulting in more efficient and effective breeding programs. Another application of AI in animal biotechnology is in the genetic engineering of animals. Genetic engineering involves modifying an animal's DNA to produce desirable traits (Faber *et al.,* 2003).

AI can be used to analyze and interpret large amounts of genetic data, which can help researchers identify genes that are responsible for specific traits. This can lead to more targeted genetic engineering efforts, resulting in animals with specific desired traits.AI can also be used to develop more effective animal disease control strategies. For example, machine learning algorithms can be used to analyze data on animal health and identify patterns that could indicate the presence of a disease. This can help veterinarians and animal health professionals identify and respond to outbreaks more quickly, reducing the spread of disease and minimizing its impact on animal populations. The development of new animal-based products, such as medicines or food. For example, researchers could use AI to analyze the genetic makeup of different animals to identify compounds that could be used to develop new medicines (Faber *et al.,* 2003).

In conclusion, AI has the potential to make a significant impact on animal biotechnology. Its applications range from improving animal breeding and genetic engineering to developing more effective animal disease control strategies and identifying new uses for animal-based products. As AI technology continues to advance, we will likely see even more innovative applications of this technology in animal biotechnology in the future (Bhardwaj *et al.,* 2022).

Limitations and challenges

Reproducibility problems have plagued biotechnological and medical research in recent decades. It is generally known that the detailed documentation of research objects and their chains greatly improves the traceability, reproducibility, and reliability of research results, and many fields of life sciences and biotechnology are working to overcome this problem. The lack of sufficient quantities of data that are suitable, or of acceptable quality, is a major barrier to the creation of AI models. In many circumstances, the supply of biological material that is needed for the task at hand will determine the data's availability. The accessibility of software and the reproducibility of its runs are two other characteristics unique to the creation of AI models (Holzinger *et al.,* 2023).

To ensure that an AI solution is trustworthy, explainability is a crucial component in addition to the required reliability. AI systems are frequently used to analyze vast amounts of data and generate predictions or suggestions that could have a big impact on environmental safety, public health, and other crucial areas. People may find it challenging to have faith in the precision and dependability of an AI system's outputs if it is unable to clearly explain how it came to a given conclusion or prediction. Explainability enables people to comprehend and assess the decision-making processes of the AI system and ensures that it is being used in a trustworthy and ethical manner in the biotechnology domain (Holzinger *et al.,* 2023).

Conclusion:

According to historical accounts, humans continue to advance technology and adapt to new techniques. We could see that artificial intelligence had been evolving since its introduction in 1943. It has continued to expand quickly ever since as a new, cutting-edge technology. Machine-learning-based methods are more accurate, dependable, and robust than conventional methods. These techniques will play a considerably more significant role in the discovery of novel therapies owing to the extraordinary increase in the use of AI and ML in almost all areas of technology, science, and medicine. In the future, AI will continually discover wide-ranging applications in various unexplored domains. The amount of change it causes in people's lives is ultimately used to measure success. Given how quickly people embrace AI technologies today, the future appears bright.

References:

1. Auslander, N., Gussow, A. B., & Koonin, E. V. (2021). Incorporating machine learning into established bioinformatics frameworks. *International journal of molecular sciences*, *22*(6), 2903.
2. Athanasopoulou, K., Daneva, G. N., Adamopoulos, P. G., & Scorilas, A. (2022). Artificial intelligence: the milestone in modern biomedical research. *BioMedInformatics*, *2*(4), 727-744.
3. Bhardwaj, A., Kishore, S., & Pandey, D. K. (2022). Artificial intelligence in biological sciences. *Life*, *12*(9), 1430.
4. Brogi, S., & Calderone, V. (2021). Artificial intelligence in translational medicine. *International Journal of Translational Medicine*, *1*(3), 223-285.
5. Carpenter, K. A., & Huang, X. (2018). Machine learning-based virtual screening and its applications to Alzheimer's drug discovery: a review. *Current pharmaceutical design*, *24*(28), 3347-3358.
6. Chicco, D., Heider, D., & Facchiano, A. (2020). Artificial Intelligence Bioinformatics: Development and Application of Tools for Omics and Inter-Omics Studies. *Frontiers in Genetics*, *11*, 309.
7. El Karoui, M., Hoyos-Flight, M., & Fletcher, L. (2019). Future trends in synthetic biology—a report. *Frontiers in bioengineering and biotechnology*, 175.
8. Faber, D. C., Molina, J. A., Ohlrichs, C. L., Vander Zwaag, D. F., & Ferre, L. B. (2003). Commercialization of animal biotechnology. *Theriogenology*, *59*(1), 125-138.
9. Gómez-González, E., Gomez, E., Márquez-Rivas, J., Guerrero-Claro, M., Fernández-Lizaranzu, I., Relimpio-López, M. I.,& Capitán-Morales, L. (2020). Artificial intelligence in medicine and healthcare: a review and classification of current and near-future applications and their ethical and social Impact. *arXiv preprint arXiv:2001.09778*.
10. Hamamoto, R. (2021). Application of Artificial Intelligence for Medical Research. Biomolecules 2021, 11, 90.
11. Lee, D., & Yoon, S. N. (2021). Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges. *International Journal of Environmental Research and Public Health*, *18*(1), 271.
12. Holzinger, A., Keiblinger, K., Holub, P., Zatloukal, K., & Müller, H. (2023). AI for life: Trends in artificial intelligence for biotechnology. *New Biotechnology*, *74*, 16-24.
13. Khan, M. H. U., Wang, S., Wang, J., Ahmar, S., Saeed, S., Khan, S. U., ... & Feng, X. (2022). Applications of Artificial Intelligence in Climate-Resilient Smart-Crop Breeding. *International Journal of Molecular Sciences*, *23*(19), 11156.

14. Lancellotti, C., Cancian, P., Savevski, V., Kotha, S. R. R., Fraggetta, F., Graziano, P., & Di Tommaso, L. (2021). Artificial intelligence & tissue biomarkers: advantages, risks and perspectives for pathology. *Cells*, *10*(4), 787.
15. Lin, E., Lin, C. H., & Lane, H. Y. (2020). Precision psychiatry applications with pharmacogenomics: artificial intelligence and machine learning approaches. *International journal of molecular sciences*, *21*(3), 969.
16. Lu, Y. F., Goldstein, D. B., Angrist, M., & Cavalleri, G. (2014). Personalized medicine and human genetic diversity. *Cold Spring Harbor perspectives in medicine*, *4*(9).
17. Lee, S. Y., Lee, D. Y., & Kim, T. Y. (2005). Systems biotechnology for strain improvement. *Trends in biotechnology*, *23*(7), 349-358.
18. Oliveira, A. L. (2019). Biotechnology, big data and artificial intelligence. *Biotechnology journal*, *14*(8), 1800613.
19. Prasad, P. (2020). *Role of artificial intelligence in biotechnology*.
20. Ranjha, M. M. A. N., Shafique, B., Khalid, W., Nadeem, H. R., Mueen-ud-Din, G., & Khalid, M. Z. (2022). Applications of biotechnology in food and agriculture: a mini-review. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences*, *92*(1), 11-15.
21. Sarker, I. H. (2022). Ai-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems. *SN Computer Science*, *3*(2), 158.
22. Tran, K. A., Kondrashova, O., Bradley, A., Williams, E. D., Pearson, J. V., & Waddell, N. (2021). Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Medicine*, *13*(1), 1-17.
23. Verma, M. (2012). Personalized medicine and cancer. *Journal of personalized medicine*, *2*(1), 1-14.
24. Vaish, S., Agarwal, A., & Raheja, R. ROLE OF ARTIFICIAL INTELLIGENCE IN BIOTECHNOLOGY.
25. Wanerman, R. E., Javitt, G. H., & Shah, A. B. (2020). Artificial intelligence in biotechnology: A framework for commercialization. In *Biotechnology Entrepreneurship* (pp. 419-427). Academic Press.
26. Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., ... & Zhang, J. (2021). Artificial intelligence: A powerful paradigm for scientific research. *The Innovation*, *2*(4).
27. Zheng, P., Wang, S., Wang, X., & Zeng, X. (2022). Artificial Intelligence in Bioinformatics and Drug Repurposing: Methods and Applications. *Frontiers in Genetics*, *13*, 870795.

TRANSMISSION EMBEDDED COST ALLOCATION USING PROPORTIONAL NUCLEOLUS METHOD IN A LIBERALIZED ELECTRICITY MARKET

Murali Matcha*, Siddheswar Kar and Neha Verma

Department of Electrical Engineering,

Medi-Caps University, Indore, M.P, India

*Corresponding author E-mail: murali_233@yahoo.com

Abstract:

In deregulated electricity markets there is a strong need for effective allocation of transmission embedded costs to market participants. The conventional usage-based methods such as MW-Mile and Zero Counter Flow (ZCF) methods which are currently employed in market scenario may fail to send right economic signals. Hence in this paper, cooperative game theory based approaches are demonstrated. The existing game theory based approaches like Nucleolus and Shapley Value methods are found to be inefficient for transmission embedded cost allocation, due to their own pros and cons. Therefore Proportional Nucleolus (P-N) method which is also a cooperative game theory approach is proposed in this paper to overcome the drawbacks of aforementioned methods. All the methods presented in this paper are tested on IEEE 14 bus system, New England 39 bus system and Indian 75 bus system and a comparative study was carried out with the obtained results. Concept of Monotonicity is also explained on IEEE 14 bus system.

Keywords: Core, Coalition, Cooperative game, Imputation, Nucleolus, Payoff, Proportional Nucleolus, Shapley, Monotonicity

Introduction:

In the present-day power markets, the issue of allocating fixed costs to the market participants is of great significance. Fixed costs make up the largest part of transmission charges; hence there is a great demand for a fair and effective allocation of these costs to the market participants [1, 2, 3]. Different allocation schemes have been formulated in recent years based on the "natural economic use" of the transmission system [1, 10, 12, 17, 18]. In three variations of the MW-Mile method for pricing counter flows are investigated for the cost allocation method [19]. But it failed in providing incentives to users of the grid who causes counter flows. Therefore, all the conventional usage based methods like MW-Mile method, Zero counter flow (ZCF) method are advantageous from an engineering point of view, but they may fail to send right economic signals. The fixed cost allocation is a typical case where cooperation between the agents produces incentives and economies of scale [4]. These benefits can in turn be shared among the network participants. Thus the cooperative game theory (C.G.T) provides interesting concepts, methods and models that may be used when assessing the interaction of different agents in competitive markets and in the solution of conflicts that arise in that interaction, such as those of the electricity markets [12]. In particular, cooperative game theory arises as a most convenient tool to solve cost allocation problems [12]. The solution methodologies of cooperative game theory behave well in terms of fairness, efficiency, stability, and qualities required for the correct allocation of transmission costs [13]. C.G.T also suggests reasonable fixed cost allocations that may be economically efficient as well as advantageous from engineering point of view [11]. Hence in this paper three C.G.T methods namely Nucleolus, Shapley and Proportional Nucleolus are attempted for transmission fixed cost allocation problem. Proportional Nucleolus method is proposed in this paper which is already in use for transmission loss cost allocation problem[18].

In this paper, section II gives the basic introduction of Game theory. Section III discusses about different methods of C.G.T. Section IV presents usage based methods for transmission fixed cost allocation. Section V details about general algorithm of C.G.T and presents a case study on IEEE 14 bus system, New England 39 bus system and Indian 75 bus system. Section VI explains about Monotonicity concept. Section VII concludes the paper.

Game theory

Game Theory is the formal study of decision making where several players must make choices that potentially affects the choice of other players. The Game theory is a kind of mathematical analysis designed to predict what the outcome or to predict the most likely result of a dispute between two individuals [17]. Game theory deals with any problem in which each player's strategy depends on what other players do. It is assumed that the rationality of all players is of common knowledge. A player is said to be rational if he seeks to play in a manner which maximizes his own payoff. Payoff is nothing but the payment received at the end of game. It is mainly employed in power systems to prevent collusion due to market power i.e. discourage collusions that could minimize payoff [5]. The game theory methodologies can be used to identify non-competitive situations (from market co-coordinator point of view) and minimize the risks in price decisions (from participant's point).

1. Characteristics of a game

In order to play a game, the players must exhibit the following characteristics [17]:

1. **Symmetry:** Two players, A and B, are symmetrical if all coalitions in which they can participate are met:

   $v(S \cup \{A\}) = v(S \cup \{B\}) \forall S \subset N \ \textit{such that} \ A, B \notin S$ (1)

2. **Attractiveness:** One player, A, is more desirable than another player, B, if it satisfies

   $v(S \cup \{A\}) \geq v(S \cup \{B\}) \forall S \subset N \ \textit{such that} \ A, B \notin S$ (2)

3. **Additivity:** If two or more coalitions agree to co-operate with each other then the payoff with the union can guarantee itself should be at least equal to the sum of their individual payoffs. That is

$$v(S \cup T) \geq v(S) + v(T), \forall S, T \subset N, \ if\ S \cap T = \phi \quad (3)$$

4. **Inessential:** For all the subsets of N if the equality holds in (3), then it is indifferent for the players to form any coalition and the game is called inessential.

$$v(S \cup T) = v(S) + v(T), \forall S, T \subset N, \ if\ S \cap T = \phi \quad (4)$$

In this case the characteristic function 'v' is just additive. For the inessential game it is:

$$v(N) = \sum_{i=1}^{n} v(i) \quad (5)$$

where n is the number of players in the game.

5. *Monotony:* Monotony means that when the characteristic function value v(S) of a coalition increases then the payoff to the members of this coalition is getting larger.

Solution methodologies in cooperative game theory

Terminology

Consider a game of N players with a characteristic function v. These players can form 2^N coalitions including the φ (null) coalition. The characteristic function v(S) assigns to each coalition say 'S' is the minimum payoff under any adverse conditions. This can be found by applying max-min criteria to S and (N-S) players. The set of all possible distribution of payoffs to the participants are called Imputations. A payoff vector $y = (y_1, y_2, \ldots\ldots y_n)$ is an imputation if it holds the following two conditions i.e., global rationality given by (6) and individual rationality given by (7).

$$\sum_{i=1}^{n} y_i = v(N) \quad (6)$$

$$y_i \geq v(i), i = 1, 2, \ldots. n \quad (7)$$

There are numerous methods for the allocation of benefits among the participants or players of a cooperative game. Some of them are briefly described below:

The core

One of the first solutions suggested for cooperative game is the core concept [7]. It is based on domination of imputations. That is, the core of a game is the set of all the imputations that are not dominated over any coalition. For an imputation to belong to the core, it must satisfy global rationality and coalitional rationality given by (8), (9).

$$\sum_{i=1}^{n} y_i = v(N) \quad (8)$$

$$\sum_{i=1}^{n_s} y_i \geq v(S) \quad \forall S \subset N \tag{9}$$

where n_s is the number of players in coalition 'S'.

It is clear that the core may include one or more than one imputation or may be even empty. Thus to choose a single solution whenever the core is non empty, Nucleolus concept was introduced in [6, 8].

The nucleolus

It is based on the idea of minimizing the dissatisfaction of the most dissatisfied groups. For a coalition S, measure of its dissatisfaction is the excess e(S):

$$e(S) = v(S) - y(S) \tag{10}$$

where $y(S) = \sum_{i=1}^{n_s} y_i$

Thus the larger the excess, the more dissatisfied the coalition is with this Imputation. Thus it reduces to the following optimization problem.

$$\min e \tag{11}$$

$$y(S) + e \geq v(S) \tag{12}$$

$$y(N) = v(N) \tag{13}$$

One main drawback of Nucleolus is that it is not monotonic i.e., even though the characteristic function v(S) of a coalition S is increased, the payoff to the members of this coalition is not affected.

The shapley value

For the foundation of Shapley value [8-9], [14-15] three axioms have to be settled.

i) Symmetry: $\phi_i(v)$ is independent of the labelling of the players.

$$\phi_{\Pi(i)}(v) = \phi_i(v) \tag{14}$$

ii) Efficiency: The sum of the expectations must be equal to the characteristic functional value for the grand coalition N.

$$\sum_{i=1}^{n} \phi_i(v) = v(N) \tag{15}$$

iii) Additivity: The sum of expectations, for a player, by playing two games with characteristic values v_1 and v_2 must be equal to the value if he played both games together.

$$\phi_i(v_1 + v_2) = \phi_i(v_1) + \phi_i(v_2) \tag{16}$$

Thus the Shapley value which satisfies three axioms is given by

$$\phi_i(v) = \sum_{S,i\in S} \frac{(n_s - 1)!(n - n_s)!}{n!} [v(S) - v(S - \{i\})] \quad (17)$$

Its main advantage is it exhibits monotonicity. However its main disadvantage is that it may or may not lie inside the core.

Proportional nucleolus

This solution concept coincides with the cases where core is nonempty. It is an important characteristic of extended core solution concept. It gains greater importance, as there are considerable number of games in which the core concept cannot be applied. As the extended core is a multiple valued concept, it is important to find a unique solution among its imputations. Hence Proportional Nucleolus (P-N) method is proposed in this paper for transmission fixed cost allocation problem, which is previously used for transmission loss cost allocation problem [18]. P-N method always chooses an imputation from the extended core in a similar way to the concept of Nucleolus. P-N method differs from the original Nucleolus in the definition of excess concerned with coalitions. It is defined as

$$e(Y:S) = \frac{v(s) - \sum_{i\in s} y_i}{v(S)} \quad (18)$$

The Proportional Nucleolus can expand the core to obtain a unique solution in both cases of empty core and large core. Thus the P-N method provides a better solution to both the extended core and core selection problem. The solution approach for P-N is to solve a linear program of the following problem formation:

$$Min\ e \quad (19)$$

subjected to

$$\sum_{i\in S} y_i \geq v(S)\ (1-e) \quad (20)$$

$$\sum_{i\in N} y_i = v(N) \quad (21)$$

The P-N of a game satisfies the properties namely:

- non emptiness
- single valuedness
- solution belonging to extended core.

P-N solution always lies inside the core and it is always monotonic as the excess value is proportional to the characteristic functional value. Thus it is the most efficient and plausible method among all the discussed game theoretic approaches.

Fixed cost allocation using usage based methods

The MW-Mile method

MW-Mile method takes into account the transacted power flow on all transmission lines, it can reflect not only the amount of wheeled energy, but also the path and distance of transfer.

However this method does not consider the economies of scale (The cost advantage that arises with increased output of a product. Economies of scale arise because of the inverse relationship between the quantity produced and per-unit fixed costs; i.e. the greater the quantity of a good produced, the lower the per-unit fixed cost because these costs are shared over a larger number of goods) of transmission network facilities and does not argue the stability of the solution.

For each transaction 'i'

$$f_{i,l} = C_l \left| P_{i,l} \right| \tag{22}$$

where, C_l = specific transfer cost of branch 'l'

$f_{i,l}$ =use of branch 'l' by participant 'i'

$P_{i,l}$ = power flow on branch 'l' by participant 'i'

Thus the total network usage for 'nl' number of lines is given by

$$f_i = \sum_{l=1}^{nl} f_{i.,l} \tag{23}$$

Thus the cost allocation by MW-Mile method is given by

$$MWM_i = K * (f_i \Big/ \sum_{j=1}^{n} f_j) \tag{24}$$

where 'K' is the total fixed cost to be allocated.

The drawback of this method is, it does not consider the direction of line flow.

The zero counter flow method

MW-Mile method does not consider the direction of power flow of each transaction. However, it is often argued that power flows having opposite direction from the net flow (the power flow due to all transactions) contribute positive in the system situation by relieving congestions and increasing the Available Transfer Capacity. Using Zero Counter Flow method transmission users are charged or credited based on whether their transactions lead to flows or counter flows with regard to the direction of net flows. The method suggests that if a particular transaction results in flows in the opposite direction of the net flow, then the transaction should be credited. Hence to accommodate this concept, Zero Counter Flow (ZCF) method is introduced. According to this method, the usage of a line by a particular transaction is set to zero if the power flow due to the transaction goes in the opposite direction of the net flow for the line.

Thus the change for each transaction 'i' is given as

$$f_{i,l} = \begin{cases} C_l P_{i,l} & P_{i,l} > 0 \\ 0 & P_{i,l} \le 0 \end{cases} \tag{25}$$

Thus the total network usage is given by

$$f_i = \sum_{l=1}^{nl} f_{i.,l} \tag{26}$$

Thus the cost allocation by ZCF is given by

$$ZCF_i = K*(f_i \Big/ \sum_{j=1}^{n} f_j) \quad (27)$$

But this method may fail to send right economic signals, i.e. it is well established from engineering point of view but subsidizes the largest network users with comparatively smaller users due to the counter flows of former. The savings due to counter flows are not allocated as payoffs to participants, which is a major drawback of ZCF method.

Hence to overcome the drawbacks of usage based methods, Game theory based methods are attempted in this paper.

Fixed cost allocation using cooperative game theory and a case study on ieee 14 bus system

Game definition

Many of the fixed cost allocation methods are based on the network usage from the side of market participants. The payment R_i allocated to each participant 'i' or player 'i' may be given by one of the following forms:

$$R_i = K*(f_i \Big/ \sum_j f_j) \quad (28)$$

$$R_i = f_i b \quad (29)$$

where b= cost of 1 MW power flow

General algorithm

Step1: Consider the number of possible coalitions that can be formed using the players (n) of the game.

Step2: Run DCOPF for each transaction 'i' and then calculate corresponding fixed cost f_i.

Step 3: Calculate characteristic function v(S) of each coalition

$$v(S) = (\sum_{i=1}^{n_s} f_i) - f_s \quad (30)$$

where f_s = usage of the network by coalition S

From (30) it is explicit that the characteristic function represents the savings that can be achieved in case of cooperation. It is obvious that for individual player i, it is v(i) = 0.

Step4: Using Nucleolus, Shapley Value and Proportional Nucleolus methods allocate the savings to all the players i.e., payoffs of the players y_i arose from the solution of the game.

Step 5: These payoffs are resulting in a reduction of f_i for each player:

$$f_i^{'} = \begin{cases} f_i - y_i & if \ f_i > y_i \\ 0 & if \ f_i < y_i \end{cases} \quad (31)$$

where $f_i^{'}$ is the new use of network by player i. If the savings assigned to player i are larger than the original f_i then the $f_i^{'}$ is set at zero. Thus, a player does not have the opportunity to receive money back from the network operator. The reason of making this adjustment is to prevent the misuse of game from the side of players.

Step 6: Calculate the amount that player 'i' has to pay. The cost allocation is done using the given formula

$$R_i = K \frac{f_i'}{\sum_{j=1}^{n} f_j'} \qquad (32)$$

When the electricity market operates in an environment of bilateral transactions then each transaction agent or player is responsible to pay a part of power system fixed cost. The formulation of a coalition between some players can be profitable by the existence of counter flows.

Case study on IEEE 14 bus system

The above algorithm is implemented on IEEE 14 bus system [16]. The loads are aggregated based on their Locational Marginal Prices (LMP) and then 4 transactions are formed in the system. The transactions are shown in the Fig. 1.

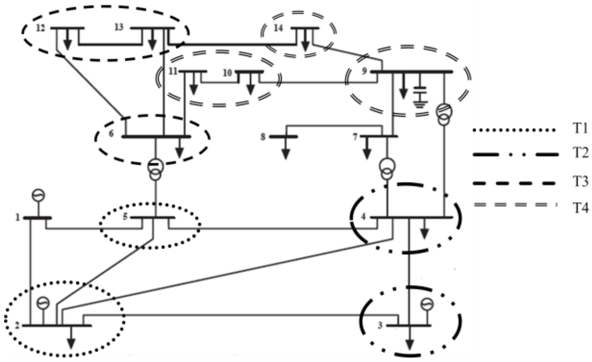


Figure 1: IEEE 14 bus test system

From Fig.1 it can be observed that first transaction is near to the generators and does not use much of the network. From Fig.1 it can be observed that first transaction is near to the generators and does not use much of the network. The generator power outputs are obtained by running DCOPF and thus the obtained transactions (Players) are as given in Table 1.

Table 1: Transaction data of IEEE 14 bus system

Transaction / Player (i)	Load demand (MW)	S (j, k)	B (i)
1	29.3	(1 → 24.070508), (2 → 5.229492)	2,5
2	142	(1 → 75.247070), (2 → 66.752930)	3,4
3	30.8	(1 → 19.452344), (2 → 11.347656)	6,12,13
4	56.9	(1 → 21.694922), (2 → 35.205078)	9,10,11,14

where S(j,k) = Bus 'j' supplying load 'k' for transaction 'i'. B(i) = Load Buses. In the above table, row 1, the first transaction comprises of a total load of 29.3 MW (buses 2 and 5 are grouped together based on their LMP's evaluated using Power World Simulator). This load is met by both generators with 1st generator generating 24.07 MW where as 2nd generator 5.23 MW.

By running a DCOPF for each transaction, the network usage and characteristic functional values of each coalition are obtained considering counter flows and are presented in table 2. The last row shows the grand coalition in which all players are present, which assures maximum savings.

From table 2 for coalition 5:

Players 1 and 2 forms coalition.

$f_s = 353.8507$

$f_1 = 31.1526$

$f_2 = 326.3217$

v(S) = $(f_1 + f_2) - f_s$ = (31.1526 + 326.3217) - 353.8507 = 3.6236 €

Similarly v(S) is calculated for each coalition. v(S) is the minimum amount which the coalition can assure itself. v(S) value obtained for grand coalition in table 2 is the maximum total savings i.e., 68.5833 € which is allocated to players in the game as their payoffs. In table 2 network usage values for each coalition are calculated by (26). v(S) values are calculated by (30).

Table 2: Characteristic functional values of IEEE 14 bus system

S. No.	Coalition	f_s	v(S)
1	1 0 0 0	31.1526	0
2	0 1 0 0	326.3217	0
3	0 0 1 0	133.0911	0
4	0 0 0 1	230.2376	0
5	1 1 0 0	353.8507	3.6236
6	1 0 1 0	161.3797	2.864
7	1 0 0 1	258.0073	3.3829
8	0 1 1 0	433.6315	25.7813
9	0 1 0 1	538.2311	18.3282
10	0 0 1 1	320.8602	42.4685
11	1 1 1 0	461.7372	28.8282
12	1 1 0 1	566.2091	21.5028
13	1 0 1 1	348.8808	45.6005
14	0 1 1 1	623.9727	65.6777
15	1 1 1 1	652.2197	68.5833

Next is to calculate the minimum and maximum values of payoffs y_i. $y_{i\,\min}$ is taken as v(S) when player 'i' acts alone i.e., zero for all 4 players. $y_{i\,\max}$ is taken as $v(S \cup \{i\}) - v(S)$ [1].

For player 1:

$y_{1\,\max} = v(15) - v(14) = 2.9056$

Similarly for the remaining 3 players maximum limits are determined and are shown in table 3.

Table 3: Maximum limit of Payoffs

Player	y_i **max**
1	2.9056
2	22.9828
3	47.0805
4	39.7551

The payoffs and the new usage of network of player 'i' obtained by Nucleolus, Shapley Value and Proportional Nucleolus methods are shown in tables 4, 5 and 6.

Table 4: Payoffs and new network usage of 4 players in Nucleolus method

Player	y_i	f_i'
1	1.45	29.7026
2	3.62	322.7017
3	45.63	87.4611
4	17.88	212.3576
Total	68.583	652.223

Table 5: Payoffs and new network usage of 4 players in Shapley Value method

Player	y_i	f_i'
1	0.0	31.1526
2	3.78	322.5417
3	26.32	106.7711
4	38.48	191.7576
Total	68.58	652.223

Table 6: Payoffs and new network usage of 4 players in Proportional Nucleolus method

Player	y_i	f_i'
1	0.0	31.1526
2	3.78	322.5417
3	26.32	106.7711
4	38.48	191.7576
Total	68.58	652.223

From tables 4, 5 and 6, it is observed that the sum of the payoffs of 4 players is equal to v(S) of grand coalition in table 2. That means the payoffs satisfied the condition shown in (6). New usage of network by player 'i' is $f_i^{'}$ and is calculated using (31). From these tables 4, 5 and 6, the total network usage by all the 4 players is equal to the value obtained for f_i of grand coalition value in table 2. This indicates that when the 4 players acting individually the total network usage is 720.803 € whereas when 4 players forms a grand coalition the total network usage is reduced to 652.2197 €. Finally the allocations to players can be computed by (32).

Next step is to calculate the total fixed cost to be covered by the market participants i.e., K. This cost is calculated by multiplying the power flows with their corresponding line lengths and line costs. Table 7 shows the allocation of K = 2773.35 € to four players with all the above discussed methods.

Table 7: Cost allocation using various methods in IEEE 14 bus system

Player	MWM (€)	ZCF(€)	Nucleolus(€)	Shapley Value(€	Proportional Nucleolus
1	104.50	119.86	126.28	122.56	121.13
2	1340.23	1255.5	1372.16	1322.38	1331.65
3	528.08	512.07	371.91	450.01	464.96
4	800.52	885.85	902.98	878.39	855.59

Players are in the ascending order of 1,3,4,2 w.r.to demand as can be seen from table 1. 3rd player utilizes more network compared to 4th player because 3rd player accounts for higher line lengths than 4th player. 4th player got 902.9843 € by Nucleolus method & 878.3903 € by Shapley value method, whereas by P-N method the cost is still reduced to 855.5968 €. 3rd player got 371.9095 € by Nucleolus method and 450.0096 € by Shapley value method whereas by P-N method this player got 464.9632 €. As 3rd player uses more network, the cost is increased by P-N method. Remaining cost is allocated to other 2 players. The comparison graph is shown in Fig. 2 further supports the results obtained using various methods.

From Fig. 2, we can infer that the cost allocated using Game Theory is in tune with other fixed cost allocation methods.

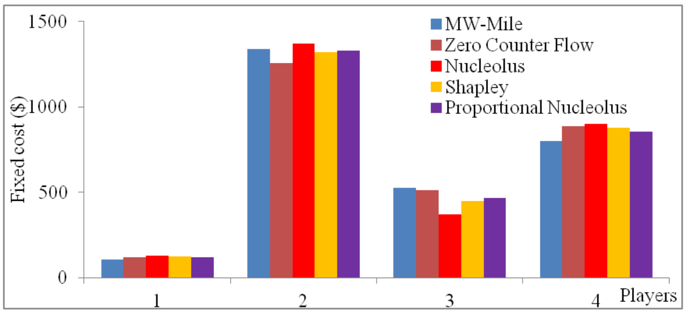


Figure 2: Comparison of allocations in IEEE 14 bus system with different methods

Case study on New England 39 bus system

The developed algorithms are tested on New England 39 bus system. Loads are aggregated to form two transactions using LMP's computed in Power World Simulator. The resulting transactions are as shown in table 8. The loads are aggregated based on their Locational Marginal Prices (LMP) and then 2 transactions are formed in the system. The data for New England 39 bus system is taken from [20].

a) Nucleolus method: The network usage and characteristic functional values for each of the three coalitions are shown in table 9.

Table 8: Transaction data of New England 39 bus system

Transaction / Player (i)	Load demand (MW)	S (j → k)	B(i)
1	4184.8	(30→90.0), (31→600.0), (32→460.0), (33→650.0), (34→608.0), (35→357.95), (36→279), (37→297.85), (38→387), (39→455)	3,4,7,8,12,15,16, 18,20,21,23,24,27
2	1965.7	(30→90.0), (31→249.0), (32→170.0), (33→307.8), (34→232.9), (35→170), (36→152), (37→148), (38→206), (39→240)	25,26,28,29,31,39

v(S) value obtained for global coalition in table 9 is the maximum total savings i.e., 39671.854 $ allocated to both the players in the game as their payoffs. Maximum values of payoffs y_i are shown in table 10 and these are calculated in the same way as in IEEE 14 bus system. y_i min is zero for both the players.

The payoffs and the new usage of network by player 'i' obtained by Nucleolus method are shown in table 11. The payoffs satisfied the condition shown in equation (6) i.e., they met the global rationality.

Table 9: Network usage by each coalition and characteristic functional value

S. No.	Coalition	f_i	v(s)
1	10	26707.491	0
2	01	48448.721	0
3	11	35484.358	39671.854

Table 10: Maximum limit of payoffs

Player	y_i max
1	39671.854
2	39671.854

Table 11: Payoffs and new network usage of 2 players in Nucleolus method

Player	y_i	$f_i^{'}$
1	19835.927	6871.564
2	19835.927	28612.794
total	39671.854	35484.358

In table 11, the total network usage by both the players is equal to the value obtained for f_i corresponding to global coalition value in table 9. Here also total network usage is reduced (i.e., 35484.358 $) when the 2 players form global coalition rather than when they act individually (i.e., 75156.212 $). This same observation was made in Shapley Value and Proportional Nucleolus methods also.

b) The Shapley Value method: The network usage and characteristic functional values of each coalition are same as shown in table 9. The min and max limits of payoffs are same as in Nucleolus method. The payoffs ϕ_i and the new usage of network by player 'i' obtained by Shapley Value method are shown in table12. The payoffs meet the global rationality. The solution lies inside the core because "*coalitional rationality*" i.e., equation (9) is satisfied.

Table 12: Payoffs and new network usage of 2 players in Shapley Value method

Player	ϕ_i	$f_i^{'}$
1	19835.927	6871.564
2	19835.927	28612.794
total	39671.854	35484.358

From table 12, it can be observed that the total network usage by both the players is equal to the value obtained for f_i corresponding to global coalition value in table 9.

c) Proportional Nucleolus method: The network usage and characteristic functional values of each coalition are same as shown in table 9. The min and max limits of payoffs are same as in Nucleolus method. The payoffs y_i and the new usage of network by player 'i' obtained by Proportional Nucleolus method are shown in table 13. The payoffs meet the global rationality.

Table 13: Payoffs and new network usage of 2 players in Proportional Nucleolus method

Player	y_i	$f_i^{'}$
1	16131.2815	10576.2095
2	23540.5725	24908.1485
total	39671.854	35484.358

From table 13, it can be observed that the total network usage by both the players is equal to the value obtained for f_i of global coalition value in table 9.

All the payoffs obtained by Nucleolus, Shapley and Proportional Nucleolus methods are within the min., and max. limits and all these payoffs satisfy individual, collective, global rationalities. It indicates the payoff solutions obtained with all the three methods are within the core. Table 14 shows the allocation of K= 34430 $ (calculated in a similar manner as was done in 14 bus system) to two players.

Table 14: Transmission fixed cost allocation using various methods in New England 39 bus system

Player	MW-Mile Method ($)	Zero Counter Flow method	Nucleolus Method ($)	Shapley Value Method ($)	Proportional Nucleolus Method ($)
1	15523.229	12235.035	6667.387	6667.387	10261.955
2	18906.77	22194.964	27762.613	27762.613	24168.044

Transaction 1 i.e., T1 has more load than T2, but T2's network usage is more than T1; that means T2 accounts for higher line lengths than T1. So cost allocated to T2 is more than T1 in MW–Mile method. In ZCF method counter flows are more for T1, hence cost allocation for T1 is still reduced. For T2 counter flows are very less and the remaining cost is allocated to T2. In Nucleolus and Shapley value methods; due to payoffs, T1's share is further reduced and T2's share is raised in transmission cost allocation. But as Proportional Nucleolus method is an extended core method; even though T2 accounts for higher line lengths, T1's load is more than T2 and therefore T1's share is improved, T2's share is reduced when compared to Nucleolus and Shapley methods in transmission cost allocation.

The comparison graph is shown in Fig. 3.

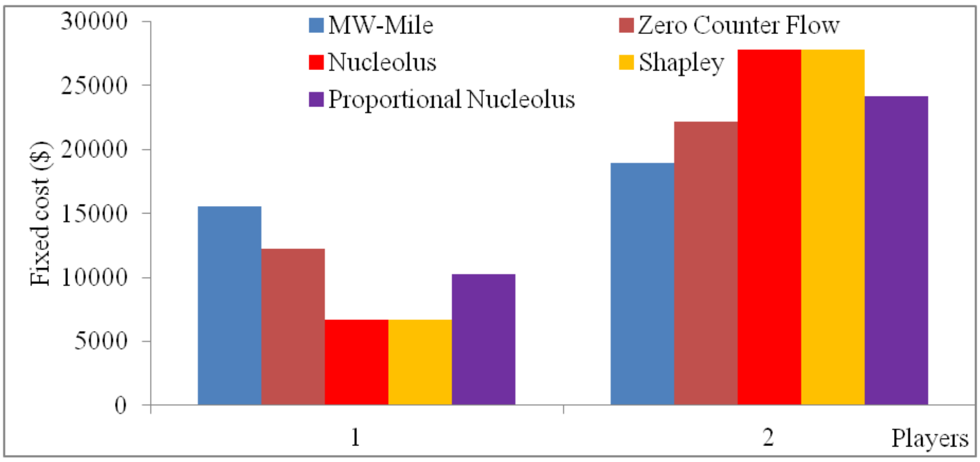


Figure 3: Comparison of allocations in New England 39 bus system with different methods

Case study on Indian 75 bus system

The developed algorithms are also tested on Indian 75 bus system. Loads are aggregated to form two transactions using LMP's computed in Power World Simulator. The formed transactions are as shown in table 15. The data for Indian 75 bus system is taken from [21].

a) Nucleolus method: The network usage and characteristic functional values for each of the three coalitions are shown in table 16.

v(S) value obtained for global coalition in table 16 is the maximum total savings i.e., 265.195421 \$ allocated to both the players in the game as their payoffs. Maximum values of payoffs y_i are shown in table 17. y_i min is zero for both the players.

Table 15: Transaction data of Indian 75 bus system

Transaction / Player (i)	Load demand (MW)	S(j → k)	B(i)
1	5199.26	(1 → 847.74), (2 → 331.63), (3 → 258.77), (4 → 25.91), (5 → 93.98), (6 → 205.75), (7 → 90.60), (8 → 68.56), (9 → 296.25), (10 → 62.00), (11 → 19.52), (12 → 1704.82), (13 → 806.88), (14 → 216.66), (15 → 170.17)	16, 20, 24, 25, 27, 28, 30, 32, 34, 37, 39, 42, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75
2	368.86	(1 → 15), (2 → 15), (3 →15), (4 → 15), (5 → 15), (6 → 15), (7 → 15), (8 → 15), (9 → 15), (10 → 15), (11 → 15), (12 →15), (13 → 15), (14 → 158.86), (15 → 15)	57,58,59

The payoffs and new usage of network by player 'i' obtained by Nucleolus method are shown in table 18. The payoffs meet the global rationality.

In table 18, the total network usage by both the players is equal to the value obtained for f_i corresponding to global coalition value in table 16. Here also total network usage is reduced (i.e., 4180.0393 \$) when the 2 players form global coalition rather than when they act individually (i.e., 4445.239012 \$). The same observation was made in Shapley Value and Proportional Nucleolus methods also.

Table 16: Network usage by each coalition and characteristic functional value

Sl. no.	Coalition	f_i	v(s)
1	10	3993.270032	0
2	01	451.968980	0
3	11	4180.043591	265.195421

Table 17: Maximum limit of payoffs

Player	y_i max
1	265.195421
2	265.195421

Table 18: Payoffs and new network usage of 2 players in Nucleolus method

Player	y_i	$f_i^{'}$
1	132.60	3860.670032
2	132.60	319.36898
total	265.2	4180.0393

b) The Shapley Value method: The network usage and characteristic functional values of each coalition are same as shown in table 16. The min and max limits of payoffs are same as in Nucleolus method. The payoffs ϕ_i and the new usage of network by player 'i' obtained by Shapley Value method are shown in table 19.

Table 19: Payoffs and new network usage of 2 players in Shapley Value method

Player	ϕ_i	$f_i^{'}$
1	132.597711	3860.672321
2	132.597711	319.371269
total	265.195422	4180.04359

The payoffs meet the global rationality. The solution lies inside the core because "*coalitional rationality*" is satisfied. From table 19, it can be observed that the total network usage by both the players is equal to the value obtained for f_i corresponding to global coalition value in table 16.

c) Proportional Nucleolus method: The network usage and characteristic functional values of each coalition are same as shown in table 16. The min and max limits of payoffs are same as in Nucleolus method. The payoffs y_i and the new usage of network by player 'i' obtained by Proportional Nucleolus method are shown in table 20. The payoffs meet the global rationality.

From table 20, it can be observed that the total network usage by both the players is equal to the value obtained for f_i of global coalition value in table 16.

All the payoffs obtained by Nucleolus, Shapley and Proportional Nucleolus methods are within the min., and max. limits and all these payoffs satisfy individual, collective, global rationalities. It indicates the payoff solutions obtained with all the three methods are within the core. Table 21 shows the allocation of K= 3515.03 $ (calculated in a similar manner as was done in 14 bus system) to two players.

Table 20: Payoffs and new network usage of 2 players in Proportional Nucleolus method

Player	y_i	f_i'
1	118.9481	3874.3219
2	146.2473	305.7216
total	265.1954	4180.0435

Table 21: Transmission fixed cost allocation using various methods in Indian 75 bus system

Player	MW-Mile Method ($)	Zero Counter	Nucleolus Method ($)	Shapley Value Method ($)	Proportional Nucleolus Method ($)
1	3271.62	3157.63	3246.46	3246.46	3257.94
2	243.4	357.39	268.56	268.56	257.08

The comparison graph is shown in Fig. 4.

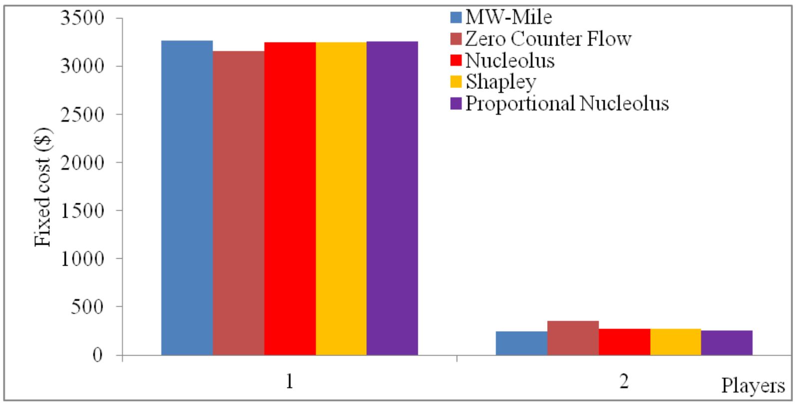


Figure 4: Comparison of allocations in Indian 75 bus system with different methods

From Fig. 4 it can be understand that, in 75 bus system 2nd player demand is less and it uses less network than 1st player; so cost allocated to 2nd player is less than 1st player in all the methods. With P-N method the cost allocated for 2nd player is minimum.

Monotonicity

If v(S) of a coalition is increased, then the payoffs of all the players belonging to that coalition should also increase. 8th coalition of table 2 is considered in all methods for monotonicity testing, in which players 2 and 3 forms a coalition. Base case payoffs of players 2 and 3 are given in tables 4, 5 and 6. The base case characteristic functional value of 8th coalition is v(S) = 25.78 $. Assume v(S) value is increased to 29.78 $ for monotonicity testing. Then the new payoffs are given in tables 22, 23 and 24.

Table 22: New payoffs in Nucleolus method

Player	y_i
1	1.45
2	21.53
3	9.70
4	35.90

Table 23: New payoffs in Shapley value method

Player	ϕ_i
1	1.9950
2	15.6645
3	27.5939
4	23.3297

Table 24: New payoffs in Proportional Nucleolus method

Player	y_i
1	0.0
2	20.97
3	44.09
4	3.53

- From the new payoffs it can be observed that in Nucleolus method player 2 payoff is improved but player 3 payoff is reduced, so Nucleolus method does not exhibit monotonicity. This method favors player 2 but not player 3.
- In Shapley value method payoffs of both 2nd and 3rd players are improved slightly; hence Shapley value method exhibits monotonicity. Shapley value method payoffs are non-zero for all the case studies.
- In Proportional Nucleolus method payoffs of both 2nd and 3rd players are improved significantly; hence Proportional Nucleolus method exhibits monotonicity.

Conclusions:

In this paper cooperative game theory is implemented for power system fixed cost allocation in a transaction based market model in an equitable manner and the results obtained are compared with conventional usage based methods like MW-Mile method and Zero counter flow method. The study of the cost allocation is of high importance and it is at the origin of discussion and study all over the world. Since players behave in rational way, the cost allocation problem becomes a matter of conflict. Cooperative game theory is used to deal with such matters of conflict. In MW-Mile method counter flows are not accounted. In ZCF method counter flows are accounted but the savings due to counter flows are not allocated to players which could be achieved with game theory methods. Hence game theory methods give correct economic signals about the allocations of transmission fixed cost to players in the system. In the case of a pool market, concerning the whole system, there is no obstacle for such an implementation. However, negative characteristic function values may arise if the game is played at each system branch. For a bilateral transaction market, the fixed cost allocation can take place in the entire network as well as at each single branch. Monotonicity for different game theory methods is tested on 8^{th} coalition of IEEE 14 bus system. All the results of IEEE 14 bus system satisfy individual, coalition and global rationalities. Nucleolus is not monotonic, solution always lies within the core if the core is non-empty and it may favor some players only. If the core is empty Nucleolus method cannot produce solution. Shapley value method is monotonic and always assigns a non zero payoff to the players. But the solution with Shapley Value method may or may not lie within the core. Hence Proportional Nucleolus method is proposed for fixed cost allocation problem in this paper to overcome the drawbacks of Nucleolus and Shapley Value methods. P-N method is also monotonic and the solution is always lies within the core for both empty and non empty core cases due to the extended core concept used in P-N method. Due to its inherent property of extended core concept, a better solution is obtained by P-N method in the presented case study.

References:

1. G.C.Stamsis and I.Erlich, "Use of Cooperative Game Theory in Power System Fixed Cost Allocation", *IEE Proc. Generation, Transmission and Distribution*, Vol.151, No.3, May 2004.
2. Research thesis on "Power Transmission cost calculation in deregulated electricity market" by George Stamsis.
3. Green R, "Electricity Transmission Pricing: an International Comparison", *Utility Policy*, 1997, 6, (3), pp. 177-184.
4. Lima.J.W.M, "Allocation of Transmission Fixed Charges: An Overview", *IEEE Trans. Power System*, 1996, 11, (4), pp.1409-1418.
5. "Introduction to Game Theory and its application in Electrical Power Markets", *IEEE Computer Applications in Power*, CAP Tutorial.
6. Thomas L.C, "Game Theory and its application" (Ellis Horwood Limited, Chichester, 1984).
7. D.B.Gillies, "Solution to general non zero sum games" in Contribution to the Theory of Games, Vol IV, A.W. Tuckerand D.R.Luce, Ed Princeton: Princeton University Press, 1959, pp 47-85.

8. Schmeidler. D, “The nucleolus of a characteristic Function game”, SIAM J. Appl. Math., 1969, 17, pp. 1163 -1170.
9. L.S.Shapley, “A value of n person games”, in Contribution to the Theory Of Games.
10. Tsukamoto. Y, and Iyoda. I, “Allocation of fixed transmission cost to wheeling transactions by cooperative game theory”, *IEEE Trans.Power Syst*., 1996, 11, (2), pp. 620–629.
11. Singh. H, Hao. S, Papalexopoulos. A, and Obessis. M, “Cost allocation in electric power networks using cooperative game theory”, in *Proc. 12th Power System Computation Conf. (PSCC)*, Dresden, Germany, 1996, pp. 802–807.
12. Zolezzi. J. M, and Rudnick. H, “Transmission cost allocation by cooperative games and coalition formation”, *IEEE Trans. Power Syst*., 2002, 17, (4), pp. 1008–1015.
13. Contreras. J, and Wu. F.F, “A kernel-oriented algorithm for transmission expansion planning”, *IEEE Trans. Power Syst*., 2002, 15, (4), pp. 1434–1440.
14. Shapley. L.S, “Contributions to the theory of games II” (Princeton University Press, Princeton, NJ, 1953), pp. 307–317.
15. Tan. X, and Lie. T.T, “Application of the Shapley value on transmission cost allocation in the competitive power environment”, *IEE Proc., Gener. Transm. Distrib*., 2002, 149, (1), pp. 15–20.
16. P.K.Iyambo, R.Tzoneva, “Transient Stability Analysis of the IEEE 14-Bus Electric Power System”, *IEEE conference, Africon* 2007.
17. Filipe Azevedo, H.M. Khodr, Zita A. Vale, “Transmission cost allocation using cooperative game theory: A comparative study”, *IEEE 6th International Conference on the European Energy Market*, 2009.
18. P.V. Satya Ramesh, C. Radhakrishna, “Use of cooperative game theory concepts for loss allocation in multiple transaction electricity markets”, *Journal of Electrical Systems*, Vol. 5, no.1, 2009.
19. George A. Orfanos, Pavlos S. Georgilakis, Nikos D. Hatziargyriou, “A More Fair Power Flow Based Transmission Cost Allocation Scheme Considering Maximum Line Loading for N-1 Security”, *IEEE Trans. Power Syst*., 2013, 28, (3), pp. 3344–3352.
20. http://sys.elec.kitami-it.ac.jp/ueda/demo/WebPF/39-New-England.pdf
21. Indian 75 bus power system data “obtained from Indian Institute of Technology Kanpur”.

SOCIAL WELFARE MAXIMIZATION IN A RESTRUCTURED POWER SYSTEM USING BAT ALGORITHM

Murali Matcha*, Siddheswar Kar and Neha Verma

Department of Electrical Engineering,

Medi-Caps University, Indore, M.P, India

*Corresponding author E-mail: murali_233@yahoo.com

Abstract:

The basic key issue in restructured power market is to design an appropriate auction mechanism. This paper discusses the influence of elastic demand on congestion and Locational marginal prices (LMP's) in double auction electricity market. This double auction problem of the pool based day-ahead electricity market is formulated as maximization of social welfare. This problem is subject to equality constraint given by real power balance equation and security constraint in the form of active power flow limits over the congested lines. DCOPF is adapted to solve this pool market problem by computing LMP's at all buses under three different loss cases such as without loss case, concentrated loss case and distributed loss case. Linear bids are used for both generators and consumers. DCOPF is executed with Genetic algorithm and Bat algorithm. The algorithm is tested on three test systems e.g., IEEE 14 bus system, New England 39 bus system and Indian 75 bus system (practical data). In all the three test systems Bat algorithm has produced better social welfare than Genetic algorithm under all the three loss cases.

Keywords: Elastic demand, Social Welfare, Bat algorithm, LMP, Concentrated loss, Distributed loss

Introduction:

In restructured electric power industry, one of the essential components for electric power market design is to find an appropriate auction mechanism, capable of creating more fair competition and more efficient operation [5, 6]. In microeconomic principles, if an allocation of resources maximizes the total social welfare, it is said that the allocation exhibits efficiency [19]. In inelastic market ISO does not provide sufficient signals or incentives to a customer to adjust its demand in response to the price, i.e., the customer does not have any motivation to adjust its demand for electrical energy to adapt the market conditions and customer does not respond to market price. This means that the demand side is unable to protect itself by curtailing demand in response to high prices, and strategic behaviour of power suppliers will lead to price spikes. Lack of demand elasticity confers market power on suppliers. In a market that has fixed demand, MCP for energy is determined by the price structure of supply offers. Customers use the concept of *elastic demand* when they are exposed to and aware of the price of energy and arrange their activities in such a fashion as to reduce their demand as the price of the next available offer exceeds a certain level.

The assumption that each supplier or consumer cannot affect the price by its actions i.e., all market participants take the price as given, and then the market is said to be a *perfectly competitive market*. This assumption is usually not true for electricity markets. In a competitive market, it is the combination of all the consumers on one side and all the suppliers on the other side that determines the price. A graphical representation of demand and supply curves is shown in figure 1 with no transmission constraints. In figure 1 the market displays elastic demand, where the load responds directly to the price of supply offers, i.e., the demand varies with the price of offer. The intersection of demand and supply curves in figure 1 indicates that the operating point (the equilibrium point) of the market which defines the current MCP as well as the current level of electrical load. The former task is achieved by solving an optimal power flow (OPF), with the objective of maximizing social welfare or social surplus. Elastic markets can also be called as double auction model markets because both suppliers and consumers involve in the auction [12]. The explicit demonstration about social welfare maximization in double auction model with elastic demand during congestion in the transmission system and its nodal pricing due to congestion and losses in a pool market is presented in this paper.

Supply-demand equilibrium will set the market price and quantities. This equilibrium point is the point where the social welfare, subject to constraints is optimal [21]. Social welfare is the difference between total consumer benefit B(d) and total supplier cost C(s) i.e., (B(d) – C(s)). Graphically this equilibrium point can be described as the point where the incremental aggregate cost curve and the incremental aggregate utility curve cross. At this point any additional trade between the supply and demand sides will reduce the social welfare. Aggregating the individual incremental (marginal) cost bids of the generators forms the supply curve and the demand curve is formed by aggregating the incremental (marginal) utility of the loads. Some markets, in England and the PJM interconnection consists of only supply-side bidding. Few other markets, in New Zealand and California incorporate demand-side bidding also, allowing the consumers in the market to react to pricing [13].

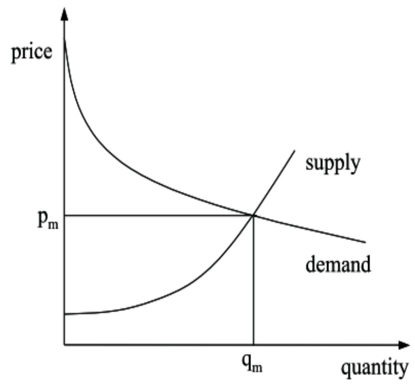


Figure 1: Elastic market with unconstrained transmission

In an unconstrained market, where the network is not considered, the solution of the optimization problem is characterized by a certain value of social surplus and the demand prices and generator prices at each bus are equal. In this situation a *market clearing price* is defined and it is unique for the whole system. On the contrary, if the network is constrained congestion can arise causing, along with losses, a different optimum, with a lower level of social surplus. In this situation a single market clearing price does no longer exist. The nodal price at each bus may vary according to the determined MW demand and supply, and their locations which reflect system loss and network congestion [1], [2]. The consumer located at bus 'i' will pay at nodal price 'i' for its marginal energy consumed and the supplier located at bus 'j' will be paid at nodal price j for its marginal energy supplied [3], [4] and [17].

Supplier and customer functions representation

Most of the electricity markets throughout the world include supplier bidding only, but some markets include consumer bidding also. So, it is very important to define the consumer utility function and supplier cost function. In literature most of the papers presented Social welfare maximization (SWM) problem with linear elastic supplier and customer curves [13], [15], [16] and [20]. In this paper a continuous bid curve model i.e., linear bid (corresponding to quadratic curves) is assumed for both suppliers and consumers. According to economic theory, a firm in a perfectly competitive market maximizes its profit when it sells/buys at its marginal cost/benefit. This implies that profit maximizing generators/loads will bid at marginal costs/marginal benefits, so ideally generators/loads should bid their true marginal cost/marginal benefit curves [20]. In this paper for bidding purpose, marginal cost curves are used for suppliers and consumers.

1. Supplier cost function

Supplier cost is the actual cost of producing power by a generator. Generally two types of generators are used by the suppliers: a hydro generator and a thermal generator. Hydro generators have a constant marginal cost and have different capacities depending on the water levels. The present study considered thermal generators only. The marginal cost for thermal generators vary with the amount of power generated and the data can be obtained from hourly generator bid data obtained from ISO. The capacity of a thermal generator is constant, independent of time [20]. The supplier bid curve is a positive linearly increasing price function w.r.to generated power. The generation cost function is the integration of the marginal cost function [20] and is represented as:

$$C_i(P_{Gi}) = a_i + b_i P_{Gi} + c_i P_{Gi}^2 \ \text{(\$/hr)} \quad (1)$$

where a_i, b_i, c_i are the positive coefficients.

The supplier bid = $dC_i(P_{Gi})/dP_{Gi} = b_i + 2c_i P_{Gi}$ (\$/MWh) (2)

The continuous bid curve of the supplier (linear bid) is shown in figure 2.

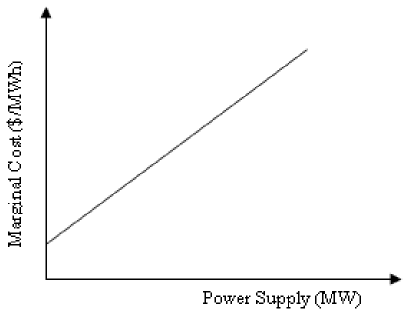


Figure 2: Linear bid curve of a thermal generator

2. Customer utility function

Customer's utility is a measure of the customer's benefit from the utilization of power, which can be obtained from the consumer's bid curve. The bid curve represents the relationship between the price a customer is willing to pay and the amount of electric power demanded. The bid curve indicates how much is the additional MWh worth to the customer as the customer considers using one more MWh of electric energy. Every customer has his/her own bid curve. The aggregated bid curve represents different utility levels of the customers in the aggregation. Customer can be modelled in a manner analogous to the suppliers. A demand function is used, which is mathematically similar to the supply function, except that the demand function decreases as the price increases. The willingness to pay of the customer at bus k can be expressed through a positive linearly decreasing function [20]. The utility function is the integration of the marginal utility function and is represented as:

$$B_i(P_{di}) = b_i P_{di} + c_i P_{di}^2 \text{ (\$/hr)} \tag{3}$$

where b_i is positive coefficient and c_i is negative coefficient.

The customer bid = $dB_i(P_{di})/dP_{di} = b_i + 2c_i P_{di}$ (\$/MWh) (4)

The bid curve of the customer (linear bid) is shown in figure 3.

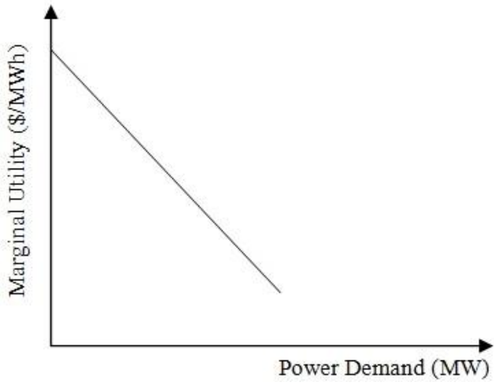


Figure 3: Linear bid curve of the customer

Concept of dead weight loss

The sum of the net consumers' surplus and of the producers' profit is called the *social welfare*. It quantifies the overall benefit that arises from trading. The global welfare is maximum when a competitive market is allowed to operate freely and the price settles at the intersection of the supply and demand curves. The market clearing price is π^* and the market clearing volume is q^*. Under these conditions, figure 4 shows that the consumers surplus is equal to the sum of the areas labelled A, B and E and the producers' profit is the sum of the areas labelled C, D and F. Producers surplus is defined as the amount of revenue received from selling the power to ISO, minus the cost of supplying the power. Consumer's surplus is defined as the amount consumer is willing to pay, minus actual amount paid by the customer to ISO for consuming the power.

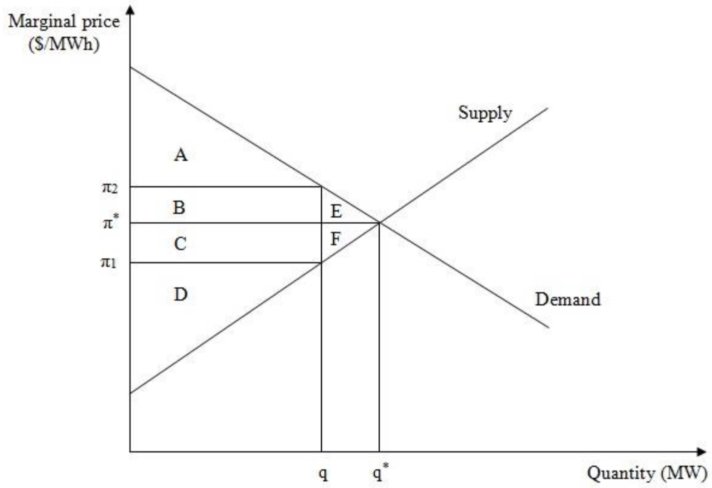


Figure 4: Global welfare and deadweight loss

When congestion occurs in the transmission system the amount of power traded reduces from q^* to q. The corresponding prices are π_1 and π_2, called Locational Marginal prices which includes congestion price and loss price. The consumers' surplus shrinks to area A and the producers' surplus shrinks to area D. The ISO collects the difference π_2 - π_1 for each MW traded. The total amount collected by ISO in the form of congestion taxes is equal to the sum of areas B and C, which is also called as merchandising surplus. Merchandising surplus is defined as the amount paid by the customer to ISO for consuming the power, minus amount of revenue received by producer from selling the power to ISO. Due to congestion the global welfare reduces by an amount equal to the sum of areas E and F. This drop in global welfare is called the *deadweight loss* and is the result of the reduction in the amount traded caused by the price distortion [14]. The area due to Deadweight loss is neither useful to supplier nor useful to customer or ISO. This is one of the major drawbacks in electricity trading during system congestion.

Under the minimum cost centralized pool schedule, the generator or supplier profit can lead to an incongruity, i.e., units that are scheduled on, make negative profit. If allowed to self-schedule, the ISO presumes that such units would rather turn off in order to function at zero profit [18].

Social welfare maximization (SWM) for LMP calculation with different loss cases

Social welfare is originally defined by the Bergson – Samuelson [20] social welfare function. The new OPF is characterized by a composite objective function and is formulated with two terms, the total benefits and the total costs, whose difference represents the Social welfare (S.W) or Social surplus (S_S).

$$S.W \ or \ S_S = \sum_{k=1}^{N_e} B_k(D_k) - \sum_{i=1}^{N_g} C_i(P_i) \qquad (5)$$

In the function S_S, N_e is the number of buses with elastic load demand; N_g is the number of generators in the studied system. The ISO maximizes the social welfare subjected to an OPF solution based on all bids in the market [11]. In the transmission management, where the system operation and transmission functions are managed by a regulated transmission organization, the objective of transmission enhancement is to maximize social welfare. For the other transmission management practices, social welfare is also an important consideration and needs to be well defined [22].

In this paper LMP computation [7-10] with social welfare maximization as objective using DCOPF, with various loss cases has been studied. DCOPF (linear model) is used due to its fast convergence, with an assumption that the voltage profile is sufficiently flat, and the R/X ratio is less than 0.25. LMPs' using DCOPF are computed with GA and Bat algorithms under three different loss cases such as without loss, concentrated loss and distributed loss cases. The obtained power generations and demands are used in calculation of LMP's with and without loss for the congested transmission system. Generation Shift factors (GSF) have been used for the calculation of transmission line flows. Delivery factors (DF) at all buses have been used to consider the impact of marginal losses on LMPs' [9]. In double auction model only major suppliers and customers are allowed for bidding. Hence all the suppliers and customers are not accounted for bidding. It is assumed that remaining suppliers and consumers will be engaged in bilateral transactions. This type of market is called as Coexisting Pool and Bilateral Model or Hybrid model as is shown in figure 5.

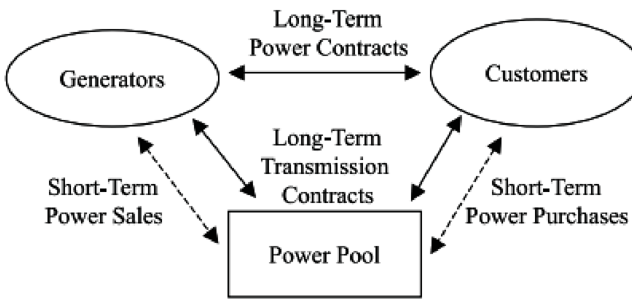


Figure 5: Coexisting Pool and Bilateral Model

The following section describes the mathematical model of different cases studied.

Case 1: Without loss

In this method the objective function is maximization of total social welfare subjected to power balance and line flow constraints. Active power generations of the generators and all the elastic customer demands are considered as set of variables for optimization problem. LMP's are calculated after obtaining generator power outputs and the customer demands from the DCOPF problem. Supplier surplus, Customer surplus, Merchandising surplus or ISO surplus and finally Social surplus (which is a sum of all the previous three surpluses) are calculated.

The objective function is

$$Maximize \;\; J = B(d) - C(s) \tag{6}$$

$$s.t \;\; \sum_{i=1}^{N} P_{Gi} = \sum_{i=1}^{N} P_{Di} \tag{7}$$

$$F_k \leq \text{limit}_k, \quad \text{for k=1, 2 …M} \tag{8}$$

$P_{Gi}^{\min} \leq P_{Gi} \leq P_{Gi}^{\max}$ for i=1, 2 …N (9)

$P_{Di}^{\min} \leq P_{Di} \leq P_{Di}^{\max}$ for i=1, 2 …N (10)

where, N is number of buses, M is number of lines, B(d) is the total benefit function of all the elastic customers, C(s) is the total fuel cost function of all the suppliers, P_{Gi} is real power output of generator at bus i (MW), P_{Di} is the real power demand at bus i (MW), F_k is line flow of line k, $limit_k$ is thermal limit of line k.

After getting power outputs of generators, slack bus power is calculated using equation (7) and the price at the reference (slack) bus needs to be calculated by substituting slack bus power in its linear bid curve. At the reference bus, both loss price and congestion price are always zero. Therefore, the price at the reference bus is equal to the energy component. Now the LMP formulation at a bus B can be written as

$LMP_B = LMP^{energy} + LMP_B^{cong} + LMP_B^{loss}$ (11)

The decomposition of LMP is shown here

$LMP^{energy} = \lambda$ = price at the reference bus (12)

$MP_B^{cong} = -\sum_{k=1}^{M} GSF_{k-B} \times \mu_k$ (13)

where μ_k is the constraint cost or shadow price of line k, defined as:

$$\mu_k = \frac{change\ in\ total\ \text{cost}}{change\ in\ constra\text{int's}\ flow}$$

$LMP_B^{loss} = \lambda \times (DF_B - 1)$ (14)

($LMP_B^{loss} = 0$ for lossless power system)

If the losses and congestion are not considered (or they are not active at the optimal point), the solution of the optimization problem is characterized by a certain value of social surplus. The demand prices and the generator prices at each bus are equal. In this situation a market clearing price is unique for the whole system.

Social Welfare (S.W) = Supplier surplus (S^G) + Customer surplus (S^D) (15)

where S^G = [MCP ($/MWh) × power generated (MW)] – C(s) (16)

S^D = B(d) – [MCP ($/MWh) × power consumed (MW)] (17)

On the contrary, if network constraints are considered, even for a lossless system, congestion may arise due to any constraint violation. Then market clearing price no longer exists and the prices are different at all buses. Congestion may result in electricity price volatility and may lead to price spikes. System losses and congestion introduce *merchandising surplus* (S^M) or ISO surplus. For a lossless system, with congestion S^M may not be zero and can be either positive or negative. If the two effects, losses and congestion, are considered jointly, S^M is usually greater than zero. S^M can be adopted as a measure of congestion costs and is a reasonable metric to compare the congestion impact under different load elasticity conditions. The absolute value of S^M decreases with an increase in elasticity. In a lossless system, for infinite elasticity,

S^M is zero as in an unconstrained market [15]. The demand responsiveness can play a major role in competitive electricity markets, particularly in the case of congestion [16].

$$S.W = S^G + S^D + S^M \tag{18}$$

where S^G = [LMP ($/MWh) × power generated (MW)] – C(s) (19)

S^D = B(d) – [LMP ($/MWh) × power consumed (MW)] (20)

S^M = [LMP ($/MWh) × power consumed (MW)] – [LMP ($/MWh) × power generated (MW)] (21)

Case 2: With concentrated loss

The main objective is maximization of social welfare subjected to energy balance and line flow constraints. Generators' active power outputs and elastic customer demands are taken as variables for optimization. However, in nodal price based electricity markets, system marginal losses have significant impact on the economics of power system operation. So, system marginal losses have to be taken into account for obtaining accurate prices. In this model it is assumed that total system loss is supplied by slack bus generator. The problem is to

$$Maximize\ \ J = B(d) - C(s) \tag{22}$$

$$\text{s.t.}\ \sum_{i=1}^{N} DF_i \times (P_i) + P_{loss} = 0 \tag{23}$$

$$F_k \leq \text{limit}_k, \qquad \text{for k=1, 2 …M} \tag{24}$$

$$P_{Gi}^{\min} \leq P_{Gi} \leq P_{Gi}^{\max} \quad \text{for i=1, 2 …N} \tag{25}$$

$$P_{Di}^{\min} \leq P_{Di} \leq P_{Di}^{\max} \quad \text{for i=1, 2 …N} \tag{26}$$

where DF_i is the delivery factor at bus 'i'.

P_i is the real power injection at bus 'i'

P_{loss} is the total system loss used to offset doubled average system loss caused by the marginal loss factor (LF) and the marginal delivery factor (DF). The derivation of (23) is given in [23].

Case 3: With Distributed loss

If the load demand is in GW, then loss will be of the order of hundreds of MW, and it is not feasible to add the total system loss to slack bus as this may cause the slack bus power output to violate its maximum limit. It was observed that concentrated loss case for social welfare maximization problem is a major drawback and therefore distributed loss case is proposed for social welfare maximization in this paper. E_i is the extra load at bus 'i' (for each bus, the total extra load is the sum of half of line losses which are connected to that bus) and it is defined as follows:

$$E_i = \sum_{k=1}^{M_i} \frac{1}{2} \times F_k^2 \times R_k \tag{27}$$

where, M_i is number of lines connected to bus i.

The line flow F_k for this loss case is calculated as in (28).

$$F_k = \sum_{i=1}^{N} GSF_{k-i} \times (P_{Gi} - P_{Di} - E_i) \qquad (28)$$

where GSF_{k-i} is the generation shift factor of line 'k' to bus 'i'.

The algorithm for this problem is the same as in case 2. After getting power outputs of generators, slack bus power is calculated using (23) for both the cases 2 and 3 and supplier surplus can be calculated using (19). LMP's can be computed at all buses in the system using (11). With the obtained customer demands customer surplus can be calculated using (20). ISO surplus is calculated using (21). Then total social surplus is calculated using (18).

To check the correctness of the result, the obtained social welfare from (6) or (22) should be equal to the total social surplus obtained from (18). This is proved on all the test systems studied in this paper.

Algorithm for LMP calculation in double auction model

1. GA and BAT algorithms

Step 1: Read no. of buses, no. of lines, slack bus number, and Bus data. Read GA parameters like population size, chromosome length, no. of units, maximum no. of generations, elitism probability, crossover probability, mutation probability, and epsilon. Read a, b, c coefficients; min. and max. limits of generators. Read line data including line thermal limits.

Step 2: Generate randomly power generations of all generators except slack generator and decode them.

Step 3: Generate randomly power demands of all customers and decode them.

Step 4: Calculate Generation shift factors (GSF) using system 'X' matrix.

Step 5: Calculate initial line flows using GSF.

Step 6: Calculate the system loss i.e., P_{loss} in each line for cases2 and 3.

Step 7: Calculate the extra load at each bus 'i' using (27) from initial line flows for case 3, and then calculate new line flows using (28).

Step 8: Calculate loss factors and then delivery factors at each bus.

Step 9: Calculate P_{gen} of slack bus using (7) of case 1 or (23) of cases 2 and 3.

Step 10: Check for line flow limits (8). If the line limits are violated add penalties to objective function.

Step 11: Check for slack bus power limits (9). If it violates the limits add penalties to objective function.

Step 12: Calculate the social welfare with the randomly generated power generations and power demands using (6) for all cases and then calculate the fitness function = (objective function+ penalties).

Step 13: Sort the chromosomes in the descending order of fitness.

Step 14: Is iteration = max. no. of iterations. If yes go to step 20 else go to step 15.

Step 15: If fitness (1) == fitness (psize)→problem converged.

If no go to step 19.

Calculate the energy price of the reference bus either with fixed bids or with linear bids and then calculate LMPs' at all buses and the decomposition of LMP.

Step 16: Calculate ISO payment to generators ($\sum_i (power$ generation at bus i * LMP at bus i)) and then calculate suppliers surplus using (19).

Step 17: Calculate customers' payment to ISO ($\sum_j (power$ demand at bus j * LMP at bus j)) and then calculate customers surplus using (20).

Step 18: Calculate ISO profit using (21) for all the cases, calculate Social welfare using (18) & STOP.

Step 19: Use selection, crossover and mutation operators. Generate new population.
iteration = iteration +1; Go to step 4.

Step 20: STOP and print maximum number of iterations reached.

Using BA, the algorithm mathematically is similar to GA except the BA operators used. The detailed flowchart of LMP calculation with BA is given in next section.

2. Bat algorithm (BA) flow chart in double auction model

The flow chart of all the steps involved in LMP computation with Bat algorithm is presented in figure6.

Results and Discussion:

Genetic algorithm (GA) and Bat algorithm (BA) are used to solve the DCOPF for all the three loss cases with Social Welfare Maximization as objective function. BA is first time proposed for SWM problem in this paper. Both the algorithms GA and BA with linear bids for cases 1-3 are applied on IEEE 14bus system [24], New England 39 bus system [25] and Indian 75 bus system [26]. The solution reported is the best solution over 20 different runs for both GA and BA. BA gives better solution of social welfare (S.W) (i.e., maximum S.W value) than GA for all the three cases of all test systems. Supplier surplus, Customer surplus, Merchandising surplus and Social surplus obtained are positive values for all loss cases and for all test systems studied. Some of these values are negative in single auction model [23] which should be avoided in competitive electricity markets. This is one major drawback in single auction model which can be overcome in double auction model. LMPs' are also computed for all loss cases and all test systems while congestion is present in the system.

1. Case study 1: IEEE 14 bus system

The two algorithms GA and BA are tested initially on IEEE 14 bus system. This system consists of 2 generators and 11 loads. Both the generators and all the 11 loads are considered as suppliers and customers participating in market. It is observed that 9th line connecting 4-9 buses and 10th line connecting 5-6 buses are congested in all loss cases in both the methods. The shadow price or constraint cost is taken as 91.75 $/MWh for case 1 and 109.25 $/MWh for cases 2 and 3. These values are obtained from Power World Simulation. The generator power outputs and the corresponding social welfare are shown in table 1. The LMPs' are shown in table 2. Figure 7 presents a comparison of social welfare or social surplus for all the three loss cases. Figures 8, 9 and 10 show the LMPs' comparison for cases 1, 2 and 3. From the LMPs' comparison graphs of all the cases, it is observed that LMP at 5th bus is minimum and maximum at 6th bus. 10th line connecting 5th and 6th buses is congested which is a very crucial line in IEEE 14 bus system because this line divides the system into two zones. Therefore the impact of this line is high on the system. The LMPs' at both ends of 10th line have large difference and these LMPs' emerged as minimum and maximum of the system LMPs'. It can be observed that even though BA takes more iterations than GA, per iteration time of BA is less than GA and hence the CPU time with BA is less than with GA.

The minimum and maximum LMP's are indicated in red color in table 2. The highest LMP is reduced with Bat algorithm which leads to more customer surplus and less ISO surplus. From figure 7 it can be observed that for all the three cases, BA has given better social welfare than GA.

Table 1. Generators' active power outputs and social welfare for IEEE 14 bus system

Supplier bus No.	**Power Generations in MW without loss CASE 1**		**Power Generations in MW with concentrated loss CASE 2**		**Power Generations in MW with distributed loss CASE 3**	
	GA	BA	GA	BA	GA	BA
1(slack)	223.853	146.952	228.638	150.597	226.552	134.45
2	41.1438	114.988	41.1438	114.988	42.94	130.837
System loss MW)			4.785	3.6454	4.745	3.4984
Supplier surplus ($/hr)	4933.56	2958.32	5150.64	3134.05	5115.71	2489.33
Consumer surplus ($/hr)	37561.4	40198.8	35932	38787.7	36007.8	39315.2
Merchandising surplus or ISO surplus ($/hr)	5658.9	5586.63	6905.75	6737.79	6902.64	6685.34
Social surplus ($/hr)	48153.8	48743.7	47988.4	48659.5	48026.2	48489.9
Iterations	264	281	264	281	293	489
CPU time (sec)	48.1071	20.2850	48.1071	20.2850	36.4839	25.5068

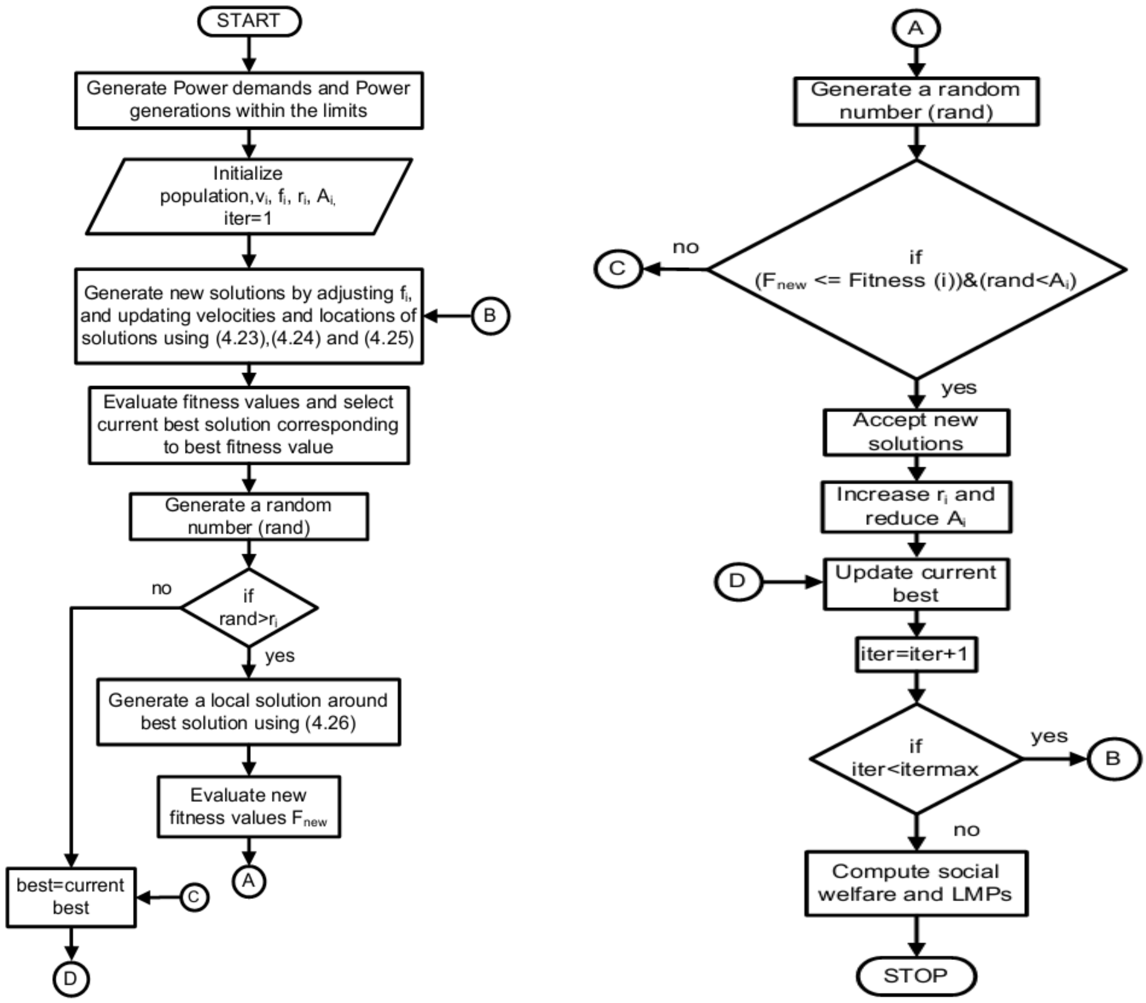


Figure 6: Flow chart of Bat algorithm for LMP calculation in double auction model

Table 2: Bus LMPs' for IEEE 14 bus system

Bus No.	LMPs' at all buses ($ / MWh)					
	Without loss CASE 1		**Concentrated loss CASE 2**		**Distributed loss CASE 3**	
	GA linear bids	BA linear bids	GA linear bids	BA linear bids	GA linear bids	BA linear bids
1	34.213	22.831	34.921	23.371	34.612	20.981
2	34.478	23.096	35.86	23.927	35.539	21.479
3	35.23	23.849	37.737	25.492	37.408	22.981
4	35.88	24.498	38.225	26.092	37.898	23.6
5	33.214	21.832	34.87	22.808	34.545	20.33
6	106.04	94.662	121.629	109.555	121.304	107.07
7	67.773	56.392	76.186	64.06	75.859	61.57
8	67.773	56.392	76.186	64.06	75.859	61.57
9	84.551	73.17	96.156	84.034	95.829	81.545
10	88.371	76.989	100.742	88.608	100.415	86.117
11	97.053	85.671	111.03	98.916	110.704	96.428
12	104.34	92.963	119.755	107.629	119.429	105.13
13	103.01	91.636	118.193	106.061	117.866	103.56
14	92.625	81.244	105.951	93.77	105.623	91.268

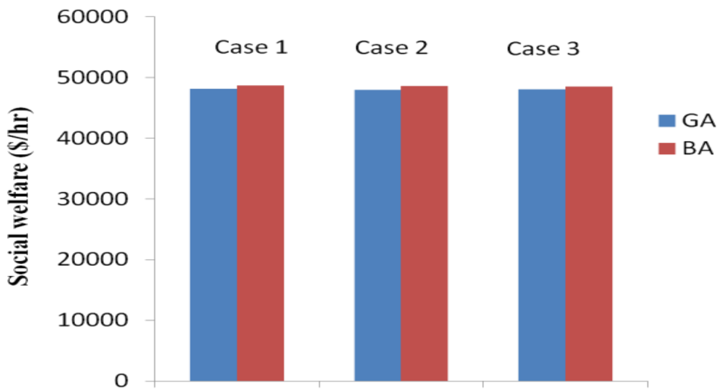


Figure 7: Social welfare comparison for all the three loss cases of IEEE 14 bus system

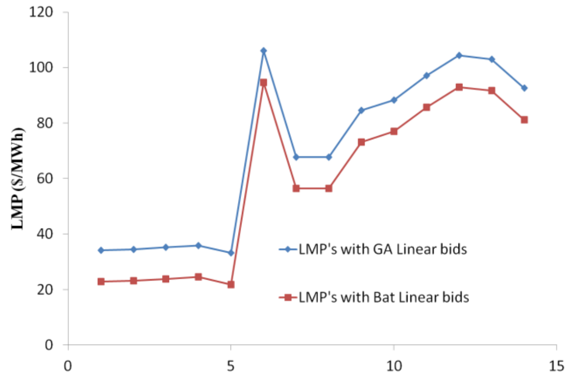


Figure 8: LMPs' comparison for case 1 of IEEE 14 bus system

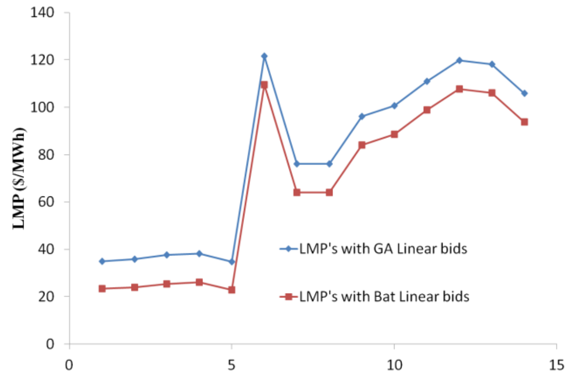


Figure 9: LMPs' comparison for case 2 of IEEE 14 bus system

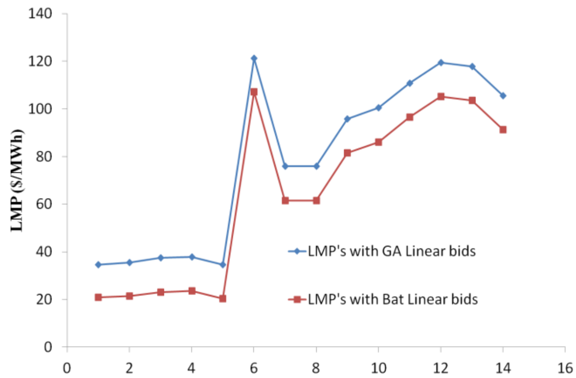


Figure 10: LMPs' comparison for case 3 of IEEE 14 bus system

2. Case study 2: New England 39 bus system

Both GA and BA are tested on New England 39 bus test system. This system has 10 generators and 19 loads. 4 generators are considered as suppliers and 4 major loads are considered as customers in the market. It was observed that only 46th line connecting 19–20 buses is congested for both methods in all loss cases. The shadow price or constraint cost is taken as 2.9 \$/MWh for all cases which is obtained from Power World Simulation. The generator power outputs and the corresponding social welfare are shown in table 3. The LMPs' are shown in table 4. Figure 11 presents a comparison of social welfare or social surplus for all the three loss cases. Figures 12, 13 and 14 show the LMPs' comparison for cases 1, 2 and 3.

From figure 11 it can be observed that for all the three cases, BA has given better social welfare than GA. From the LMPs' comparison graphs, it can be observed that for case 1 all the LMPs' are same except at 20th and 34th buses. These are maximum of all and are indicated in red color in table 4. As the congested line connects 19th and 20th buses, LMP at 20th bus (i.e., sink node of the congested line) is raised to maximum value. 40th line connects 20th bus and 34th bus. As 34th bus is a dead end bus, LMP of 20th bus is reflected at 34th bus. For cases 2 and 3 also, LMPs' are maximum at 20th and 34th buses, but for remaining buses LMPs' vary slightly due to system losses. It can be observed that in table 3, even though BA takes more iterations than GA, per iteration time of BA is less than GA and hence the CPU time with BA is reduced than with GA.

Table 3: Generators' active power outputs and social welfare for New England 39 bus system

Supplier bus No.	Power Generations in MW without loss CASE 1		Power Generations in MW with concentrated loss CASE 2		Power Generations in MW with distributed loss CASE 3	
	GA	BA	GA	BA	GA	BA
31(slack)	235.046	248.21	272.77	285.12	272.7	280.10
32	650.99	650.17	650.99	650.17	650.98	645.29
33	632.99	625.89	632.99	625.89	632.98	642.22
35	650.91	641.64	650.91	641.64	650.98	629.78
System loss (MW)			37.73	36.9	37.73	36.82
Supplier surplus (\$/hr)	696.91	913.98	542.143	765.79	540.68	739.481
Consumer surplus (\$/hr)	28630.7	28418.6	28013.5	27806.1	28014.5	27827.7
Merchandising surplus or ISO surplus (\$/hr)	1507.97	1507.17	1896.11	1889.56	1896.14	1885.62
Social surplus (\$/hr)	30835.6	30839.8	30451.7	30461.5	30451.4	30452.8
Iterations	112	186	112	186	137	214
CPU time (sec)	83.689	30.3429	83.689	30.3429	148.868	145.316

Table 4: Bus LMPs' for New England 39 bus system

Bus No.	LMPs' at all buses ($ / MWh)					
	Without loss model CASE 1		**Concentrated loss model CASE 2**		**Distributed loss model CASE 3**	
	GA linear bids	BA linear bids	GA linear bids	BA linear bids	GA linear bids	BA linear bids
1	10.06	10.139	10.543	10.621	10.543	10.586
2	10.06	10.139	10.279	10.357	10.279	10.325
3	10.06	10.139	10.224	10.302	10.223	10.27
4	10.06	10.139	10.291	10.367	10.29	10.336
5	10.06	10.139	10.306	10.38	10.305	10.35
6	10.06	10.139	10.286	10.36	10.286	10.33
7	10.06	10.139	10.364	10.439	10.364	10.409
8	10.06	10.139	10.404	10.479	10.403	10.448
9	10.06	10.139	10.581	10.658	10.581	10.624
10	10.06	10.139	10.149	10.224	10.149	10.194
11	10.06	10.139	10.194	10.269	10.194	10.239
12	10.06	10.139	10.178	10.253	10.177	10.223
13	10.06	10.139	10.162	10.237	10.161	10.207
14	10.06	10.139	10.184	10.26	10.183	10.229
15	10.06	10.139	10.05	10.129	10.05	10.098
16	10.06	10.139	9.992	10.073	9.992	10.041
17	10.06	10.139	10.078	10.157	10.077	10.125
18	10.06	10.139	10.133	10.212	10.133	10.18
19	10.06	10.139	9.947	10.03	9.947	9.997
20	12.96	13.039	12.847	12.93	12.847	12.897
21	10.06	10.139	9.875	9.957	9.875	9.925
22	10.06	10.139	9.754	9.836	9.753	9.804
23	10.06	10.139	9.799	9.881	9.799	9.849
24	10.06	10.139	9.965	10.045	9.964	10.013
25	10.06	10.139	10.255	10.334	10.255	10.301
26	10.06	10.139	10.166	10.245	10.166	10.213
27	10.06	10.139	10.125	10.205	10.125	10.172
28	10.06	10.139	10.166	10.245	10.166	10.213
29	10.06	10.139	10.166	10.245	10.166	10.213
30	10.06	10.139	10.279	10.357	10.279	10.325
31	10.06	10.139	10.286	10.36	10.286	10.33
32	10.06	10.139	10.149	10.242	10.149	10.194
33	10.06	10.139	9.947	10.03	9.947	9.997
34	12.96	13.039	12.847	12.93	12.847	12.897
35	10.06	10.139	9.754	9.836	9.753	9.804
36	10.06	10.139	9.799	9.881	9.799	9.849
37	10.06	10.139	10.255	10.334	10.255	10.301
38	10.06	10.139	10.166	10.254	10.166	10.213
39	10.06	10.139	10.704	10.781	10.703	10.745

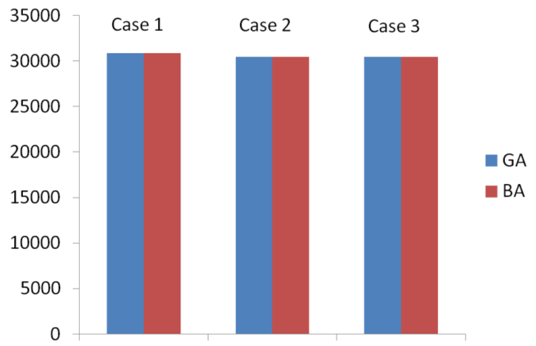


Figure 11: Social welfare comparison for all the three loss cases of New England 39 bus system

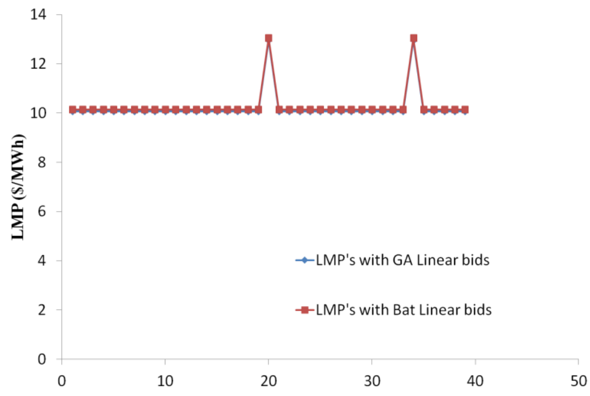


Figure 12: LMPs' comparison for case 1 of New England 39 bus system

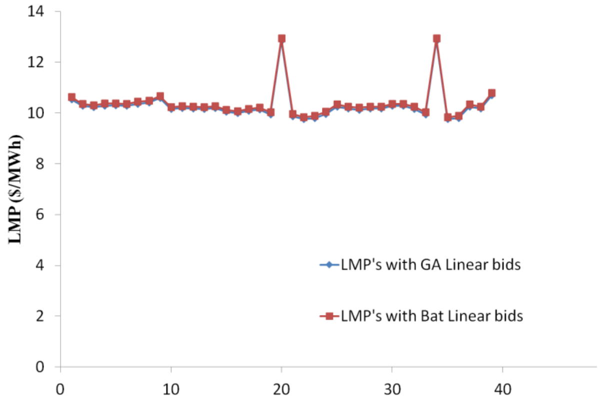


Figure 13: LMPs' comparison for case 2 of New England 39 bus system

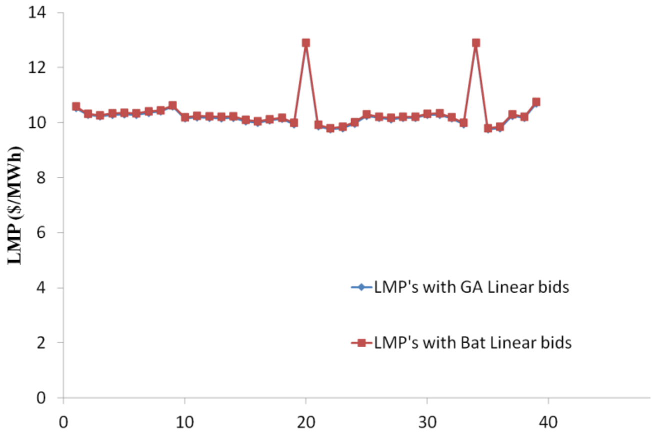


Figure 14: LMPs' comparison for case 3 of New England 39 bus system

3. Case study 3: Indian 75 bus system

GA and BA are tested on Indian 75 bus system also. This system has 15 generators and 40 loads. 10 generators are considered as suppliers and 9 major loads are considered as customers. All the remaining generators and loads are assumed to be engaged in bilateral transactions. It was observed that with GA for cases 1 and 2 line no. 9 (28-4), 11 (31-5) and 86 (25-72) are congested but for case 3 lines 9, 11, 39 (25-60), 43 (28-43) and 86 are congested. With BA for cases 1, 2 and 3 lines 9 and 86 are congested. The shadow price or constraint cost is taken as 3.35 $/MWh for case 1, 3.67 $/MWh for case 2 and 3.7 $/MWh for case 3 which are obtained from Power World Simulation. The generator power outputs and the corresponding social welfare are shown in table 5. Figure 15 presents a comparison of social welfare or social surplus for all the three loss cases. Figures 16, 17 and 18 show the LMPs' comparison for cases 1, 2 and 3.

From table 5 it is observed that the social welfare is improved with BA than with GA and also the CPU time of BA is less than GA. Even though BA takes more iterations than GA, per iteration time of BA is less than GA and hence the CPU time with BA is reduced than with GA.

From figures 16, 17, and 18, it is observed that, the LMP values at some buses (buses belong to congested lines and 70th bus) are raised to higher values, because of congestion in the system. For remaining buses LMP values vary slightly due to presence of losses in cases 2 and 3. Though 70th bus does not belongs to congested lines, the LMP at this bus is raised to a higher value because this bus is a dead end bus and is connected after bus 72 (which is belongs to congested line), so the LMP of 72nd bus is reflected at 70th bus.

Table 5: Generators' active power outputs and social welfare of Indian 75 bus system

Supplier bus No.	**Power Generations in MW without loss CASE 1**		**Power Generations in MW with concentrated loss CASE 2**		**Power Generations in MW with distributed loss CASE 3**	
	GA	BA	GA	BA	GA	BA
1(slack)	365.95	258.75	429.29	314.95	383.72	276.57
2	359.99	350.15	359.99	350.15	358.99	359.33
3	270.00	273.22	270.00	273.22	279.68	270.76
4	199.99	199.91	199.99	199.91	199.99	198.67
5	279.99	272.57	279.99	272.57	279.99	273.71
6	210.00	216.19	210.00	216.19	210	218.64
7	150.00	153.6	150.00	153.6	150.00	156.76
8	170.00	174.98	170.00	174.98	179.99	171.14
9	649.99	648.86	649.99	648.86	649.97	641.32
10	170.00	172.02	170.00	172.02	179.76	173.39
System loss (MW)			63.335	56.198	62.17	54.143
Supplier surplus ($/hr)	2403.16	3187.45	4585.53	5047.88	6686.13	4289.62
Consumer surplus ($/hr)	6866.84	10320.5	2577.56	6701.9	1986.52	9088.36
Merchandising surplus or ISO surplus ($/hr)	1971.85	248.05	3065.15	1143.64	1821.14	1106.01
Social surplus ($/hr)	11241.9	13756	10228.2	12893.4	10493.8	14484
Iterations	459	639	459	639	384	478
CPU time (sec)	2233.7	1153.79	2233.7	1153.79	1913.70	674.053

Customers who are involved in bidding will pay to ISO at the rate of its bus LMP for the consumed power. Suppliers who are involved in auction will be paid by ISO at the rate of its bus LMP for supplying power to the grid. Trading will be done for remaining generators and loads based on bilateral transactions. It can be observed from the case studies that slack bus power in distributed loss case is reduced than that in concentrated loss case and burden on the slack generator is removed. This is the main reason to propose the distributed loss case for SWM problem. Even though BA takes more iterations to converge than GA, the CPU time taken per iteration is very less in BA compared to GA as BA has less number of parameters to adjust and offers a better result with less population size. Bat algorithm works with small population of the order of 10-25 size. Therefore with BA, the total computational time for the convergence is very much reduced than with GA for all cases for all test systems studied. Hence, Bat algorithm with distributed loss case can be a best algorithm for SWM in double auction markets.

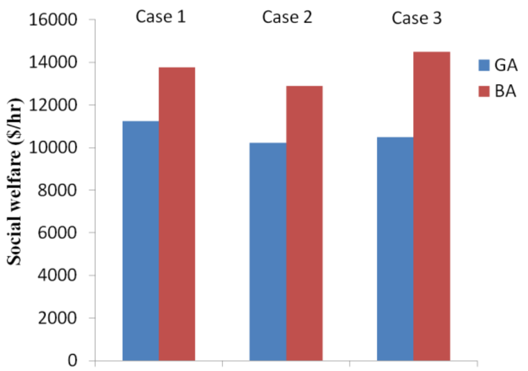


Figure 15: Social welfare comparison for all the three loss cases of Indian 75 bus system

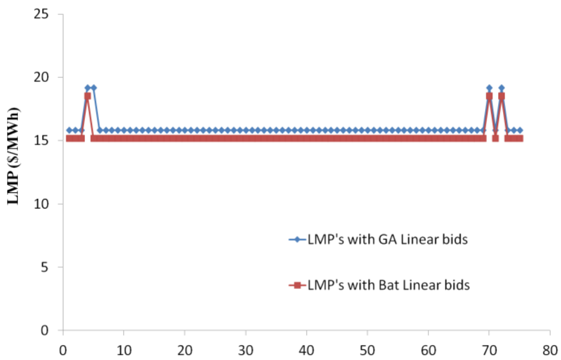


Figure 16: LMPs' comparison for case 1 of Indian 75 bus system

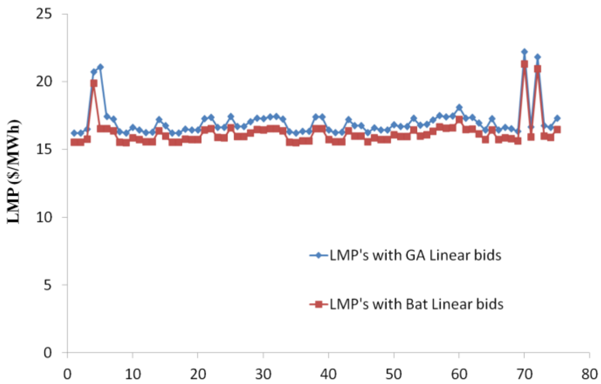


Figure 17: LMPs' comparison for case 2 of Indian 75 bus system

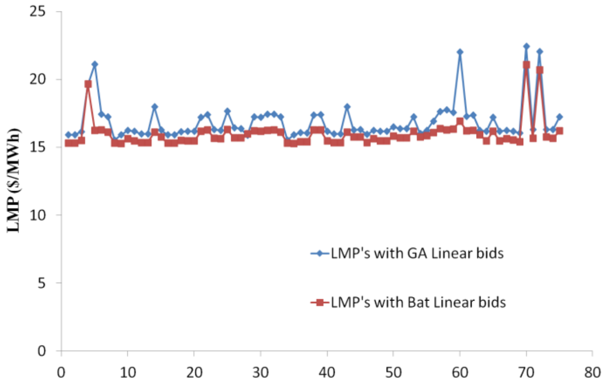


Figure 18: LMPs' comparison for case 3 of Indian 75 bus system

Conclusion:

In this paper LMP calculation with Social Welfare Maximization as objective function with three different loss cases for congested system is attempted to allocate variable cost to the users of the grid. This problem is solved by DCOPF with GA and BA with generator linear bids. Consumers are modelled as elastic and hence consumer surplus exists. Only major consumers participate in bidding and they have choice to alter their demand according to the price of the suppliers. From the results, it is observed that BA produces better results compared to GA in terms of social welfare and computation time in all the loss cases for all the test systems studied. All the surpluses (e.g., supplier surplus, consumer surplus, ISO surplus and social surplus) are positive with both GA and BA in all the loss cases for all the test systems used.

Appendix:

Generation Shift Factor

Generation shift factor is the ratio of change in power flow of line 'k' to change in power injection at bus 'i'. GSF can be computed using (i), where B^{-1} is the inverse of B matrix, x_k is reactance of line k, 'a' and 'b' are sending and receiving end buses of line k.

$$GSF_{k-i} = (B^{-1}_{(a,i)} - B^{-1}_{(b,i)}) / x_k \qquad \text{(i)}$$

Delivery Factor

The delivery factor (DF_i) at the i^{th} bus represents the effective MW delivered to the customers to serve the load at that bus. It is defined as (ii)

$$DF_i = 1 - LF_i = 1 - \partial P_{loss} / \partial P_i \qquad \text{(ii)}$$

$$P_{loss} = \sum_{k=1}^{M} F_k^2 \times R_k \qquad \text{(iii)}$$

$$F_k = \sum_{i=1}^{N} GSF_{k-i} \times P_i \qquad \text{(iv)}$$

$$\frac{\partial P_{loss}}{\partial P_i} = \sum_{k=1}^{M} \frac{\partial}{\partial P_i}(F_k^2 \times R_k)$$

$$= \sum_{k=1}^{M} R_k \times 2F_k \times \frac{\partial F_k}{\partial P_i}$$

$$= \sum_{k=1}^{M} 2 \times R_k \times GSF_{k-i} \times (\sum_{j=1}^{N} GSF_{k-j} \times P_j)$$

In (ii), LF_i represents the loss factor at bus 'i' which is calculated using (v), F_k is the power flow in line k , R_k is the resistance of line k, P_i is the injected power at bus i,GSF_{k-i} is the generation shift factor to line 'k' from bus 'i'. LF_i may be viewed as the change of total system loss with respect to 1 MW increase in injection at that bus. Interestingly, the loss factor at a bus may be positive or negative. When it is positive, it implies that an increase of injection at the bus may increase the total system loss. If it is negative, it implies that an increase of injection at that bus reduce the total loss.

References:

1. H. Singh, S. Hao, and A. Papalexopoulos, "Transmission congestion management in competitive electricity markets," IEEE Trans. Power Syst., vol. 13, No.2, pp. 672 –680, May 1998.
2. Christie R.D., Wollenberg B.F., Wangensteen I., "Transmission management in the deregulated environment", Proc. of IEEE, 88, (2), pp. 170 – 195, 2000.
3. Luonan chen, Hideki Suzuki, Tsunehisa Wachi, Yukihiro Shimura, "Components of Nodal prices for Electric Power Systems", IEEE Trans. Power. Syst., vol.17, No.1, February 2002.
4. G. Hamoud, I. Bradley, "Assessment of transmission congestion cost and Locational Marginal Pricing in a competitive electricity market", IEEE Trans. on Power Syst., Vol. 19, no.2, May 2004.
5. Thomas J. Overbye, Xu Cheng, Yan Sun, "A comparison of the AC and DC power flow models for LMP calculations", IEEE 37th Annual Hawaii International Conference on System Sciences, 2004.
6. Yuan-Kang, Wu, "Comparison of pricing schemes of several deregulated electricity markets in the world", IEEE/PES Transmission and Distribution Conference, 2005.
7. Yong Fu, Zuyi Li, "Different models and properties on LMP calculations", IEEE Power Engineering Society General Meeting, 2006.
8. Haifeng Liu, Leigh Tesfatsion, A.A.Chowdhury, "Locational Marginal Pricing basics for Restructured wholesale power markets", IEEE Power & Energy Society General Meeting, 2009.
9. Fangxing Li, Rui Bo, "DCOPF-based LMP Simulation: Algorithm, Comparison with ACOPF, and Sensitivity", IEEE Trans. Power Syst., Vol.22, no.4, November 2007.
10. E. Litvinov, "Design and operation of the Locational marginal prices-based electricity markets" IET Gener. Transm. Distrib., Vol. 4, no. 2, pp. 315-323, 2010.
11. Hongyan Li, Leigh Tesfatsion, "ISO Net Surplus Collection and Allocation in Wholesale Power Markets Under LMP", IEEE Trans. on Power Systems, Vol. 26, No. 2, MAY 2011.
12. Rajesh Rajaraman, Jayant V.Sarlashkar, Fernando L.Alvarado, "The effect of demand elasticity on security prices for the POOLCO and Multi-lateral contract models", IEEE Trans. on Power Systems, Vol. 12, no. 3, August 1997.

13. J.D.Weber, T.J.Overbye, "A Two-Level Optimization Problem for Analysis of Market Bidding Strategies", IEEE Power Engineering Society Summer Meeting, 1999.
14. James D. Weber, Thomas J. Overbye, "An Individual Welfare Maximization Algorithm for Electricity Markets", IEEE Trans. on Power Systems, Vol. 17, no. 3, August 2002.
15. Ettore Bompard, Enrico Carpaneto, Gianfianco Chicco, George Gross, "The role of load demand elasticity in congestion management and pricing", IEEE Power Engineering Society Summer Meeting, 2000.
16. Paolo Marannino, Riccardo Vailati, Fabio Zanellini, Ettore Bompard, George Gross, "OPF tools for optimal pricing and congestion management in a two sided auction market structure", IEEE Power Tech Proceedings, 2001.
17. Surachai Chaitusaney, Bundhit Eua-Arporn, "Actual Social Welfare Maximization in Pool Market", IEEE Power Engineering Society Summer Meeting, 2002.
18. Francisco D. Galiana, Alexis L. Motto, and Francois Bouffard, "Reconciling Social Welfare, Agent Profits, and Consumer Payments in Electricity Pools", IEEE Trans. on Power Systems, Vol. 18, No. 2, May 2003.
19. Jiang Wu, Xiaohong Guan, Feng Gao, Guoji Sun, "Social Welfare Maximization Auction for Electricity Markets with Elastic Demand", IEEE 7th World Congress on Intelligent Control and Automation, 2008.
20. Hongrui Liu, Yanfang Shen, Zelda B. Zabinsky, Chen-Ching Liu, Alan Courts, Sung-Kwan Joo, "Social Welfare Maximization in Transmission Enhancement Considering Network Congestion", IEEE Trans. on Power Systems, Vol. 23, no. 3, August 2008.
21. Xiaoyan Zou, "Double-sided auction mechanism design in electricity based on maximizing social welfare", Energy Policy, Elsevier, Vol.37 (2009) 4231–4239.
22. Kanwardeep Singh, Narayana Prasad Padhy, Jaydev Sharma, "Influence of Price Responsive Demand Shifting Bidding on Congestion and LMP in Pool-Based Day- Ahead Electricity Markets", IEEE Trans. on Power Systems, Vol. 26, no. 2, May 2011.
23. M. Murali, M. Sailaja Kumari, M. Sydulu, "Optimal spot pricing in electricity market with inelastic load using constrained bat algorithm", Journal of Electrical Power and Energy Systems, Elsevier, Vol.62, pp. 897-911, Nov 2014.
24. Iyambo P.K., Tzoneva R., Transient Stability Analysis of the IEEE 14-Bus Electric Power System, IEEE conference, Africon, 2007, (14 bus data).
25. New England 39 bus system data (online), Available at:
26. http://sys.elec.kitami-it.ac.jp/ueda/demo/WebPF/39-New-England.pdf, (12/02/2001)
27. Indian75 bus Power system data, Uttar Pradesh State (A state of India) Electricity Board data". (This data is obtained from Indian Institute of Technology, Kanpur)

REVIEW ON APPLICATION PROGRAMMING INTERFACES FOR INTERNET PROTOCOL

Sumit Chopra*, Lovedeep Kaur, Anchal Nayyar and Gagandeep Singh

GNA University,

Phagwara, Punjab

*Corresponding author E-mail: sumit.chopra@gnauniversity.edu.in

Abstract:

The Internet Protocol is a fundamental protocol used for communication and data transfer over the internet. Application Programming Interfaces play a critical role in enabling interactions between different components of the IP ecosystem, including routers, switches, and other networking devices. This research paper presents a comprehensive analysis of APIs used in Internet Protocol. The study starts by providing an overview of the Internet Protocol and its architecture. It provides the basic knowledge of various API's used for internet protocol such as Socket API, Network API, IP geolocation API, IP address lookup API. This paper presents a comparative analysis of these APIs, evaluating their performance in terms of scalability, efficiency, and reliability.

Keywords: Application Programming Interface, Internet Protocol,

Introduction:

API stands for Application Programming Interface. It consists of routines, protocols, and tools for developing software applications. APIs specify how software components ought to communicate and interact with one another. Software applications that integrate with other systems and services require APIs. By utilizing the functionality and data that are already present in other applications, they enable programmers to develop brand-new applications. And that implies designers can assemble applications quicker and more effectively by giving prefabricated usefulness that can be incorporated to their application [1].

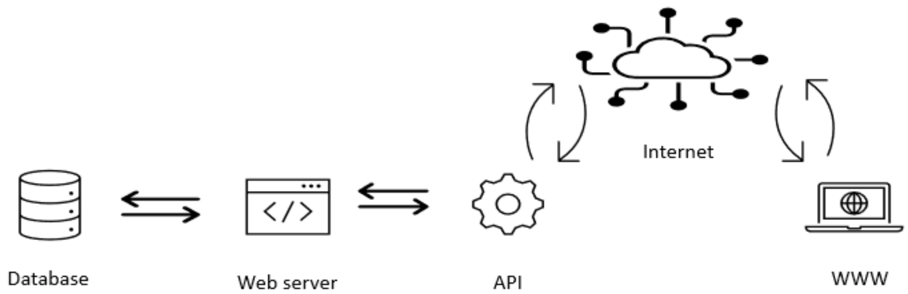


Figure 1: Application Programming Interface

The client app sends an API request to the API server in the diagram below. The Programming interface server processes the solicitation and sends back a Programming interface reaction. The client application then, at that point, gets the Programming interface reaction and can utilize the information returned in the reaction to play out an activity or show data to the client. APIs can be utilized for a wide range of tasks, including database data retrieval, calculation, and integration with third-party services. Businesses can monetize their APIs by offering them to third-party developers in the API economy, a new business model created with APIs [2].

```
+--------------------+          +--------------------+
|                    |          |                    |
|     Client App     |          |     API Server     |
|                    |          |                    |
+--------------------+          +--------------------+
          |                                |
          |                                |
          |          API Request           |
          +------------------------------->|
          |                                |
          |          API Response          |
          |<-------------------------------+
          |                                |
```

Figure 2: Working of APIs

Basics of Internet Protocol

Internet Protocol (IP) is a bunch of decides and techniques that administer the correspondence of information parcels over the web. Data packets are routed between internet-connected devices by this fundamental protocol in the internet protocol suite.

IP provides a unique address to each device which is connected to the internet, known as an IP address. Additionally, IP specifies the rules for transmitting data packets over the internet and their format.

There are currently two IP versions in use: Both IPv4 and IPv6 IPv4 utilizes a 32-digit address design and can uphold up to 4.3 billion special locations. The address format of IPv6 is 128 bits, making it capable of supporting an almost infinite number of distinct addresses.

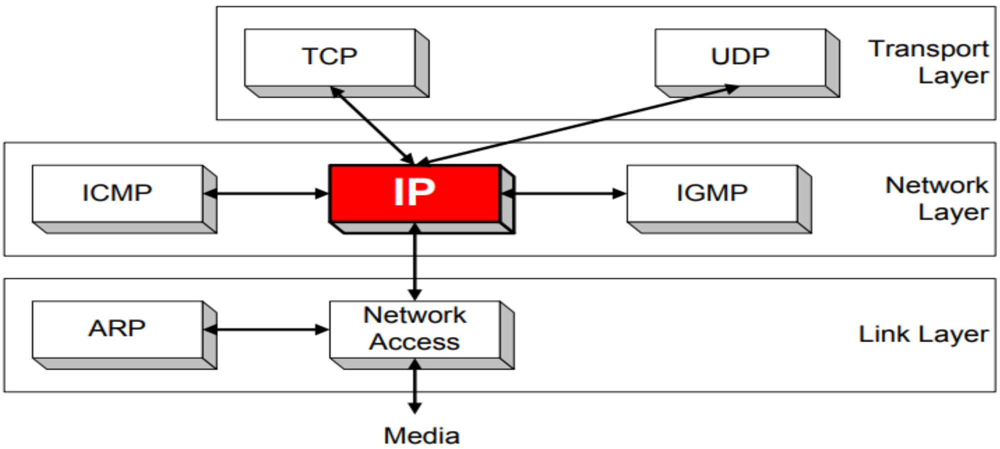


Figure 3: Internet Protocol in Network Layer

API for Internet Protocol

APIs can be categorized into different types, such as web APIs, operating system APIs, library APIs, and more. Web APIs are the most common type of API and are used to communicate with web-based applications, such as social media platforms, online marketplaces, and more.

The Socket API is the Internet Protocol (IP) API that is utilized the most frequently. Using the TCP/IP protocol suite, this API provides a set of functions that enable applications to create, send, and receive data over the Internet. The Attachment Programming interface is accessible in most programming dialects and working frameworks, including C, C++, Java, Python, and .NET. It is a low-level API that can be used to implement a wide range of network applications, including web browsers, email clients, and file transfer programs. It gives direct access to the underlying network protocols [3].

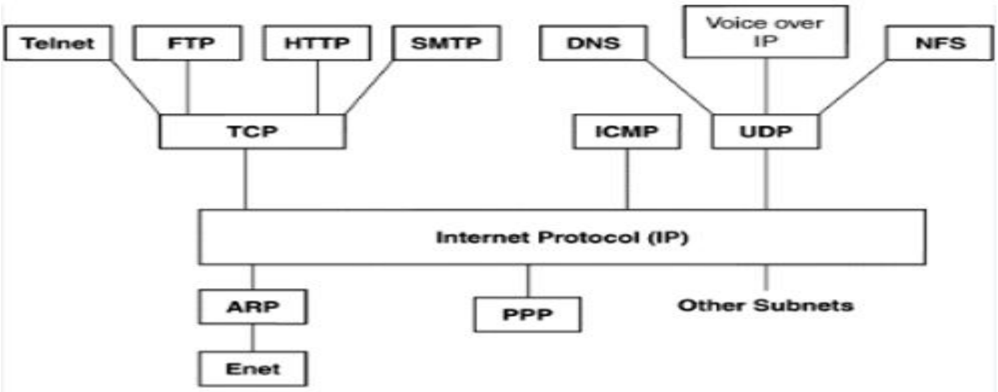


Figure 4: Different APIs used in Internet Protocol

There are several APIs available for Internet Protocol (IP), which is a protocol used for communication between devices on a network. Some of the commonly used APIs for IP include:

Socket API

Socket API, also known as Berkeley Sockets, is a low-level programming interface for network communication that provides a standardized way for software applications to communicate over a network. It was originally developed for Unix-based operating systems, but it is now available on a wide range of platforms. Applications can use the Socket API's set of functions to create, connect, send, and receive data over a network as shown in Figure 5.

These functions are typically implemented using TCP or UDP protocols, which are the most used transport protocols for network communication. The Socket API is based on the concept of sockets, which are endpoints for communication between two processes over a network. Each socket is identified by a unique IP address and port number, which allows data to be sent and received between the two processes. The Socket API provides a flexible and powerful mechanism for network communication, allowing applications to implement a wide range of protocols and communication patterns. It is frequently used in client-server applications like email clients, file transfer programs, and web servers [4].

In the client-server paradigm, we need to define local host, local process, remote host, and remote process in order to achieve process communication. Additionally, we require a second identifier known as a port number. The integers that make up a port number range from 0 to 65535. The ephemeral port number serves as the client program's self-definition. This term is used because clients typically only have a short lifespan. The IP address is what distinguishes the local host from the remote host. To establish a connection, the socket address contains both the IP address and the port number.

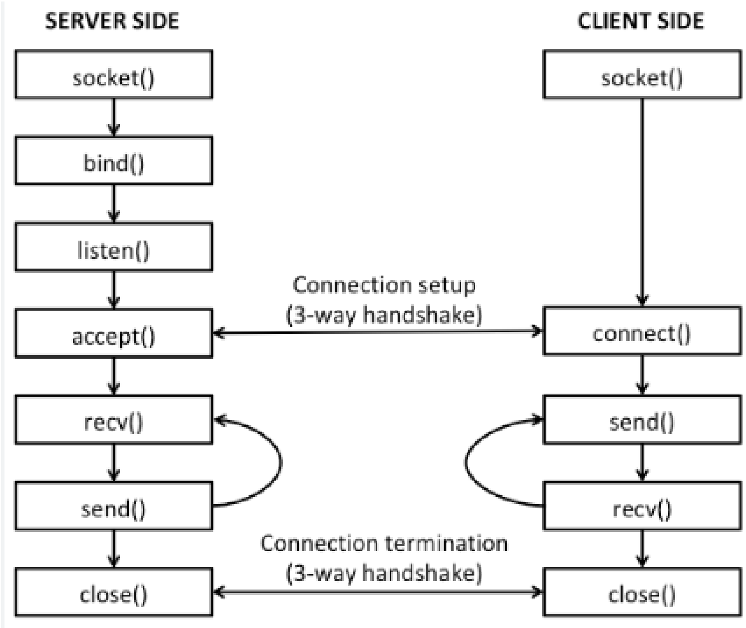


Figure 5: Working of Socket APIs

Transmission Control Protocol

TCP (Transmission Control Protocol) is a broadly involved convention in PC organizing. It is one of the core protocols that make up the Internet Protocol (IP) suite and oversees how computers on the Internet communicate with each other. TCP gives a dependable, association situated correspondence administration between two applications on an organization.

The three-way handshake is how TCP establishes a connection between two endpoints, also known as sockets. Data is sent between the two endpoints in packets called segments once the connection is established [5]. TCP gives a few components to guarantee that these portions are dependably conveyed and aligned correctly. Checksums for error detection, acknowledgments for data receipt confirmation, and retransmissions for lost data are examples of these mechanisms. The 3-way handshake is a process that establishes a reliable connection between two devices using the Transmission Control Protocol (TCP). It is a critical step in establishing a connection before any data is transmitted between the devices. The 3-way handshake process involves the following steps:

Step 1: SYN The process begins with the client, the first device, sending a SYN (synchronize) packet to the server, the second device, or client. This parcel contains an irregular grouping number that is utilized to synchronize the succession quantities of the gadgets. A flag in the SYN packet also indicates that the device wants to establish a connection.

Step 2: SYN-ACK The second device responds with a SYN-ACK (synchronize-acknowledge) packet after receiving the SYN packet. A random sequence number and an acknowledgment of the client's sequence number are included in this packet. A flag in the SYN-ACK packet also indicates that the device is prepared to establish a connection and has acknowledged the client's request.

Step 3: ACK The first device sends an acknowledgement packet back to the server after receiving the SYN-ACK packet. The server's sequence number is acknowledged in this packet. At this point, the connection is established, and both devices can begin to transmit data over the connection.

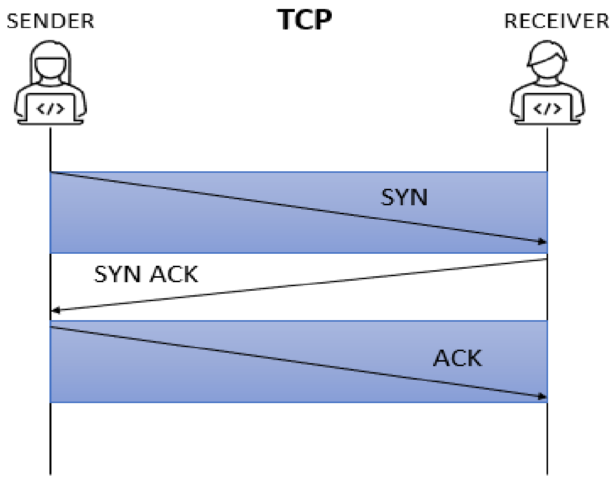


Figure 6: Working of Transmission Control Protocol

It is important to note that the 3-way handshake is a reliable process, meaning that it ensures that both devices are aware of each other's sequence numbers and that they are in sync. This helps to prevent data loss or corruption during transmission, as both devices can keep track of the data that is sent and received.

Figure 6 shows that the client sends the server a SYN (synchronize) packet to start the connection. To acknowledge the request and synchronize sequence numbers for data transmission, the server sends a SYN-ACK (synchronize-acknowledge) packet response. To conclude the handshake and ensure that the sequence numbers are in sync, the client sends an ACK (acknowledgement) packet. This makes it possible for data to be transferred between the two devices.

User Datagram Protocol

The transport layer protocol known as UDP, or User Datagram Protocol, is extensively utilized in computer networking. It transmits data without first establishing a connection because it is a connectionless protocol. Instead, it sends datagrams, or individual packets, to their destination without guaranteeing their delivery or sequence.

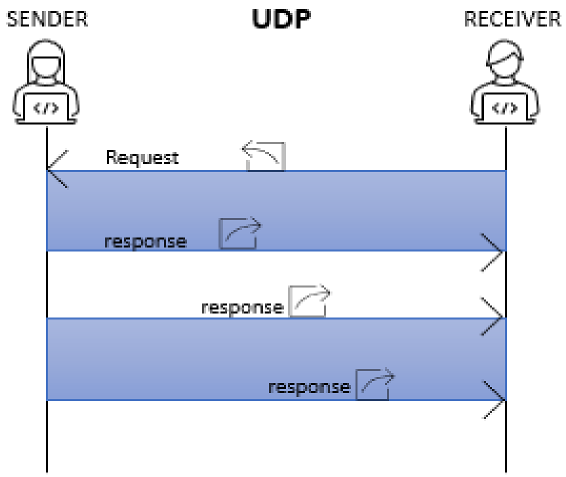


Figure 7: Working of User Datagram Protocol

UDP provides a lightweight and fast alternative to TCP (Transmission Control Protocol), another popular transport layer protocol. Unlike TCP, which guarantees reliable data delivery by establishing a connection and performing error-checking, UDP sends data packets without any guarantee of delivery or sequencing. This makes UDP ideal for applications that can tolerate some loss or duplication of data, such as real-time communication, streaming media, and online gaming [6].

UDP is a straightforward protocol that requires fewer computational resources than TCP and is simple to implement. It has a more modest bundle header than TCP, which decreases how much information communicated over the organization. For high-throughput applications, this makes UDP more efficient. However, UDP has a few drawbacks. Since it gives no assurance of conveyance or sequencing, not appropriate for applications require dependable information move, for example, document move or email. Furthermore, it has no blockage control instruments, which can prompt organization clog and corrupted execution. Additionally, it is susceptible to DoS (Denial of Service) attacks, in which an adversary can flood the network with UDP packets, resulting in service disruption and congestion.

Applications that necessitate high throughput and low latency, such as real-time communication, streaming media, and online gaming, frequently employ UDP. Additionally, it is utilized in DHCP (Dynamic Host Configuration Protocol) and DNS (Domain Name System) protocols that do not require dependable data transfer.

The Socket API is a low-level API that provides a set of functions for creating and managing sockets. Sockets are endpoints for communication between devices connected to the internet. The Socket API is widely used for developing network applications [7].

Hypertext Transfer Protocol API

HTTP (Hypertext Transfer Protocol) is a convention utilized for correspondence on the Internet. A Programming interface (Application Programming Connection point) is a bunch of conventions, schedules, and instruments for building programming applications. A HTTP Programming interface, thusly, is a Programming interface that involves HTTP as its convention for correspondence between two programming applications.

A set of rules and protocols for how two applications use the HTTP protocol to communicate is known as an HTTP API. The API defines a set of endpoints, or URLs, that represent specific actions that can be performed by the API. Each endpoint typically represents a specific resource or object that can be manipulated or retrieved by the client application [8]. When a client application wants to use a HTTP API, it sends a request to the API using the HTTP protocol. The request contains information about the action the client wants to perform, such as retrieving data, creating a new resource, or updating an existing resource.

The request typically includes a method, such as GET, POST, PUT, or DELETE, which specifies the type of action to perform. Once the API receives the request, it processes it and response back to the client application. The response typically includes a status code, such as 200 OK or 404 Not Found, which indicates whether the request was successful or not. It also includes any data or information requested by the client application.

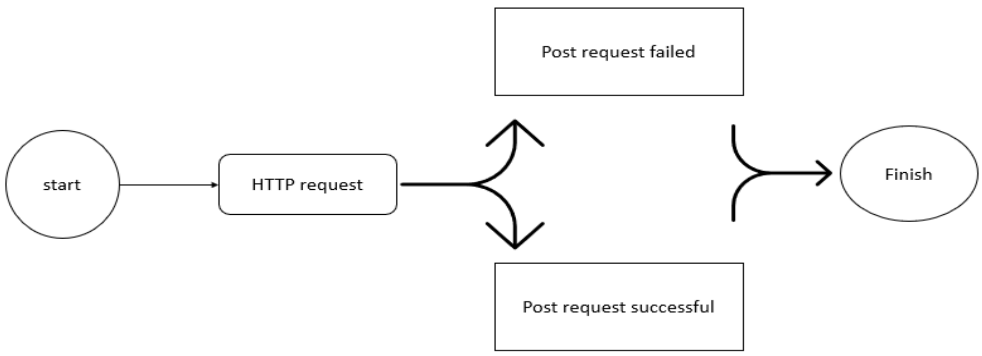


Figure 8: Hypertext Transfer Protocol

HTTP APIs can be used to enable integration between different software applications or to provide a service to external developers or clients. They are commonly used in web development, mobile application development, and in the development of other software applications that require communication over the internet.

HTTP APIs are typically documented using a specification like OpenAPI or Swagger, which describes the available endpoints, their parameters, and the expected responses. This allows developers to easily understand how to interact with the API and build applications that use it. Additionally, numerous HTTP APIs are tied down utilizing confirmation and approval instruments to guarantee that main approved clients can get to the Programming interface and its assets.

Representational State Transfer API

The architectural style known as REST (Representational State Transfer) is used to construct distributed systems and makes use of the HTTP communication protocol. Using the REST architectural style, a REST API (Application Programming Interface) is a set of rules and protocols that control how a client application can interact with a server application over the internet. A set of resources, or objects, that the client application can manipulate, or retrieve are defined by a REST API. A unique URL, also known as a resource identifier or URI (Uniform Resource Identifier), is used to identify each resource [9]. REST APIs use HTTP methods like GET, POST, PUT, and DELETE to perform operations on resources. For instance, a REST API for managing books might have a resource for retrieving all books at the URI /books and a resource for retrieving a particular book at the URI /books/id. Every technique compares to an alternate activity that can be performed on an asset, for example, recovering information, making another asset, refreshing a current asset, or erasing an asset.

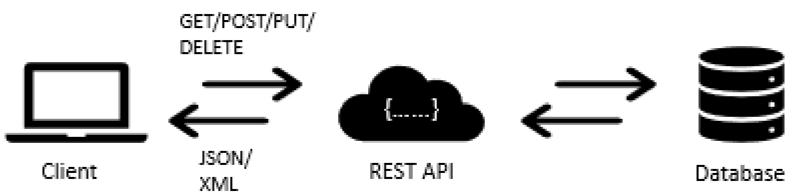


Figure 9: Working of REST API

To use a REST API, a client application sends an HTTP request to the API server, specifying the appropriate method and URI for the desired resource. The request may also include parameters, such as query strings or request bodies, that provide additional information about the desired operation. For example, a GET request to the /books resource might include a query string parameter to specify the number of results to return.

Since REST APIs are designed to be stateless, each request from a client application contains all of the information that the server needs to process the request. The API's architecture is made simpler and can easily scale because of this. Additionally, clients can cache responses to enhance performance and reduce network traffic when using REST APIs that are frequently designed to be cacheable.

Simple Object Access Protocol API

The internet uses the SOAP (Simple Object Access Protocol) protocol to exchange structured data. A Cleanser Programming interface (Application Programming Point of interaction) is a bunch of decides and conventions that oversee how two programming applications can speak with one another utilizing the Cleanser convention.A SOAP API defines a set of XML-based messages that can be used to perform operations on remote objects or services. Each message typically includes a request and a response, with each message being sent over an HTTP or HTTPS connection [10]. The SOAP messages are structured using an XML schema, which defines the structure of the message and the data it contains. When a client application wants to use a SOAP API, it sends a request to the API using a SOAP message. The request typically includes information about the operation to be performed, such as retrieving data, creating a new resource, or updating an existing resource [9]. The SOAP message also includes a header and a body, which contain additional information about the request and the data being sent.

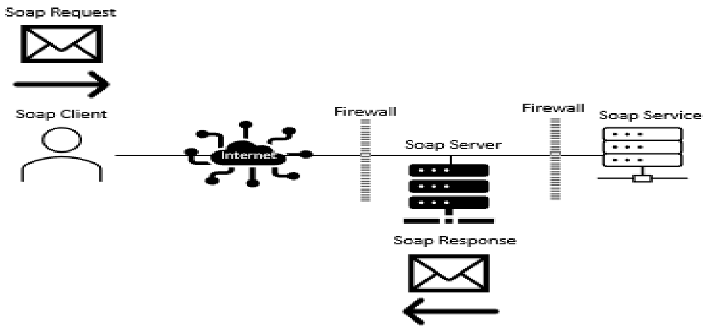


Figure 10: Working of SOAP API

Once the API receives the request, it processes it and sends a SOAP response back to the client application. The response typically includes a status code, such as Success or Failure, which indicates whether the request was successful or not. It also includes any data or information requested by the client application. SOAP APIs can be used to enable integration between different software applications or to provide a service to external developers or clients. They are commonly used in web development, enterprise applications, and in the development of other software applications that require communication over the internet.

SOAP APIs are typically documented using a WSDL (Web Services Description Language), which describes the available operations, their parameters, and the expected responses. This allows developers to easily understand how to interact with the API and build applications that use it [11]. Additionally, many SOAP APIs are secured using authentication and authorization mechanisms to ensure that only authorized clients can access the API and its resources.

Conclusion:

In conclusion, developing software applications that make use of the internet protocol necessitates the use of APIs. Software developers can interact with the internet and exchange data thanks to the various APIs for IP that are available. The particular API that is best suited to a given application will be determined by the application's requirements and use case. The significance of APIs in the creation of cutting-edge software applications will only grow with the development of the internet.

References:

1. Ofoeda, J., Boateng, R., & Effah, J. (2019). Application programming interface (API) research: A review of the past to inform the future. *International Journal of Enterprise Information Systems (IJEIS)*, *15*(3), 76-95.
2. Fávero, L. P. L., Belfiore, P., & De Freitas Souza, R. (2023). Using application programming interfaces to collect data. In *Elsevier eBooks* (pp. 157–167). Krishna, S. M., & Sharma, R. (2021). Survey on application programming interfaces in software defined networks and network function virtualization. *Global Transitions Proceedings*, *2*(2), 199–204. https://doi.org/10.1016/j.gltp.2021.08.018
3. Lindman, J., Horkoff, J., Hammouda, I., & Knauss, E. (2018). Emerging perspectives of application programming interface strategy: A framework to respond to business concerns. *IEEE software*, *37*(2), 52-59.
4. Nawaz, M. S., Khan, S. U. R., Hussain, S., & Iqbal, J. (2023). A study on application programming interface recommendation: state-of-the-art techniques, challenges and future directions. *Library Hi Tech*, *41*(2), 355-385.
5. Jolley, K. A., Bray, J. E., & Maiden, M. C. (2017). A RESTful application programming interface for the PubMLST molecular typing and genome databases. *Database*, *2017*, bax060.
6. Muri, M. F. A., Utomo, H. S., & Sayyidati, R. (2019). Search Engine Get Application Programming Interface. *Jurnal Sains dan Informatika*, *5*(2), 88-97.
7. Gadre, D. V., Gupta, S., Gadre, D. V., & Gupta, S. (2018). Application Programming Interface (API). *Getting Started with Tiva ARM Cortex M4 Microcontrollers: A Lab Manual for Tiva LaunchPad Evaluation Kit*, 95-100.
8. Wahyuni, F., & Fitriati, R. (2021, February). Why is the Application Programming Interface the backbone of a Smart City?. In *Journal of Physics: Conference Series* (Vol. 1783, No. 1, p. 012029). IOP Publishing.
9. Ogorodnyk, O., Larsen, M., Martinsen, K., & Lyngstad, O. V. (2021b). Development of application programming interface prototype for injection molding machines. *Procedia CIRP*, *97*, 453–458.
10. Liu, X., Meneghin, M., & Shapiro, V. (2021). An application programming interface for multiscale shape-material modeling. *Advances in Engineering Software*, *161*, 103046.
11. Kemer, E., & Samli, R. (2019). Performance comparison of scalable rest application programming interfaces in different platforms. *Computer Standards & Interfaces*, *66*, 103355.

PROCESS MANAGEMENT IN OPERATING SYSTEM

Sumit Chopra*, Diksha Rani, Rajesh Sharma and Simranjot Kaur
GNA University, Phagwara, Punjab
*Corresponding author E-mail: sumit.chopra@gnauniversity.edu.in

Abstract:

The execution of the program with complete control over the system resources was one of the scenarios used in the early era of computing. This program was privileged with all the resources it needed from the system. Nowadays, we have the privilege to load and simultaneously execute multiple programs into the computer memory. It results in describing a process as a program under execution in the computer system. Programs defined by the users are taken into more consideration in the complex operating systems. The system tasks which are kept outside of the kernel along with the user's program are executed by the operating system. Multiple processes are executed simultaneously in the operating system. The performance of the computer can be increased by switching and distributing CPU among different processes so that all the processes have easy access to the resources.

Keywords: deadlock, process, multiprogramming, synchronization, throughput, semaphores, scheduling, pseudo parallelism, CPU Scheduling

Introduction:

An operating system can be defined as a software that acts as an interface between the user and the computer hardware. It works to manage and coordinate usage of hardware among multiple applications as well as users. Common functions of programs in computers and the management of hardware resources is done by the collection of programs commonly known as the operating system. That's why it is considered as the most crucial part of the computer system. The most crucial operations performed by an operating system are generation of process, process state, process synchronization, avoidance of deadlock and process communication. The multitasking characteristic of an operating system allows it to execute or run two or more processes simultaneously. The environment that supports multitasking of multiple programs helps to manage the accessing of the available resources like CPU to the processes and contains the details about the processes, resource allocation time, and the execution time of the process. This scenario leads to the need to manage different running processes in the operating system [1]. There are different concerns occurred due to multithreading of an operating system which are discussed. The complication of the interface between the communication layers and the processes is one of the concerns. The second obstacle is that there is no reduction in the counts of process switches that increase the processing time. Due to the unspecified order of the events, it is so difficult to program and debug the interface of non-blocking transactions which is again a matter of concern in the process management. The simple and light weight processes having shared state can implement the parallelism. A single address space is being shared by the multiple threads in the operating system which lead to the fast switching of the threads. The action of servers serving one process to a single client at a particular time is possible due to the process's light weightage. It is possible to serve multiple clients simultaneously by generating multiple parallel processes having light weightage. Normally, the mechanism of processes having shared memory and the same data structure is possible due to the process synchronization [2,3].

Process

A program under execution running on an operating system is termed as a process. Since program instructions need resources like CPU, memory, access to files and Input/Output devices to be executed and accomplish the task programmed in the process. The process requests the processor for resources either during the creation of the process or execution of the process. There are situations when different processes in the system need the same resources simultaneously for their execution. Therefore, the operating system must manage all the processes and the resources in order to keep smooth functioning of all tasks in the system. We can also define a process as a text section due to its feature of showing current activities just as shown by a program counter. A process takes space in the computer main memory and holds the following sections of the memory. The temporary data like local variables, function parameters, return addresses of a function are stored in the stack. The process holds the data section of the memory that stores the global variables of the process program. The dynamic memory is allocated to the process during its run time in the heap section of the main memory[4].

Classification of processes in an operating system

There is a pool of processes in a system. It is the type of an operating system that decides whether the processes execute simultaneously or sequentially. The processes are categorized as following:

- **Operating system process:** This system executes system code. It runs in kernel mode. It performs administrative and housekeeping functions such as memory allocation, process swapping, scheduling etc.
- **User process:** User process executes user code or user applications. These processes execute in user mode to run the user program and its utilities. The processes belonging to the kernel are executed in kernel mode.
- **Event-specific based category of process**

  a) CPU bound process: Processes that spend the majority of their time simply using the CPU (doing calculations).

b) I/O bound process: Processes that are associated with input/output-based activity like reading from files, etc.

- **Nature based category of process**

  a) Independent process: A process that does not need any other external factor to get triggered is an independent process.

  b) Cooperative process: A process that works on the occurrence of any event and the outcome affects any part of the rest of the system is a cooperating process.

Operating system operations in process management

The operating system plays a major role in deciding the burst time, resource allocation time of any process. It decides which and when a process can access the CPU for how much period of time in the multiprogramming environment. The following operations are performed by the operating system in process management: The creation and deletion of any process is managed by the operating system. It provides the method for process communication, process synchronization and the mechanism of pausing and resuming the processes is provided by the operating system.

Process synchronization: In process synchronization, the order of cooperative processes can be maintained for smooth running of all the processing available in the ready queue. The cooperative processes can affect or can be affected by the execution of other processes. This is done to prevent interference of any process in the execution of another process. Send() and receive() primitives are sent to allow communications between multiple processes through calls. Various design choices are available to carry out each primitive. Synchronous and asynchronous are the types of messages passing which can be blocking or non-blocking respectively. When it is blocking send, until the receiving process or the mailbox receives the message, the sending process is blocked. When it is non-blocking send, the message is sent by the sending process and operation is resumed. When it is Blocking receive, until a message is made available, the receiver is blocked. When it is non-blocking, a valid or a null message is fetched by the receiver. Numerous blends of send() and receive() are available. The sender and the receiver communicate with each other when we block the send() and receive() at the same time.

Computer process

The state of a machine and operations that allow us to switch from one state to another is called the Machine State. The main memory holding a process is one of the components of a machine state that constitutes a process. The memory holds the instructions and the data read and written by the running program. Consequently, the memory holded by the process comprises it.

In the single-core CPU processor, only a single instruction or a process is allowed to execute at a time. Therefore, two processes must execute in turns to access the resources and to be accessed by the CPU for their execution. If only one process is running at a particular time, then it must be termed as to be either in running or not running state. Then the process in the non-running state is called to be in the Ready state which describes that the process is currently not running due to execution of some other process but will be likely to run [5, 6].

There are two terms - Dispatch and Interrupt. Dispatch describes that the operating system decided that the process should begin to run now. Where Interrupt – means that the operating system makes a decision to stop the running process so that another process can be executed. The operating system generally keeps a list of all processes currently loaded in a table known as the process table. The process table needs to keep the current state value. So if an operating system is running four programs such as Word, Notepad, Microsoft Edge and Excel, then the process table is shown in Table 1.

Table 1: Process Table at a given instance of time

Processor ID	Name	State
4	Notepad.exe	Running
6	Microsoft Edge.exe	Ready
7	Word.exe	Ready
8	Excel.exe	Ready

CPU scheduling

Multiprogrammed operating system is based on CPU scheduling. CPU scheduling is a technique of switching CPU among multiple processes in order to increase the performance and productivity of the system. It is basically fundamental function of an operating system that helps to maximize the utilization of CPU. In a single core CPU, a single process is executed at a particular time and other processes have to wait for their execution and CPU allocation until the current process releases the CPU on completing its execution. On the other side, in multiprogramming operating system, multiple processes executed simultaneously to maximize the utilization of CPU. This mechanism is called pseudo parallelism. In this mechanism, a process keeps executing until the completion of I/O request. But in a single core CPU operating system, the CPU waits idle and is not allocated to any process until the completion of I/O request [7,8]. This leads to wastage of time while CPU sits idle. Meanwhile, no useful task is accomplished. In contrast, this waiting time is accomplished productively in multiprogramming operating system. At a particular time, multiple processes are kept in the main memory of the computer. In case, one process waits for the completion of I/O request for resources, the CPU is taken away from the process by the operating system and is allocated to another process which is ready for CPU allocation. This task is continuously carried out by the operating system. This mechanism is called CPU scheduling and is diagrammatically shown in Fig. 1.

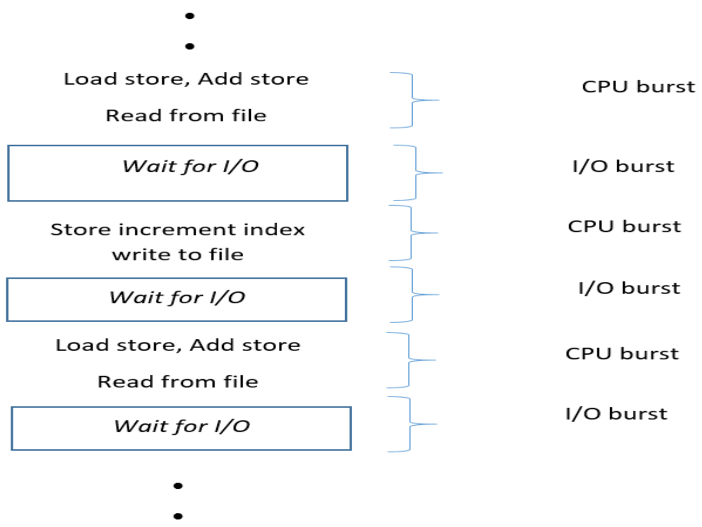


Figure 1: I/O and CPU burst

CPU scheduler - An operating system has a kernel function known as CPU Scheduler which is used to select which process in the ready queue is to be executed next. This selection of a process is based on a predefined scheduling algorithm.

CPU dispatcher - The operating system has a kernel function known as CPU dispatcher. It is used to give the control of CPU to the process which is selected by the CPU scheduler.

CPU scheduling algorithms:

First Come First Served (FCFS) Scheduling: In this scheduling algorithm, all the processes are given equal priority. The first process in the ready queue is scheduled by the algorithm. For CPU scheduling this scheduling mechanism is considered as the simplest scheduling algorithm. The cons of FCFS scheduling is that if the ready queue contains the long process in the front, then the short processes have to wait for a long time for their execution until the long process completes its execution

Shortest Job First (SJF) Scheduling: In this scheduling, the shortest job or process is given the highest priority. This results in ideal average waiting time.

Round Robin (RR) Scheduling: In this scheduling mechanism, the CPU is allocated to the first process in the ready queue for a particular predefined time slice. If the running process is not completed yet and its time slice got over, then it is put back at the ready queue's tail.

Priority (PR) scheduling: In this scheduling, all the processes are given different priorities. The CPU is scheduled to the process with highest priority. But starvation is the disadvantage of PR scheduling. Starvation is a term used to describe the situation when processes with lower priorities are blocked under high CPU load [9,10] .

Linux commands for process management

ps: ps command stands for **Process Status.** It shows the details regarding the processes running on an operating system. This command shows the process IDs (PIDs) of processes for the current shell and lists the processes which are currently running on the system. The various Linux commands associated with processes are shown in Table 2.

Table 2: Various Linux commands associated with Processes

Command	**Description**
PID	Provides the unique process ID assigned to the process.
TIME	Represents the amount of time for which the process executes.
CMD	Represents a command executed to launch the process.
TTY	Represents the type of terminal into which the user is logged in.

To kill a process: In case the computer processes become unresponsive or are consuming a lot of resources, in that case we need to kill that process. There are running background processes that are unresponsive and malfunction thus are not allowed to shut down. So we need to kill them explicitly. The root user has the permission to kill all the processes. When a process is killed, a termination message like **SIGKILL** or **SIGTERM** is sent to the given process. Most of the time, **SIGKILL** is a faster, resultant and efficient method of terminating a process. To terminate a process, we must first locate the process either by using **top** or **ps** command that shows the complete list of running processes on the system along with the detailed information of each process. The **pgrep -u root** command views the list of the processes possessed by the root user. If name of the process is known, then **pidof** is used to find the process ID. There are different commands to kill a process depending on factors like process name, process ID or the running duration of the process which is to be killed. The command **kill by pid** is used to kill a process whose process ID is known as shown in Fig 2. In **killall** command, the name of the process to be terminated is required. For instance, **kill sleep** command will kill all the sleep processes active on the system as shown in Fig 3. **bg** command is used to list and manage background jobs **fg** command is used to bring the most recently running job or process to foreground. To bring a specific job or process to the foreground, we will use the **fg** command with the name of the process.

The various extensions of the ps commands are shown in Table 3.

Table 3: Extensions used with ps command

COMMAND	**DESCRIPTION**
ps -A or ps -e	views all the running processes in the system.
ps -a	view all processes excluding session leaders and the processes which are not linked with the terminal.
ps -d	list all processes excluding session leaders.
ps -a -N or ps -a –deselect	list session leaders and the processes linked with the terminal.
ps -T	list all the processes which are linked with the terminal.
ps -r	list all the processes which are currently running in the system.

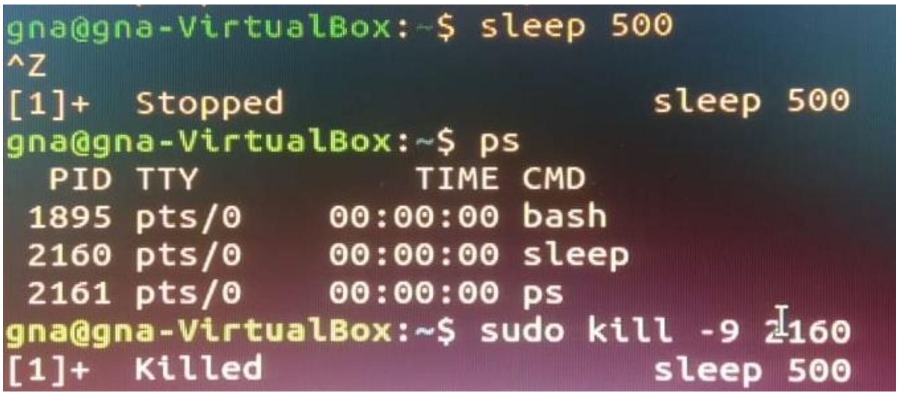


Figure 2: Snapshot of Sleep command in Ubuntu

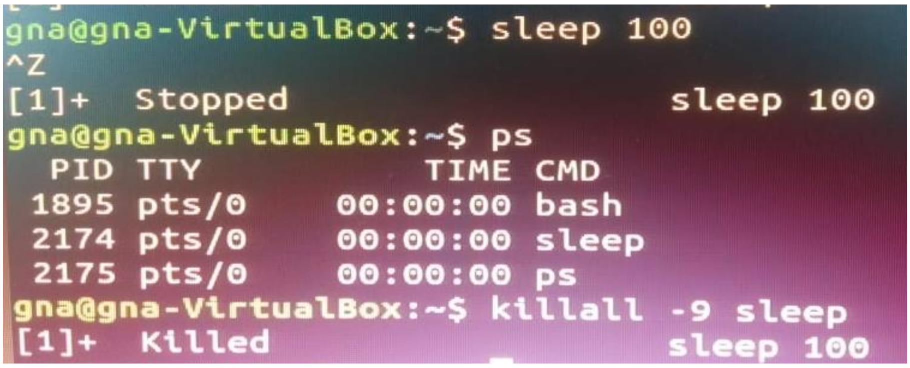


Figure 3: Snapshot of kill Command in Ubuntu

Conclusion:

In the chapter, the various aspects of Process Management in Operating system have been reviewed. To deal with the process management in any Operating system, certain set of commands can be used and the same has been elaborated in Linux. The various commands used for Process Management with their extensions have been elaborated in Ubuntu and how the same have graphically represented. If proper knowledge of command line is there, working and dealing with the processes becomes easier.

References

1. Behie, S. W., Halim, S. Z., Efaw, B., O'Connor, T. M., & Quddus, N. (2020). Guidance to improve the effectiveness of process safety management systems in operating facilities. *Journal of loss prevention in the process industries*, *68*, 104257.
2. Silberschatz, A., Peterson, J. L., & Galvin, P. B. (1991). *Operating system concepts*. Addison-Wesley Longman Publishing Co., Inc..
3. Comer, D. (2011). *Operating system design: the Xinu approach, Linksys version*. CRC Press.
4. Golm, M., Felser, M., Wawersich, C., & Kleinöder, J. (2002, June). The JX Operating System. In *USENIX Annual Technical Conference, General Track* (pp. 45-58).

5. Saraswat, P. (2023, February). An inclusive analysis of Google's android operating system and its security. In *AIP Conference Proceedings* (Vol. 2427, No. 1). AIP Publishing.
6. Harshvardhan, H., & Irabatti, S. (2023). Study of Kernels in Different Operating Systems in Mobile Devices. *Journal of Online Engineering Education*, *14*(1s), 21-26.
7. Barkin, J. (2023). *Specifying and Verifying Imperative Models of Operating System Structures* (Doctoral dissertation).
8. Singh, S. S. (2023). *A comparative approach for understanding, analyzing, and predicting the behavior of instruction set architectures and operating systems* (Doctoral dissertation, IIT Delhi).
9. Liu, P., Cui, B., Han, C., & Yang, Z. (2023, June). A fuzzing scheme for microkernel operating system based on relationship between system calls. In *International Conference on Computer Network Security and Software Engineering (CNSSE 2023)* (Vol. 12714, pp. 254-260). SPIE.
10. Aguilera, M. K., Amaro, E., Amit, N., Hunhoff, E., Yelam, A., & Zellweger, G. (2023). Memory disaggregation: why now and what are the challenges. *ACM SIGOPS Operating Systems Review*, *57*(1), 38-46.

INTRODUCTION TO SENSORS

M. W. Bhade

Department of Chemistry,

Amolakchand Mahavidyalaya, Yavatmal, Maharashtra, India

Corresponding author E-mail: madhuri.bhade@gmail.com

Introduction:

Sensor, there is no commonly accepted definition for sensor. However, in order to simplify, sensor can be defined as, "a sensor is a device that receives a signal or stimulus and responds with an electrical signal". The reason for the output of a sensor to be limited to electrical signals is related to the present development of signal processing, that is almost exclusively performed using electronic devices. Given this definition a sensor should be a device that receives a physical, chemical or biological signal and converts it into an electrical signal that should be compatible with electronic circuits. This definition may also be supported from the etymological origin of the word sensor. Sensor seems to come from the word 'sense' given that usually sensor devices try to mimic or reproduce human sense's characteristics. In the biological senses the output is also an electrical signal that is transmitted to the nervous system.

Sensors characteristics can be grouped into static or dynamic parameters, environmental conditions and structural related characteristics. Static parameters are the ones that describe the transfer function of a sensor, i.e., the relation between the input and the output of a sensor when the input does not vary significantly with time. On the other hand, dynamic characteristics try to describe the performance of the sensor taking account of the variation of the stimulus with time. Environmental conditions are all those factors that interfere with the sensor mechanism and thus change its response to the input stimulus. Finally, structural related characteristics are those that result from the specific design and components of the sensor. In these last characteristic could be included: cost, weight, power consumption, lifetime and compatibility with silicon based manufacturing technologies.

Resolution is the smallest increment in the input stimulus that results in a detectable increment in the output. This parameter is obviously limited by the precision, but also by noise, sensitivity and repeatability. The smallest increment of the measurand from a zero value that causes a detectable output is often called the detection threshold and the span of the input stimulus that produces a meaningful output is called range or span.

Although sensitivity has no standard definition, in gas sensors research it is a widely used parameter. In the majority of the publications dealing with the gas sensors, sensitivity is estimated using the ratio of the output response of the device to the output in a reference atmosphere, and is a dimensionless value. However, in many other fields of research, sensitivity is defined as the ratio of the measurand variation to the variation of the output. The former estimate will be called the response ratio. Given the absence of a common definition one has to be very careful when comparing the sensitivities for different devices. Noise is a random fluctuation in the sensor output that might be caused by random fluctuations in the measurand or by external interferences in the conversion mechanism. Repeatability is the difference in the output of the sensor when the input value is consecutively reached using identical procedures. This is a very limiting factor, since a low repeatability can lead to a very high uncertainty in the output value. A difference in the output of a sensor, for a given input value, can otherwise be caused by hysteresis. This is the difference in the output of a sensor, when the input value is reached from opposite directions. If a sensor shows hysteresis, the output value depends on the history of the sensor, i. e., if the previous input stimulus was lower than the present one. The output may be different from the one observed when the previous value was higher. Nevertheless this does not change the uncertainty. The error, though, can increase if this factor is not counter balanced by this signal processing.

Finally there are some parameters much relevant to the coupling with the electronic interface that should also be included in the static characteristics. These are the form of the electrical output, the form of the modulating signal (we needed), the output impedance, the leakage current and the grounding. These characteristics do not generally alter the performance of the sensor if the electronic circuit is properly designed, nevertheless they can change its operation.

Chemical sensors:

These are the devices used to ascertain the chemical properties of materials. This is a fast growing field of research where steadily appear new ideas and prototypes. Unlike other sensor fields, the number of mass-produced devices does not generally follow the research activity. Many Chemicals sensor devices how large batch-to-batch irreproducibility which originates high costs due to the need to calibrate each single device; skilled staff are really needed for accurate calibration and general troubleshooting, and in many cases, devices are not even suited for mass production. The large number of interfering parameters make it very difficult to complete characterization of chemical sensors. Thus frequently the progress in technology surpasses science and reduces sensor optimization to a trial and error process.

For the purpose of presenting the state-of-the-art of chemical sensor technologies it is necessary to put some order in the available devices. As in the case of general sensors, considering the detection principles, chemical sensors can be classified into electrical, magnetic, thermal, optical, mechanical or radiation sensors. A classification could also be tried having in minds the materials or the technologies; however, the former is preferable since it can give a broader view of this field of research.

In the sensors based on electrical properties the output signal is either generated by a reaction involving charge transfer or is modulated by that reaction. This implies that there must be an electric current flow through the active sensor material in order to make a measurement, and thus, at least two electrodes. One of the first devices used to sense a chemical quantity was the electrochemical cell. Over the years, several modifications have been introduced in the original configuration of the cell in order to widen the measuring scope and optimize the performance of the device. An electrochemical cell is usually built with two or more electronic conductors, called the electrodes, immersed in an electrolyte, which is an ionic conductor. The electrolyte might be a solid, a liquid or a gas. A more recent example is the field effect chemical sensor. This is a solid state device similar in its working principle to the JFET or the MOSFET devices. In these devices a conducting channel exists between the source and the drain electrodes, which are modulated by the potential of a third electrode that is called the gate. In field effect chemical sensors the potential generated at the gate depends on the chemical quantity being measured. This quantity can therefore be determined by observing the current flow from the source to the drain of the device. In this same class of chemical sensors based on electric properties are included the thin film chemo-sensitive sensors. These are very promising devices because they can be made very small and cheap and are compatible with silicon technology, which opens the possibility to integrate the sensors in the processing electronics circuit.

Neither the magnetic nor the radiation sensors are usually included in the chemical sensors field. This kind of sensors, may be used to detect magnetic or radiation properties of materials. This alone does not justify the inclusion in the chemical sensors class since these are not chemical but rather physical properties. However, if these properties are used to determine composition they can be included in that class, and thus it is worth to include them here. Chemical sensors based in thermal properties measurement are also an important class. These are mostly used to detect flammable gases because of the exothermic character of their reaction with oxygen. The most common of these devices have a catalytic layer that promotes the reaction at low temperature, and a temperature sensor that measures the temperature variation caused by the heat exchange. Both the presence of a substance and its concentration in atmosphere can be determined by this kind of sensors. When there is a temperature difference between the sensor and the surrounding atmosphere, the temperature variation can be related to a change of the thermal conductivity of the atmosphere and therefore also used to monitor its composition. Obviously, not only exothermic but also endothermic reactions can be monitored using such a device.

Optical characteristics provide good fingerprints to distinguish different substances and are widely used in materials characterization. Recent advances in optoelectronics and fiber optic techniques have brought some promising new ways to use these properties in chemical sensors.

A fiber-optic chemical sensing device consists usually in a light source, a fibre coupler to lead the light into the fibre, the light guide, a decoupler where the returning light is separated from the exciting light and light detection and amplification system. The measure properties may be, for instance absorbance, reflectance, fluorescence, light scattering or refractive index.

Finally there are two types of devices that may be included in the mechanical sensors class, that are the bulk acoustic wave sensors (BAW) and the surface acoustic wave sensors (SAW). The advantages of SAW technology over BAW include the compatibility with planar silicon technology and the possibility of using higher frequencies and potentially higher sensitivity.

Metal oxide based gas sensors:

Under the influence of ambient gases; metal oxide layers used in gas sensors may undergo either surface or bulk conductance changes. Usually surface conductance changes are associated with electron transport processes, while bulk conductance changes generally imply ion transport. Depending on the type of charge carrier involved in the sensing processes, these devices may be divided into three different groups: electronic conductance sensors, if the charge carriers are electrons or holes; ion conductance sensors, if the charge carriers are exclusively ions and mixed conductance sensors if the charge carriers are electrons and ions. Electrochemical sensors belong to the class of the ion conductance sensors.

Thin film resistive gas sensors:

Resistive sensors have been used to measure a wide variety of physical and chemical properties and are among the most common and cheap sensors, which use materials that change conductivity with light absorption; thermo-resistive sensors in which resistivity variation is controlled by the temperatures; piezoresistive sensors, that use the change in resistance with mechanical stress; magneto-resistive sensors based on the resistivity change in the presence of an external magnetic field and chemo-resistive that measure the resistivity change produced by the interaction of a chemical substance with the sensing material. The materials employed in these sensors are frequently produced in thin film form and in many cases can be produced by magnetron sputtering.

References:

1. https://en.wikipedia.org/wiki/Sensor
2. https://www.analog.com/media/en/training-seminars/design-handbooks/Practical-Design-Techniques-Sensor-Signal
3. https://en.wikipedia.org/wiki/Chemical_sensor_array

PROTEIN ENGINEERING FOR PRODUCTION OF INDUSTRIAL IMPORTANT ENZYMES

Kirti Yadav*[1,2], Neeraj K. Aggarwal[1] and Awtar Singh[2]

[1]Department of Microbiology, Kurukshetra University, Kurukshetra, Haryana

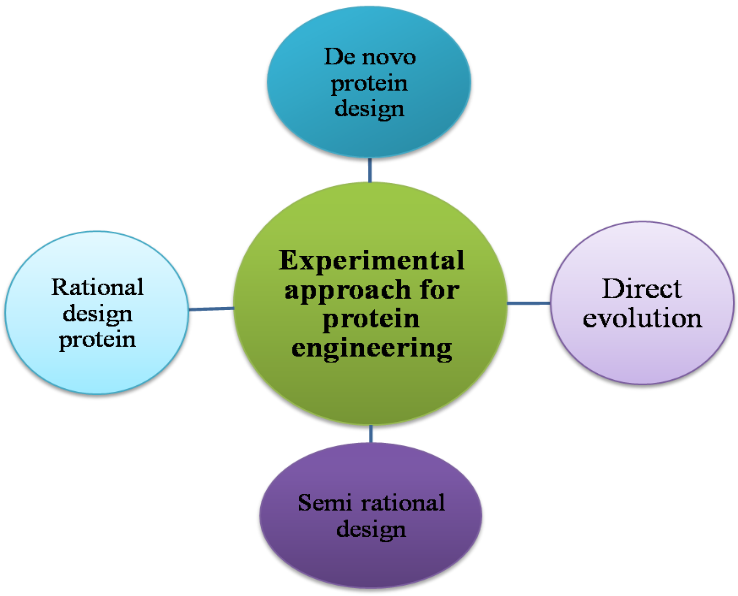
[2]ICAR–Central Soil Salinity Research Institute, Karnal, Haryana

*Corresponding author E-mail: kirtiyadav021@gmail.com

Introduction:

Protein engineering is the design of new enzymes or proteins with new or desirable functions. It is based on the use of recombinant DNA technology to change amino acid sequences. The first papers on protein engineering date back to early 1980ies: in a review by Ulmer (1983), the prospects for protein engineering, such as X-ray crystallography, chemical

DNA synthesis, computer modelling of protein structure and folding were discussed and the combination of crystal structure and protein chemistry information with artificial gene synthesis was emphasized as a powerful approach to obtain proteins with desirable properties (Ulmer, 1983). In a later review in 1992, protein engineering was mentioned as a highly promising technique within the frame of biocatalyst engineering to improve enzyme stability and efficiency in low water systems (Gupta, 1992). Today, owing to the development in recombinant DNA technology and high-throughput screening techniques, protein engineering methods and applications are becoming increasingly important and widespread.



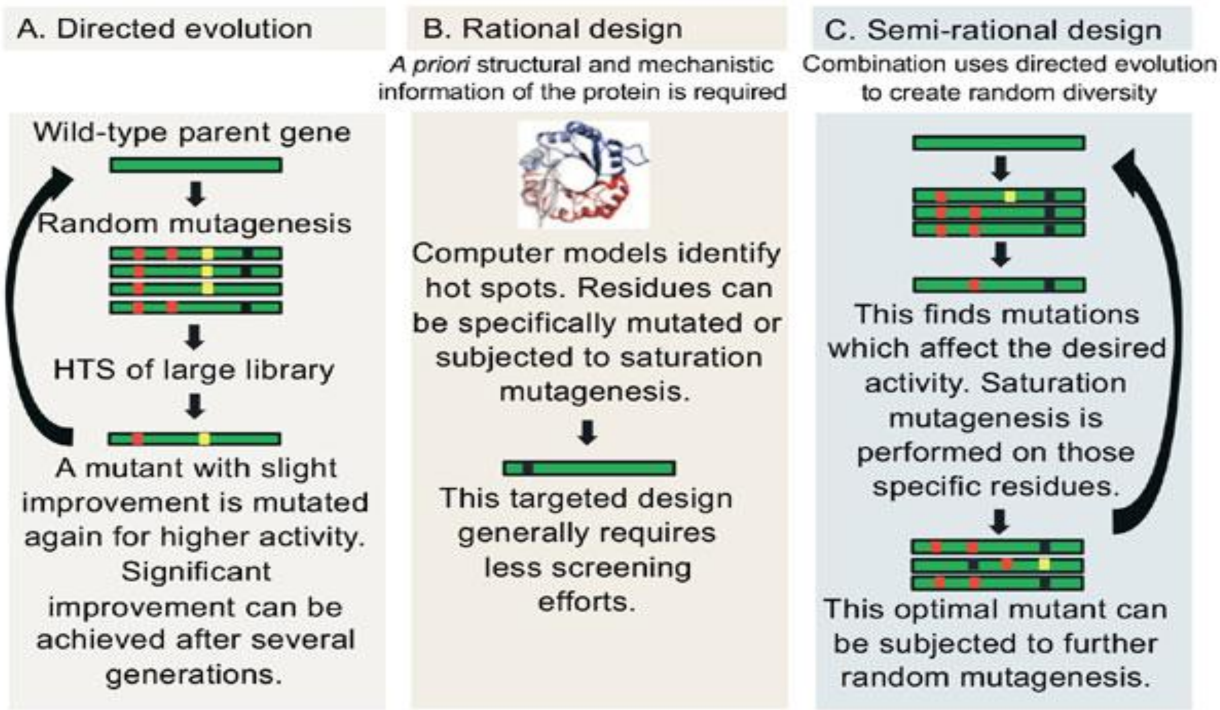
Protein engineering methods

Many different protein engineering methods are available today, owing to the rapid development in biological sciences, more specifically, recombinant DNA technology.

The most classical method in protein engineering is the so-called "rational design" approach which involves "site-directed mutagenesis" of proteins (Arnold, 1993). Site-directed mutagenesis allows introduction of specific amino acids into a target gene. There are two common methods for site-directed mutagenesis. One is called the "overlap extension" method. This method involves two primer pairs, where one primer of each primer pair contains the mutant codon with a mismatched sequence. These four primers are used in the first polymerase chain reaction (PCR), where two PCRs take place, and two double-stranded DNA products are obtained. Upon denaturation and annealing of them, two hetero-duplexes are formed, and each strand of the hetero-duplex involves the desired mutagenic codon.

DNA polymerase is then used to fill in the overlapping 3' and 5' ends of each hetero-duplex and the second PCR takes place using the non-mutated primer set to amplify the mutagenic DNA. The other site-directed mutagenesis method is called "whole plasmid single round PCR". This method forms the basis of the commercial "Quick-change Site-Directed Mutagenesis Kit" from Strata gene. It requires two oligonucleotide primers with the desired mutation(s) which are complementary to the opposite strands of a double-stranded DNA plasmid template. Using DNA polymerase PCR takes place, and both strands of the template are replicated without displacing the primers and a mutated plasmid is obtained with breaks that do not overlap. *Dpn*Imethylase is then used for selective digestion to obtain a circular, nicked vector with the mutant gene. Upon transformation of the nicked vector into competent cells, the nick in the DNA is repaired, and a circular, mutated plasmid is obtained (Antikainen& Martin, 2005).

Protein engineering methods comprise three main strategies; rational design, directed evolution and a combination of both methods, semi rational design (site saturation mutagenesis).

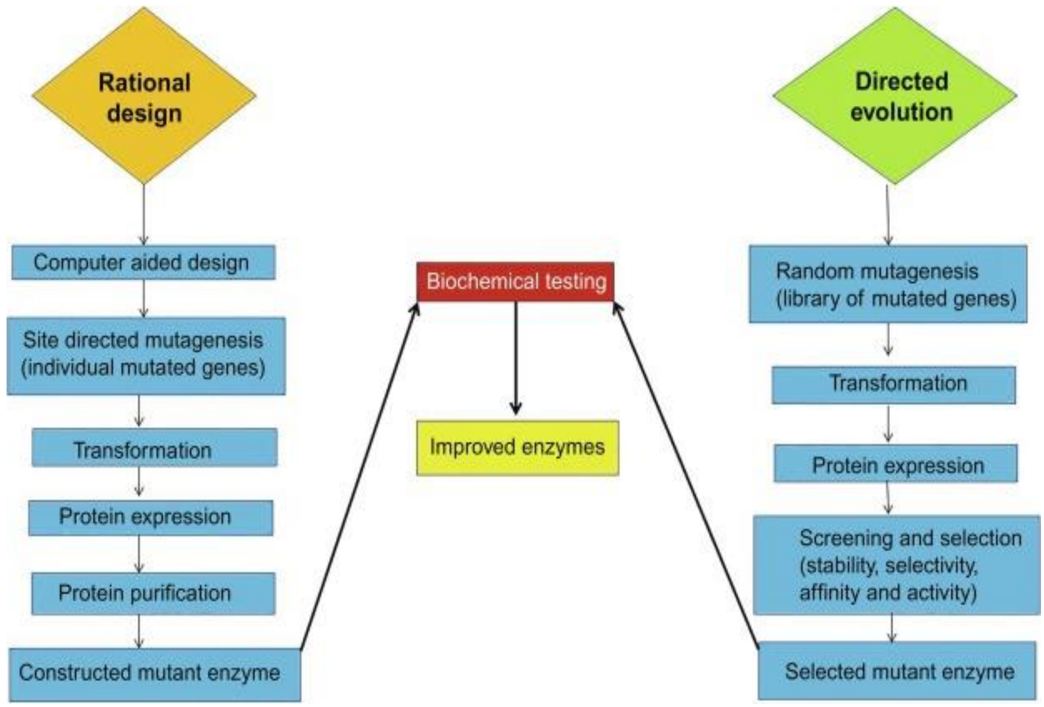


Rational protein design

In rational protein design, a scientist uses detailed knowledge of the structure and function of a protein to make desired changes. In general, this has the advantage of being inexpensive and technically easy, since site-directed mutagenesis methods are well-developed. However, its major drawback is that detailed structural knowledge of a protein is often unavailable, and, even when available, it can be very difficult to predict the effects of various mutations since structural information most often provide a static picture of a protein structure (Farmer *et al.,* 2017).

Directed evolution

In directed evolution, random mutagenesis, e.g. by error-prone PCR or sequence saturation mutagenesis, is applied to a protein, and a selection regime is used to select variants having desired traits. Further rounds of mutation and selection are then applied. This method mimics natural evolution and, in general, produces superior results to rational design. An added process, termed DNA shuffling, mixes and matches pieces of successful variants to produce better results. Such processes mimic the recombination that occurs naturally during sexual reproduction. Advantages of directed evolution are that it requires no prior structural knowledge of a protein, nor is it necessary to be able to predict what effect a given mutation will have. Indeed, the results of directed evolution experiments are often surprising in that desired changes are often caused by mutations that were not expected to have some effect. The drawback is that they require high-throughput screening, which is not feasible for all proteins. Large amounts of recombinant DNA must be mutated and the products screened for desired traits. The large number of variants often requires expensive robotic equipment to automate the process. Further, not all desired activities can be screened for easily.



Semi-rational design

Semi-rational design uses information about a proteins sequence, structure and function, in tandem with predictive algorithms. Together these are used to identify target amino acid residues which are most likely to influence protein function. Mutations of these key amino acid residues create libraries of mutant proteins that are more likely to have enhanced properties (Lutz, Stefan, 2010).

Advances in semi-rational enzyme engineering and de novo enzyme design provide researchers with powerful and effective new strategies to manipulate biocatalysts. Integration of sequence and structure based approaches in library design has proven to be a great guide for enzyme redesign. Generally, current computational de novo and redesign methods do not compare to evolved variants in catalytic performance. Although experimental optimization may be produced using directed evolution, further improvements in the accuracy of structure predictions and greater catalytic ability will be achieved with improvements in design algorithms. Further functional enhancements may be included in future simulations by integrating protein dynamics (Lutz, Stefan, 2010).

Screening and selection techniques

Once a protein has undergone directed evolution, ration design or semi-ration design, the libraries of mutant proteins must be screened to determine which mutants show enhanced properties. Phage display methods are one option for screening proteins. This method involves the fusion of genes encoding the variant polypeptides with phage coat protein genes. Protein variants expressed on phage surfaces are selected by binding with immobilized targets in vitro. Phages with selected protein variants are then amplified in bacteria, followed by the identification of positive clones by enzyme linked immune-sorbent assay. These selected phages are then subjected to DNA sequencing (Poluri *et al.,* 2017)

Cell surface display systems can also be utilized to screen mutant polypeptide libraries. The library mutant genes are incorporated into expression vectors which are then transformed into appropriate host cells. These host cells are subjected to further high throughput screening methods to identify the cells with desired phenotypes (Poluri *et al.,* 2017). Cell free display systems have been developed to exploit *in vitro* protein translation or cell free translation. These methods include mRNA display, ribosome display, covalent and non-covalent DNA display, and *in vitro* compartmentalization.

Chemical modification of enzymes

The protein synthesized under the control of gene sequence in a cell undergoes post-translational modification. This lead to stability, structural integrity, altered solubility and viscosity of individual proteins. Example - Enzyme – PEG conjugates

PEG – L- asparginase conjugates differ from the native enzyme in the following way –

- It retains only 52% f the catalytic activity of the native.

- It became resistant to proteolytic degradation it doesn't cause allergy

De novo synthesis of complete modified gene

Complete genes in some cases have been chemically synthesized in the form of several oligomers (e.g. genes for insulin), that ae ligated in correct order to produce a complete gene. The sequence of the synthetic gene can be designed in a modular fusion to get the desired function.

Protein engineering applications on industrially important enzymes

Protein engineering applications with a variety of industrially important enzymes can be found in the literature.

These includes –nitrilases (Martinkova&Kren, 2010), aldolases (Clapes*et al.,* 2010), microbial beta-D-xylosidases (Jordan &Wagschal, 2010) etc. Nitrilases are important enzymes for biotransformation, but the enzymatic reactions require improvement for higher industrial process efficiencies. For this purpose, new enzymes were screened from new isolates, medium and protein engineering methods were applied (Martinkova&Kren, 2010). Aldolases are also important enzymes for stereoselective synthesis reactions regarding carbon-carbon bond formation in synthetic organic chemistry. Protein engineering or screening methods improved aldolases for such synthesis reactions.

De novo computational design of aldolases, aldolase ribozymes etc. are promising applications (Clapes*et al.,* 2010). Microbial beta-D-xylosidases are also an industrially important group of enzymes, particularly for baking industry, animal feeding, D-xylose production for xylitol manufacturing and deinking of recycled paper. As they catalyse hydrolysis of non-reducing end xylose residues from xylooligosaccharides, they could be used for the hydrolysis of lignocellulosic biomass in biofuel fermentations to produce ethanol and butanol. Thus, improving the catalytic efficiency of beta-D-xylosidases is crucial for many industrial applications (Jordan & Wagschal, 2010). As the use of organic solvents is industrially suitable for enzymatic reactions, but has adverse effects on enzyme activity and/or stability, protein engineering of organic solvent tolerant enzymes (Gupta, 1992; Doukyu & Ogino, 2010) has become an important research area. Screening organic solvent-tolerant bacteria or extremophiles has been preferred to isolate and improve naturally solvent-stable enzymes (Gupta &Khare, 2009; Doukyu&Ogino, 2010).

Other protein engineering examples with industrially and/or pharmacologically important enzymes include studies on cholesterol oxidase (Pollegioni *et al.,* 2009), cyclodextringlucanotransferases (Leemhuis*et al.,* 2010), human butyrylcholinesterase (Masson *et al.,* 2009), microbial glucoamylases (Kumar & Satyanarayana, 2009), revealed their different structural and functional characteristics which could be exploited for biotechnological applications and improved further by protein engineering (Bjarnason*et al.,* 1993; Hough & Danson, 1999; Georlette *et al.,* 2004).

Industrial processes require extreme conditions such as high pressure, temperature and extreme pH which require a large amount of energy to achieve and may produce unwanted toxic waste. Biological enzymes do not require such conditions and produce chirally pure products often without the disadvantages of unwanted toxic by-products.

They offer a number of advantages over conventional chemical catalysts (Davies, 2003; Kirk *et al.,* 2002; Perez, 2010; Tao, 2009; Wojtasiak, 2006):

- Most enzymes catalyze their reactions under mild conditions such as physiological temperature and pH (6-8) and so are therefore often compatible with one another. Compatible enzymes can be used together either in sequence or cooperatively to catalyze multistep reactions.
- Enzymes are regioselective, and also stereospecific and this allows the production of exact chiral products from racemic mixtures. Enantiomerically pure compounds are specially demanded by the pharmaceutical, food and cosmetics industries.
- They may be cheap and easy to use because many enzymes are commercially available.
- They are regarded as environmentally friendly because catalysis is achievedwithout organic solvents or the heavy metal toxic waste.

***Candida methylica* FDH as a case study**

The FDH enzyme was first discovered in 1950 (Uversky, 2003) but it has attracted attention in recent years due to its practical application in the regeneration of NAD(P)H in the enzymatic processes of chiral synthesis. FDH is widely used for coenzyme regeneration with enzymes used for optically pure product synthesis in the pharmaceutical, food, cosmetic and agriculture industries (Patel, 2004; Jormakka *et al.,* 2003). The FDH catalysed reaction is also a suitable model for investigating the general mechanism of hydride ion transfer because of direct transfer of hydride ion from the substrate onto the C4-atom of the nicotinamide moiety of NAD^+ without stages of acid-base catalysis (Serov *et al.,* 2002).

A wide range of organisms use formate in a variety of metabolic pathways. From aerobic to anaerobic organisms formate dehydrogenase is the last enzyme in the metabolic pathway which catalyzes the oxidation of formate to CO2 and water. The use of formate and formatedehydrogenase (FDH) have been extensively studied and reviewed (Thiskov& Popov, 2004, 2006).

NAD^+-dependent FDH is a dimeric enzyme with two identical subunits each has an independent active site, containing no metal ions or prosthetic groups. They are unable to use one-electron carriers as oxidizers and are highly specific to both formate and NAD^+.

NAD+-dependent FDH from *Candida methylica*(*cm*FDH) was previously isolated by Allen &Hollbrook, 1995). Its N-terminal amino acid sequence was determined and it was cloned into pKK223-3 and overproduced in *Escherichia coli*. *cm*FDH in pKK223-3 vector has been used in several studies but purification of enzyme was a time consuming and costly process.

Therefore, in order to eliminate difficulties in the purification of FDH and to produce quick and highly purified-homogeneous-recombinant protein, *cm*FDH was subcloned into pQE-2 expression vector and the amount of purified protein improved about 3 times. It was observed that the N-terminally His tagged FDH has similar activity to the FDH enzyme without the Histag after digestion with exopeptidases (Ordu and Karagüler, 2007). Since then, the recombinant FDH from *Candida methylica*has been intensively studied to improve the properties for the NAD(P)H regeneration by using protein engineering techniques.

Conclusion:

The modification of natural enzymes and proteins by protein engineering is an increasingly important scientific field. The well-known methods of rational design and directed evolution, as well as new techniques will enable efficient and easy modification of proteins. New technologies such as computational design, catalytic antibodies and mRNA display would be crucial for *de novo* engineering of enzymes and also for new areas of protein engineering. Protein engineering applications cover a broad range, including bio-catalysis for food and industry, as well as medical, environmental and nano-biotechnological applications. The application of site saturation mutagenesis which is a combination of both strategies represent the new route to obtain the biocatalysts with the desired properties.

References:

1. Allen, S. & Holbrook, J. (1995). Isolation, sequence and overexpression of the gene encoding NAD+-dependent formate dehydrogenase from the methylotrophic yeast Candida methylica. Gene, 162, 99-104.
2. Bjarnason, JB., Asgeirsson, B. & Fox, JW. (1993). Psychrophilic proteinases from Atlantic cod. ACS Symposium Series, Vol.516, No.5, (December 1993), pp.69-82.
3. Clapes, P., Fessner, WD., Sprenger, GA. &Samland, AK. (2010). Recent progress in stereoselective synthesis with aldolases. Current Opinion in Chemical Biology, Vol.14, No.2, pp.154-167.
4. Doukyu, N. &Ogino, H. (2010). Organic solvent-tolerant enzymes.Biochemical Engineering, Vol.48, No.3, (February 2010), pp.270-282.
5. Farmer, TylarSeiya; Bohse, Patrick; Kerr, Dianne (2017). "Rational Design Protein Engineering Through Crowdsourcing". Journal of Student Research. 6 (2): 31–38.
6. Georlette, D., Blaise, V., Collins, T., D'Amico, S., Gratia, E., Hoyoux, A., Marx, JC, Sonan, G., Feller, G. &Gerday, C. (2004). Some like it cold: biocatalysis at low temperatures. FEMS Microbiology Reviews, Vol.28, No.1, (February 2004), pp.25-42.
7. Gupta, A. &Khare, SK. (2009). Enzymes from solvent-tolerant microbes: Useful biocatalystsfor non-aqueous enzymology. Critical Reviews in Biotechnology, Vol.29, No.1, (March 2009), pp.44-54.

8. Gupta, MN. (1992). Enzyme function in organic-solvents. European Journal of Biochemistry, Vol. 203, No. 1-2, (January 1992), pp.25-32.
9. Hough, DW. & Danson, MJ. (1999). Extremozymes.Current Opinion in Chemical Biology, Vol.3, No.1, (February 1999), pp.39-46.
10. Jordan, DB. &Wagschal, K. (2010). Properties and applications of microbial beta-Dxylosidases featuring the catalytically efficient enzyme from Selenomonasruminantium. Applied Microbiology & Biotechnology, Vol.86, No.6, (May 2010), pp.1647-1658.
11. Jormakka, M., Byrney, B. & Iwata, S. (2003). Formate dehydrogenase – a versatile enzyme in changing environments. Current Opinion in Structural Biology, 13, pp.418–423.
12. Leemhuis, H., Kelly, RM. &Dijkhuizen, L. (2010). Engineering of cyclodextrin glucanotransferases and the impact for biotechnological applications. Applied Microbiology and Biotechnology, Vol.85, No.4, (January 2010), pp.823-835.
13. Lutz, Stefan (December 2010). "Beyond directed evolution -semi-rational protein engineering and design". Current Opinion in Biotechnology. 21 (6): 73 743.
14. Marcaida, MJ., Munoz, IG., Blanco, FJ., Prieto, J. & Montoya, G. (2010). Homing endonucleases: from basics to therapeutic applications. Cellular and Molecular Life Sciences, Vol.67, No.5, (March 2010), pp.727-748.
15. Martinkova, L. &Kren, V. (2010). Biotransformations with nitrilases.Current Opinion in Chemical Biology, Vol.14, No.2, (April 2010), pp.130-137.
16. Patel, R. N. (2004). Biocatalytic Synthesis of Chiral Intermediates.Food Technology and Biotechnology, 42, 4, pp. 305–325.
17. Pollegioni, L., Piubelli, L. &Molla, G. (2009). Cholesterol oxidase: biotechnological. FEBS Journal, Vol.276, No.23, (December 2009), pp.6857-6870.
18. Poluri, Krishna Mohan; Gulati, Khushboo (2017). Protein Engineering Techniques.SpringerBriefs in Applied Sciences and Technology.Springer. doi:10.1007/978-981-10-2732-1.
19. Serov, A. E.; Popova, A. S &Tishkov, V. I. (2002a). The Kinetic Mechanism of Formate Dehydrogenase from Bakery Yeast.Doklady Biochemistry and Biophysics, 382, 26–30.
20. Tishkov, V. I. & Popov, V.O. (2006). Protein engineering of formate dehydrogenase. Biomolecular Engineering, 23, pp.89–110.
21. Uversky, V. N. (2003). Protein folding revisited. A polypeptide chain at the folding – misfolding – nonfolding cross-roads: which way to go? Cellular and Molecular Life Sciences, 60, pp.1852–1871.
22. Ulmer, KM. (1983). Protein engineering. Science, Vol. 219, No. 4585, (February 1983), pp.666-671, ISSN: 0036-8075

BIOMATERIALS

D. R. Nagapure

Department of Physics,

Shri Sadguru Saibaba Science & Commerce College, Ashti, Maharashtra - 442707

Corresponding author E-mail: deepaknagapure@gmail.com

Introduction:

A biomaterial is essentially a material that is used and adapted for a medical application. Biomaterials can have a benign function, such as being used for a heart valve, or may be bioactive; used for a more interactive purpose such as hydroxy-apatite coated hip implants. Biomaterials are also used every day in dental applications, surgery, and drug delivery. While a definition for the term 'biomaterial' has been difficult to formulate, more widely accepted working definitions include: "A biomaterial is any material, natural or man-made, that comprises whole or part of a living structure or biomedical device which performs, augments, or replaces a natural function".

The development of biomaterials is not a new area of science, having existed for around half a century. The study of biomaterials is called biomaterial science. It is a provocative field of science, having experienced steady and strong growth over its history, with many companies investing large amounts of money into the development of new products. Biomaterial science encompasses elements of medicine, biology, chemistry, tissue engineering and materials science.

In modern history, metals have been used as implants since more than 100 years ago when Lane first introduced metal plate for bone fracture fixation in 1895 (Lane, 1895). In the early development, metal implants faced corrosion and insufficient strength problems (Lambotte, 1909, Sherman, 1912). Shortly after the introduction of the 18-8 stainless steel in 1920s, which has had far-superior corrosion resistance to anything in that time, it immediately attracted the interest of the clinicians. Thereafter, metal implants experienced vast development and clinical use [1].

One of the most significant current discussions in orthopaedic is the total joint replacements especially hip and knee and the increasing trend to replace degraded and destroyed biological materials by artificial organs. It is estimated that approximately 1 million hip replacements and 250,000 knee replacements are carried out per year. This number is expected to double between 1999 and 2025 as a result of aging populations worldwide and growing demand for a higher quality of life. Another statistical data estimated that by the end of 2030, the number of total hip replacements will increase by 174% and total knee arthroplasties is predicted to grow by 673% from the present rate [2].

Yet increasing demand for implants makes it crucial to accelerate efforts on biomaterials. Unfortunately, the currently used materials have been found to have tendencies to fail after long term usage due to not fulfilling some vital requirements such as modulus close to that of bone, high wear and corrosion resistance and good biocompatibility. Both material and design deficiencies contribute to failure of total joint replacements. Failure of current biomaterials imposes pain for patient and after some time revision surgery should be performed.

Type of metal used in biomedical depends on specific implant applications. 316L type stainless steel (316L SS) is still the most used alloy in all implants division. Biomaterial selection is one of the most challenging issues due to crucial requirements and biocompatibility, so it has been of major interest to material designers in recent years. The present study reviewed the currently used metallic biomaterials in hip and knee; stainless steel, chromium cobalt alloys and titanium alloys. The current investigation will focus on 316L stainless steel used for total hip replacement.

Biomaterials

Biomaterials are used to make devices to replace a part or a function of the body in safe, reliably economically, and physiologically acceptable manner. A variety of devices and materials are used in the treatment of disease or injury. Commonplace examples include suture needles, plates, teeth fillings, etc.

A biomaterial is essentially a material that is used and adapted for a medical application. Biomaterials can have a benign function, such as being used for a heart valve, or may be bioactive; used for a more interactive purpose such as hydroxy-apatite coated hip implants. Biomaterials are also used every day in dental applications, surgery, and drug delivery**.**

While a definition for the term 'biomaterial' has been difficult to formulate, more widely accepted working definitions include:

"A biomaterial is any material, natural or man-made, that comprises whole or part of a living structure or biomedical device which performs, augments, or replaces a natural function" [3].

Williams (1981) defined biomaterials as “nonviable materials used in medical devices, intended to interact with the biological systems” [4].

A biomaterial is any synthetic material that is used to replace or restore function to a body tissue and is continuously in contact with body fluids [5].

Interaction of biomaterials with the human body

Clinical results in orthopaedics have demonstrated that a great need exists to find new and better biomaterials that will help satisfy the minimum requirements for orthopaedic devices to perform correctly on a long-term basis [6].

- **Biocompatibility**

Biocompatibility is the primary characteristic that a medical device should have in any orthopaedic application; that is, it must not adversely affect the local and systemic host environment of interaction (bone, soft tissues, ionic composition of plasma, as well as intra- and extracellular fluids).

- **Appropriate design and manufacturability of implants**

Finite element analysis is a powerful analytical tool used in the design of joint replacement prostheses. Currently modern manufacturing processes are necessary to guarantee the quality needed in orthopaedic devices.

- **Mechanical and biological stabilities**

The orthopaedic surgeon should seek the biomechanical stability of the implant, and the human body will take care of the biological stability.

- **Properties of biomaterials**

Some of the most important properties of biomaterials that should be carefully studied and analyzed in their applications are tensile strength, yield strength, elastic modulus, corrosion and fatigue resistance, surface finish, creep, and hardness.

- **Resistance to implant wear and aseptic loosening**

Implant wear and aseptic loosening are very important failure problems that should be taken into consideration when dealing with long-term prosthetic devices.

- **Corrosion resistance**

Corrosion of metallic implants that occur within the human body constitutes an ion source that may potentially affect the local and systemic host environment. Therefore, an important property that must be considered is the corrosion resistance of the metallic implants.

Corrosion is one of the major processes that cause problems when metals and alloys are used as implants in the body. To minimize these problems, better understanding of some of the basic principles involved in the degradative process of corrosion is required. Corrosion of implants in the aqueous medium of body fluids takes place via electrochemical reactions (Shreir 1994) and it is necessary to appreciate and understand the electrochemical principles that are most relevant to the corrosion processes. The electrochemical reactions that occur on the surface of the surgically implanted alloy are identical to those observed during exposure to seawater (namely, aerated sodium chloride). The metallic components of the alloy are oxidized to their ionic forms and the dissolved oxygen is reduced to hydroxyl ions. During corrosion process, the total rates of oxidation and reduction reactions that are termed as electron production and electron consumption respectively, must be equal. The overall reaction rate is controlled by the slowest step of these two processes. The metals and alloys used as surgical implants achieve passivity by the presence of a protective surface passive film. This film inhibits corrosion and keeps current flow and the release of corrosion products at a very low level i.e., all the implantable materials undergo corrosion at some finite rate due to complex corrosive environment of the body, while in use [7].

The types of corrosion that are pertinent to the currently used alloys are pitting, crevice, galvanic, intergranular, stress-corrosion cracking, corrosion fatigue and fretting corrosion.

Biomaterial types in orthopaedics

It is important for orthopaedic surgeons to understand the nature of biomaterials, their structural configurations, and their properties, as well as the effects of their interaction with soft and hard tissues, blood, and intra- and extracellular fluids of the human body. The orthopaedics field has benefited from the great efforts of many orthopaedic surgeons, experimental surgery laboratories, and research centres and from research work at universities, academies, societies, scientific organizations, and many interdisciplinary groups. However, many challenges remain to be conquered in the development of new biomaterials that will improve the long-term performance of clinical results in orthopaedic surgery. The main biomaterials used in orthopaedic surgery are divided into two groups: metals and non-metals.

Table 1: Materials used for implants and their mechanical properties

METALS/ ALLOYS	COMPOSITION (wt%)	DENSITY ($g.cm^{-3}$)	YIELD STRENGTH (Mpa)	YOUNGS MODULUS (Gpa)
BONE		2.0	130	17
316 L STAINLESS STEEL	Fe 16-18.5Cr 10-14Ni 2-3Mo <2Mn <1Si <0.003C	8.0	190	193
Co – Cr - Mo ALLOY	Co 33-37Ni 19-21Cr 9-10.5Mo	8.5	142	232
TI – 6Al – 4V	Ti 5.5-6.75Al 3.5-4.5V	4.40	795	116
UHMWPE		1.1	30	1

- **Metals**

The use of metals in therapeutic procedures dates back several centuries. Metallic implants were used in the 17th century. In the 18th century a metal screw implant was used for the first time. The majority of elements in the periodic table are metals. Metallic biomaterials have their main applications in load-bearing systems such as hip and knee prostheses and for the fixation of internal and external bone fractures. It is very important to know the physical and chemical properties of the different metallic materials used in orthopaedic surgery as well as their interaction with the host tissue of the human body. The metallic implants most widely used in orthopaedic surgery are:

• Low carbon grade austenitic stainless steels: 316L

• Titanium and titanium-base alloys: commercially pure titanium (CP Ti), Ti-6Al-4V, and other titanium-base alloys

• Cobalt alloys: Co-Cr-Mo, and other cobalt-base alloys

- **Polymers**

Polymers are organic materials that form large chains made up of many repeating units. Polymers are extensively used in joint replacement components. Currently the polymers most widely used in joint replacements are:

• Ultrahigh molecular weight polyethylene (UHMWPE)

• Acrylic bone cements

• Thermoplastic polyether ether ketone (PEEK)

• Bioabsorbable

- **Ceramics**

Ceramics are polycrystalline materials. The great majority are compounds made up of metallic as well as non-metallic elements; they generally have ionic bonds or ionic with some covalent bonds. The main characteristics of ceramic materials are hardness and brittleness. They work mainly on compression forces; on tension forces, their behaviour is poor. The main ceramics in orthopaedic surgery and their applications are:

• Alumina, Al_2O_3, used for acetabular and femoral components

• Zirconia, ZrO_2, used for acetabular and femoral components

• Hydroxyapatite, $Ca_{10}(PO_4)_6(OH)_2$, used for coating stem femoral components to integrate the surface material to the bone

- **Composites**

Composite biomaterials are made with a filler (reinforcement) addition to a matrix material in order to obtain properties that improve every one of the components. This means that the composite materials may have several phases. Some matrix materials may be combined with different types of fillers. Polymers containing particulate fillers are known as particulate composites.

The following composites are considered in the orthopaedic devices:

• Fiber-reinforced polymers

• Aggregates to polymethyl methacrylate (PMMA)

- **Shape memory alloys**

The term *shape memory alloys* (SMA) is applied to that group of metallic materials that demonstrates the ability to return to some previously defined shape or size when subjected to the appropriate thermal procedure. Generally, these materials can be plastically deformed at some relatively low temperature and, on exposure to some higher temperature, will return to their shape prior to the deformation. Materials that exhibit shape memory only on heating are referred to as having a one-way shape memory. Some materials also undergo a change in shape on re-cooling. These materials have a two-way shape memory.

Although a relatively wide variety of alloys is known to exhibit the shape memory effect, only those that can recover substantial amounts of strain or that generate significant force on changing shape are of commercial interest. To date, this has been the NiTi alloys and copper base alloys such as Cu-Zn-Al and Cu-Al-Ni. Of these, only NiTi alloys have been used for biomedical devices [8]. A shape memory alloy is further defined as an alloy that yields a thermoelastic martensite. In this case, the alloy undergoes a martensitic transformation of a type that allows the alloy to be deformed by a twinning mechanism below the transformation temperature. The deformation is then reversed when the twinned structure reverts on heating to the parent phase.

Applications

Biomaterials are used in Joint replacements, bone plates, bone cement, artificial ligaments and tendons, dental implants for tooth fixation, blood vessel prostheses, heart valves, skin repair devices, cochlear replacements, contact lenses, etc.

Biomaterial devices used in orthopaedics are commonly called *implants*; these are manufactured for a great number of orthopaedic applications. Biological materials such as human bone allografts (transplants of tissue between genetically different individuals) are considered to be biomaterials because they are used in many cases in orthopaedic surgery [3].

Metallic biomaterials are exploited due to their inertness and structural functions; they do not possess bifunctionalities like blood compatibility, bone conductivity and bioactivity. Hence, surface modifications are required. Improving their bone conductivity has been done by coating with bioactive ceramics like hydroxyapatite or blood compatibility by coating with biopolymers Nowadays, large number of metallic biomaterials composed of nontoxic and allergy-free elements are being developed. Even more, a new type of biodegradable metals has been proposed as temporary implants.

Generally, all metal implants are non-magnetic and high in density. These are important for the implants to be compatible with magnetic resonance imaging (MRI) techniques and visible under X-ray imaging. Most of artificial implants are subjected to loads, either static or repetitive, and this condition requires an excellent combination of strength and ductility. This is the superior characteristic of metals over polymers and ceramics. Specific requirements of metals depend on the specific implant applications. Stents and stent grafts are implanted to open stenotic blood vessels; therefore, it requires plasticity for expansion and rigidity to maintain dilatation. For orthopaedic implants, metals are required to have excellent toughness, elasticity, rigidity, strength and resistance to fracture. For total joint replacement, metals are needed to be wear resistance; therefore, debris formation from friction can be avoided. Dental restoration requires strong and rigid metals and even the shape memory effect for better results.

In overall, the use of biomaterials in clinical practice should be approved by an authoritative body such as the FDA (United States Food and Drug Administration). The proposed biomaterial will be either granted Premarket Approval (PMA) if substantially equivalent to one used before FDA legislation of 1976, or has to go through a series of guided biocompatibility assessment.

Bone composition

Bone is comprised of two main parts: the cells and intercellular matrix. As approximately 20% of bone is water, the remaining dry material consists of organic and inorganic substances. The organic parts, which make up approximately 35% of the dry material of bone, consist of collagen, proteins, and glycosaminoglycans (Pilitsis, 2003). The remaining 65% of the dry material of bone is the inorganic material, which largely consists of hydroxyapatite that contains calcium phosphate, calcium carbonate and traces of fluoride (Petchey). The cellular components of bone consist of osteoprogenitors, osteocytes, osteoblasts, and osteoclasts [9].

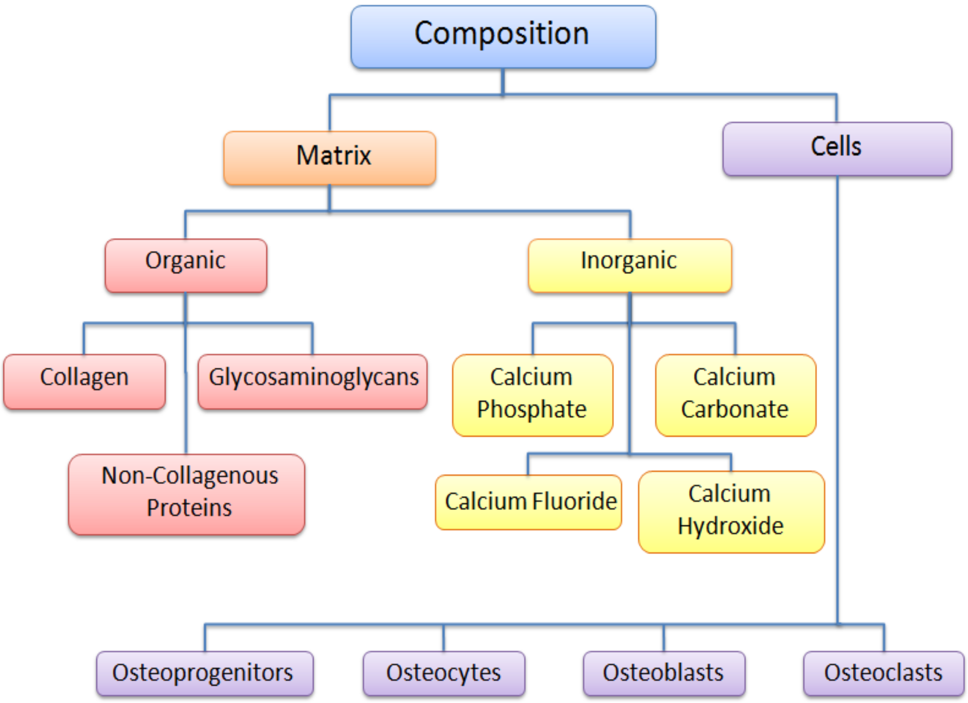


Figure 1: Composition of bone

Orthopaedic implants

Orthopaedic implants can be divided into four main groups:

• Osteosynthesis (stabilization and fixation of bone)

• Joint replacements

• Nonconventional modular tumour implants

• Spine implants

- **Osteosynthesis**

In 1949, Professor Robert Danis M.D. of the Brussels Faculty of Medicine, published his book *Théorie et Practique de L'ostéosynthèse*. His main and most remarkable contribution was the rigid fracture fixation by compressive forces of the main bone fragments previously reduced and then mechanically stabilized with a plate, resulting in an early bone fracture consolidation.

The evolution of modern osteosynthesis started with the publication of Professor Danis. He made great contributions to the scientific development of internal bone fracture fixation. The main implants used in osteosynthesis are screws, plates, nails, and pins, in a number of different shapes and forms to fulfil the required characteristics to successfully consolidate internal and external bone fracture fixation. Relevant contributions of many orthopaedic surgeons have extended their applications, including stabilizing multi-traumatized patients and correcting deformities and longitude discrepancies, among others. Patient diagnosis, surgical technique, and application of biomechanical and biomaterials knowledge are the fundamental aspects to achieve bone.

- **Joint replacements**

Prosthetic devices are implanted in the human body to replace the affected joint in order to eliminate pain and restore its normal function. This manuscript considers mainly hip and knee joint replacements because they are by far the most widely used in orthopaedic surgery. It is well known that femoral stem joint replacements have a mean useful life which, among other factors, is intimately linked to wear particles.

- **Hip joint replacements**

In the first half of the 20th century, a total hip replacement was designed and used in patients; however, the initial results were not completely satisfactory. The main concerns at that time, besides the implant design, were the surface bearing materials of the metal-on-metal and the metal-on-polymer femoral-acetabular component types. Also, methods for implant fixation (cemented and cementless femoral stem components) needed to be established.

Sir John Charnley did not use the metal-on-metal femoral acetabular component because of frictional torque in the bearing of metallic surfaces. In 1962, he found a high-density polyethylene to be a more adequate bearing surface. For the femoral stem fixation, Dr. Charnley used PMMA bone cement and finger packing as the cement insertion technique. The postoperative problems were the femoral stem subsidence.

In 1979, Carl Zweymüller, M.D., started to use a cementless tapered titanium femoral stem. A great number of total hip replacements, cemented and cementless prosthetic devices, have been developed since the relevant early design of Dr. Charnley's total hip replacement prosthetic device.

The femoral-acetabular component types currently used are:

• Metal-on-polyethylene

• Metal-on-metal

• Ceramic-on-polyethylene

• Ceramic-on-ceramic

An excellent publication that considers cemented, cementless, and hybrid implants with different designs is *The Swedish Total Hip Replacement Register.* It incorporates very important clinical results of implant survival in patients with an index diagnosis of osteoarthritis, with revision due to aseptic loosening as the end point. Currently, persistent problems remain to be solved with total hip replacements, including implant wear, aseptic loosening, and osteolysis.

- **Knee joint replacements**

There are two types of the knee joint replacement: total and unicondylar.

Types of total knee replacements are:

• Non-constrained knee replacements

• Semi-constrained knee replacements

• Constrained knee replacements

The unicondylar knee replacement is usually called half replacement. It is recommended when half of the damaged joint is to be replaced. The implant biomaterials used in total knee replacements are titanium base alloys, cobalt-chromium alloys, ceramics, and cross-linked ultrahigh molecular weight polyethylene. The improvements on implant materials and manufacturing processes have made great contributions to the long-term performance of these prosthetic devices.

- **Nonconventional modular tumour implants**

The fundamental objective in cases of oncology orthopaedics is the preservation of the affected parts. In these cases, the use of bone allografts, the tumour modular prosthesis, or a combination of both is usually required.

The main nonconventional modular tumour implants are:

• Nonconventional modular tumour implants with femoral replacement

• Nonconventional modular tumour implants with tibia replacement

For cases of children and adolescents who are still growing and have bone sarcoma tumours in the lower extremities, and where there is a length discrepancy, both non-invasive or invasive extendible tumour modular prosthesis can be used.

The technological development of orthopaedic implants has been extremely important in the successful results of many clinical cases. However, there is still a lot of work left to reach the expected results, which means the continuous search for new and better biomaterials to satisfy the day-to-day needs in orthopaedic surgery.

- **Spine implants**

Thanks to the instruments designed by Paul Harrington M.D., modern spine surgery started in 1950. From then on, it has been used to correct neuromuscular defects, especially those associated with poliomyelitis. Since the beginning of the 21st century there has been an increase in the manufacturing of sophisticated implants to treat different spine pathologies.

International standards for orthopaedic devices

The preparation and publication of material standards is an important activity. Test protocols were established by different institutions in different countries such as ASTM International and the International Organization for Standardization (ISO). The main efforts of all institutions are directed to obtain biocompatible and inexpensive orthopaedic devices, as well as safe and affordable material that responds to the medical requisites of the surgeon and, most important, to the needs of the patient. Many publications, worldwide conferences, and society and academy meetings have presented clinical results of implants used in orthopaedic surgery. The final results indicate a growing interest and considerable activity in the extraordinary and expanding field of biomaterials in orthopaedic surgery [9].

- **ASTM standards for orthopaedic devices**

The ASTM Committee F04 on Medical and Surgical Materials and Devices was created in 1962. The Committee, with a membership of approximately 880 members, currently has jurisdiction of over 250 standards, published in the *Annual Book of ASTM Standards,* Volume 13.01. Committee F04 has 34 technical subcommittees that have jurisdiction over these standards [7].

The technical subcommittees of F04 collectively encompass the following five primary areas:

- *Resources:* addresses standards for materials such as ceramics, metals, and polymers; it also includes standards to address the information needed on biocompatibility, test methodology, and magnetic resonance imaging.
- *Orthopaedic Devices:* focuses on methods and practices for osteosynthesis, arthroplasty, and spinal devices.
- *Medical and Surgical Devices:* pertain mostly to cardiology, neurology, audiology, gastroenterology, and plastic surgery.
- *Tissue Engineered Medical Products (TEMPs):* focuses on materials needed in, and practices and methods for, the development and applications of TEMPs technologies.
- *Computer-Assisted Orthopaedic Surgical Systems (CAOS):* write standards for system accuracy.

- **ISO Standards for Orthopaedic Devices**

The International Organization for Standardization, known as ISO, is an International-standard-setting body composed of representatives from various national standards organizations. ISO was founded on February 23, 1947, and its headquarters is in Geneva Switzerland. The word ISO is based on the Greek word isos meaning equal and is applied for any country [7].

Requirements of biomaterials: general issue and concerns

An implant should possess some important properties in order to long-term usage in the body without rejection. The design and selection of biomaterials depend on their mechanical and non- mechanical characteristics.

- **Mechanical properties**

The mechanical properties such as hardness, tensile strength, modulus, elongation (strain), fracture resistance and fatigue strength or life play an important role in material selection for application in the human body.

- **Long fatigue life**

The fatigue strength is related to the response of the material to the repeated cyclic loads, fatigue fracture leads some of major problems associated with implant loosening, stress-shielding and ultimate implant failure and it is frequently reported for hip prostheses. Fatigue characteristics are strongly depending on the microstructures. The microstructures of metallic biomaterials alter according to the processing and heat treatment employed.

- **Adequate Strength**

Strength of materials from which the implants are fabricated has influence the fracture of artificial organ. Inadequate strength can cause to fracture the implant. When the bone implant interface starts to fail, developing a soft fibrous tissue at the interface can make more relative motion between the implant and the bone under loading. This fact causes pain to the patient and after a certain period, the pain becomes unbearable and the implant must be replaced, by a revision procedure.

- **Modulus equivalent to that of bone**

For major applications such as total joint replacement, higher yield strength is basically coupled with the requirement of a lower modulus close to that of human bones. The magnitude of bone modulus varies from 4 to 30 GPa depending on the type of the bone and the measurement direction emphasized about the modulus and described that the large difference in the young's modulus between implant material and the surrounding bone can contribute to generation of severe stress concentration, namely load shielding from natural bone, that may weaken the bone and deteriorate the implant/bone interface, loosening and consequently failure of implant. The modulus is considered as a main factor for selection of materials

Non-mechanical requirements

In addition to the above-mentioned mechanical properties, some non-mechanical requirements which have significant role in performance of the material in the human body are describe here. High corrosion resistance Singh & Dahotre researched on corrosion resistance as an important issue in selection of metallic biomaterials because the corrosion of metallic implants due to the corrosive body fluid is unavoidable. The implants release undesirable metal ions which are non-biocompatible. Corrosion can reduce the life of implant device and consequently may impose revision surgery. In addition, the human life may be decreased by the corrosion phenomenon.

- **High wear resistance**

The low wear resistance or high coefficient of friction results in implant loosening. Wear debris are found to be biologically active and make a severe inflammatory response that led to the destruction of the healthy bone which supports the actual implant. Corrosion caused by friction is a big concern since it releases non compatible metallic ions. It should be pointed out that mechanical loading also can result in corrosion fatigue and accelerated wear processes.

- **Biocompatibility**

One of the most important non-mechanical requirements of orthopaedic biomaterials is the biocompatibility. Biocompatibility is the ability to exist in contact with tissues of the human body without causing an unacceptable degree of harm to the body. It is not only associated to toxicity, but to all the adverse effects of a material in a biological system. Navarro et al supported the study of Smallman and Bishop and with retrospect to the last 60 years, categorized three generations for evolution of biomaterials: bioinert materials, bioactive and biodegradable materials and materials designed to stimulate specific cellular responses at the molecular level.

Bioinert is related to reduce the body reaction to the implant to a minimum. Bioactivity defined as the ability of the material to interact with the biological environment to enhance the biological response. The third generation refers to the capability of the material to stimulate specific cellular responses at the molecular level. Williams defined the biomaterial requirements of total joint replacements in terms of biocompatibility as, optimizing the rate and quality of bone opposition to the material, minimizing the release rate of corrosion and the tissue response to the released particles, minimizing the release rate of wear debris and the tissue reaction to this debris and optimizing the biomechanical environment in order to minimize disturbance to homeostasis in the bone and surrounding soft tissue.

- **Osseointegration**

Osseointegration is fundamental in orthopaedic. Several literatures explained about the integration of the implant with adjacent bone and tissue. Osseointegration defined as the process of formation of new bone and bone healing. The incapability of an implant surface to join with the adjacent bone and other tissues due to micromotions, results in formation of a fibrous tissue around the implant and promote loosening of the prostheses. Thus, materials with a proper surface are extremely essential for the implant to integrate well with the surrounding bone. Surface chemistry, roughness and topography are all parameters that influence both the osseointegration and biocompatibility. It should be considered that in addition to properties of the implanted biomaterial, the characteristics and regenerative capability of the host bone affect the osseointegration of biomaterials.

Metallic biomaterials

Implants are fabricated from a wide variety of materials, including metals, polymers, ceramics and their composites. Among these materials, metals are an important group, for instance knee implant has some metallic parts. Current total knee replacement mainly has three components: femur, tibia (includes tibial tray and tibial insert) and patella or kneecap. The tibial insert and the patellar components are usually made of plastics such as ultra-high molecular weight polyethylene (UHMWPE) or cross-linked polyethylene.

The femoral component and tibial tray are metallic parts and tend to be made of titanium alloys, stainless steel or cobalt chromium with small amount of molybdenum (Co-Cr-Mo). Another application of metals is in design of hip joint implant which includes an alloy femoral stem (Ti alloy) with a metallic or ceramic femoral head moving in an acetabular cup that is normally made of UHMWPE. In the following, metallic biomaterials are divided to current and promising materials and described respectively. Current Metallic Biomaterials such as Stainless steel, Co-Cr alloys and Ti alloys are the current metals used in orthopaedics application such as knee and hip implant.

- **Stainless steel**

Stainless steel is the generic name for a number of different steels used primarily because of their resistance to a wide range of corrosive agents due to their high Cr content. The Cr in the stainless steel has a great affinity for oxygen, which allows to the formation of a film of chromium oxide on the surface of the steel at a molecular level which is passive, adhesive, tenacious and self-healing. Stainless steel implants are often degraded due to pitting, crevice, corrosion fatigue, fretting corrosion, stress corrosion cracking, and galvanic corrosion in the body. Their corrosion resistance can be modified by lowering the nickel content and alloying them with Mn or N. The wear resistance of austenitic stainless steel is relatively poor. So rapid loosening is generated by the large number of wear debris. Worse corrosion resistance as well as the danger of allergic reaction which appears in a big number of patients restricts their application in orthopaedic joint prosthesis. Moreover, the modulus of stainless steel is about 200 GPa which is much higher than that of bone [10].

Stainless steel has several types and the most mainly used for manufacturing implants is austenitic stainless steel. Stainless steel 316L is the type widely used in traumatological temporary devices such as fracture plates, screws and hip nails. Stainless steel has been used for wide range of application due to easy availability, lower cost, excellent fabrication properties, accepted biocompatibility and great strength.

- **Cobalt-Chrome**

Cobalt chromium alloys can be basically categorized into two types; one is the Co-Cr-Mo alloy (which is usually used to cast a product) and the second one is Co-Ni-Cr-Mo alloy, (which is usually wrought by hot forging). The castable Co-Cr-Mo alloy has been used in dentistry for long time and recently in making artificial joints. The wrought Co-Ni-Cr-Mo alloy is a relative newcomer material which is now used for making the stems of prosthesis of heavily loaded joints such as the knee and hip. Cobalt-based alloys are highly resistant to corrosion even in chloride environment due to spontaneous formation of passive oxide layer within the human body environment, these materials have superior mechanical properties such as high resistance to fatigue and cracking caused by corrosion with a good wear resistance. Also, they are not brittle because they have a minimum of 8% elongation. These materials have a high elastic modulus (220–230 GPa) similar to that of stainless steel (approx. 200 GPa) which is higher than that of cortical bone (20–30 GPa). Elements such as Ni, Cr and Co are indicated to be released from the stainless steel and cobalt chromium alloys due to the corrosion in the human body. It has been found that Ni, Cr and Co are the most toxic ions. The corrosion products of Co-Cr-Mo are more toxic than those of stainless steel 316L. The thermal treatments used to Co-Cr-Mo alloys modifies the microstructure of the alloy and alters the electrochemical and mechanical properties of the biomaterial.

- **Ti and Ti alloys**

Manufacturing of titanium implants date back to the late 1930s. There are three structural types of titanium alloys: Alpha (α), Alpha-Beta (α-β) or metastable β and Beta (β). The β phase in Ti alloys tends to exhibit a much lower modulus than α phase, and also it satisfies most of the other necessities or requirements for orthopaedic application. Ti alloys due to the combination of its excellent characteristics such as high strength, low density (approx. 4700 Kgm-3), high specific strength, good resistance to corrosion (due to the formation of an adhesive TiO2 oxide layer at the surface), complete inertness to body environment, enhanced biocompatibility, moderate elastic modulus of approximately 110 GPa are a suitable choice for implantation. Ti and its alloy also have this ability to become tightly integrated into bone. This high capacity to join with bone and other tissues considerably improves the long-term behaviour of the implanted devices, decreasing the risks of loosening and failure. Typically, good clinical outcome from rough surfaces of Ti and its alloy is resulted when compared to smooth-surfaced implants due to the good osseointegration between the bone and the implant.

Commercially pure Ti (CP Ti) and Ti–6Al–4V ELI (Ti64, Extra Low interstitial) are most commonly used titanium materials for implant applications. Ti-6Al-V4 is slowly replacing CP Ti due to the greater mechanical strength. According to Navarro et al.and Geetha et al, long-term performance of titanium and its alloys mainly Ti64 has raised some concerns because of releasing aluminium and vanadium.

Both Al and V ions are associated with long-term health problems, like Alzheimer disease and neuropathy. Furthermore, when titanium is rubbed between itself or between other metals, it suffers from severe wear. High friction coefficient and a rather high propensity to seizure are attributed to this alloy. Therefore, their application is limited to the locations on the implant surface where wear resistance is not of vital importance. For example, the poor wear resistance of titanium alloys avoids their use for femoral head applications in typical hip implant although the femoral stem is often made of these alloys. Rostoker & Galante measured wear rates by in vitro methods on UHMW-PE when rubbed by a oppose face of the Ti-6% AI-4% V, ELI grade (ASTM F-136) with various prepared surfaces and compared the result with stainless steel or cast Co-Cr-Mo alloy.

Two recently developed promising biomedical alloys, Ti- 35Nb-7Zr-5Ta (TNZT) and Ti-29Mo-13Ta- 4.6Zr (TNTZ), show significant improvement, in the aspect of accompanying the high yield strength and low modules, compared to previous generation alloys such as Ti-6Al-4V, stainless-steel and cobalt– chromium-based alloys .

The applications of Ti and its alloys include dental implants and parts for orthodontic surgery, joint replacement components such as knee and hip, bone fixation devices like nails, screws and plates, artificial heart valves and surgical instruments.

- **Promising materials**

Shape memory alloys (SMA) provide new insights for the design of biomaterials [9] for artificial organs and advanced surgical instruments, since they have unique characteristics and superior properties. Some unusual properties of these materials are: one-way and two-way shape memory effects, super elastic effect, high damping property and rubber-like effect.

- **316L stainless steel**

Stainless steel was first used successfully as an important material in the surgical field.

I- Type 302 stainless steel was introduced, which is stronger and more resistant to corrosion than the vanadium steel;

II- Type 316 stainless steel was introduced, which contains a small percentage of molybdenum (18-8sMo) to improve the corrosion resistance in chloride solution (salt water);

III- Type 316L stainless steel. The carbon content was reduced from 0.08 to a maximum amount of 0.03% for better corrosion resistance to chloride solution.

The inclusion of molybdenum enhances resistance to pitting corrosion in salt water. Even the 316L stainless steels may corrode in the body under certain circumstances in highly stressed and oxygen depleted region, such as the contacts under the screws of the bone fracture plate. Thus, these stainless steels are suitable to use only in temporary implant devices, such as fracture plates, screws, and hip nails.

The *austenitic stainless steels,* especially type 316 and 316L, are most widely used for implant fabrication. These cannot be hardened by heat treatment but can be hardened by cold-working. This group of stainless steels is nonmagnetic and possesses better corrosion resistance than any others. The inclusion of molybdenum enhances resistance to *pitting corrosion* in salt water. The American Society of Testing and Materials (ASTM) recommends type 316L rather than 316 for implant fabrication. The specifications for 316L stainless steel are given in the table. The only difference in composition between the 316L and 316 stainless steel is the maximum content of carbon, i.e., 0.03% and 008%, respectively.

Table 2: General chemical composition of 316L stainless steel

Element	**Composition (%)**
Carbon	0.03 max
Manganese	2.00 max
Phosphorus	0.03 max
Sulphur	0.03 max
Silicon	0.75 max
Chromium	17.oo - 20.00
Nickel	12.00 -14.00
Molybdenum	2.00 – 4.00

The nickel stabilizes the austenitic phase [γ, face cantered cubic crystal (FCC) structure], at room temperature and enhances corrosion resistance. The austenitic phase formation can be influenced by both the Ni and Cr contents for 0.10% carbon stainless steels. The minimum amount of Ni for maintaining austenitic phase is approximately 10%.

Table 3: General mechanical properties of 316L stainless steel

Condition	**U T S (MPa)**	**Yield Strength (0.2% offset) (MPa)**	**Elongation, 2 in (50.8 mm) %**	**Rockwell Hardness (HRB)**
Annealed	485	172	40	95
Cold worked	860	690	12	-

A wide range of properties exists depending on the heat treatment (annealing to obtain softer materials) or cold working (for greater strength and hardness). Figure 2.3 shows the effect of cold working on the yield and ultimate tensile strength of 18-8 stainless steels. The engineer must consequently be careful when selecting materials of this type. Even the 316L stainless steels may corrode inside the body under certain circumstances in a highly stressed and oxygen depleted region, such as the contacts under the screws of the bone fracture plate. Thus, these stainless steels are suitable to use only in *temporary* implant devices such as fracture plates, screws, and hip nails. Surface modification methods such as anodization, passivation, and glow-discharge nitrogen-implantation, are widely used in order to improve corrosion resistance, wear resistance, and fatigue strength of 316L stainless steel

Carbon at high temperatures forms $M_{23}C_6$ type carbides (M is Cr, Mo, Fe or the like) and MC- type carbides at the grain boundaries. This process results loss of Cr along the areas adjacent to the grain boundary, thus the regions get depleted in Cr content causing its susceptibility to corrosion. This process is called sensitization. For preventing this sensitization carbon content is lowered to about 0.03wt%. Addition of Nitrogen facilitates solid solution strengthening thereby compensating the loss in strength due to the decrease in C content.

Coatings

A biometric coating, which has reached a significant level of clinical application, is the use of HA as a coating on porous metal surfaces for fixation of orthopaedic prostheses. This approach combines biological and bioactive fixation. Though a wide range of methods have been used to apply the coating, plasma spray coating is usually preferred. The table below lists the bio ceramic coatings:

Substrate	Coating
316L stainless steel	Pyrolityc carbon
316L stainless steel	45S5 bioglass
316L stainless steel	αAl_2O_3-HA-Tin
316L stainless steel	HA
Co-Cr alloy	45S5 bioglass
Co-Cr alloy	HA
Ti-6Al-4V alloy	45S5 bioglass
Ti-6Al-4V alloy	HA
Ti-6Al-4V alloy	Al_2O_3
Ti-6Al-4V alloy	HA/ ABS glass (Alkali Borosilicate Glass)

- **Ion implantation**

Ion implantation is a process which involves the introduction of a small and economical number of atoms of any element into the surface of the any material with a beam of high velocity ions, without modifying the surface finish or the bulk properties of the underlying material and are independent of thermodynamic constraints. The bulk alloying techniques are limited by the equilibrium phase diagrams. Thus, if a potential alloying addition does not the surface related properties such as corrosion and wear resistance properties. Surface alloying by conventional techniques improve the corrosion resistance of metals with intrinsically inferior properties. Ion implantation is a versatile technique which at a controlled rate produces metastable-single phase-solid solution-surface alloys without any of the compositional limitations normally imposed by equilibrium phase diagrams.

Nitrogen ion implantation of stainless steels: Among the various ions implanted, nitrogen ion implantation is the most suitable technology for biomedical applications. Properties such as hardness, corrosion resistance, wear etc. can be improved without adversely affecting the bulk properties of the materials.

- **Hydroxyapatite coatings on type 316L SS**

Bioactive ceramics are dense calcium phosphate-based ceramics with a composition and structure similar to that of inorganic components of bone (Hench & Ethridge 1982). Of the various calcium phosphate ceramics, Hydroxyapatite Ca10.PO4⁄6.OH⁄2 (HAP) is the vital constituent present in bones and teeth. HAP is the most versatile material used for implantation purposes owing to their similarity with natural bone mineral and its ability to bond to bone. These materials are characterised by a certain solubility, which provokes the surrounding bone or tissue to form direct bonding to the implant. The solubility leads to gradual degradation and resorption by the surrounding tissue which stimulates the bone to grow on the material and through its pores, and in some cases, it is believed to cause total transformation of the material into living bone (Von Recum 1999). This bonding is able to transfer shear and tensile stress along the interface, that could be an advantage in anchoring the implants and reducing the stress peaks in the bone. The main restriction of these materials lies in their low strengths so that they can be used as bulk materials only for low loaded devices. The biological response to HAP has been characterised by clinical and laboratory studies. HAP granules are used as fillers in large bone defects after resection of bone tumours.

HAP is also known to have a simulating effect on bone formation, which is known as Osseo induction. It enhances the osteointegration, and there are indications that chemical bonding may occur between HAP and bone. These attractive features of HAP are offset by the lack of strength necessary for load bearing applications. Therefore, to combine the bioactivity of HAP and the strength of the materials used in orthopaedic implants, it can be applied as coatings [12]. Among the various modifications, the coating technologies have emerged as a viable process and have opened up a new possibility for implant and prosthetic devices. Ceramic coated metal implants for prosthetic applications provide the necessary porosity for bone ingrowth, while the underlying metal substrate bear the load and the full weight bearing capacity is ensured soon after surgery. Thus, bio ceramics play a twin (dual) role both in preventing the release of metal ions (rendering it more corrosion resistant) and also in making the metal surface bioactive. of non-stoichiometric phases of the powder. A number of novel methods for coating HAP have been proposed offering the potential for better control of film structure such as hot isostatic pressing, flame spraying, ion beam deposition, laser ablation and electrochemical deposition along with plasma spraying which has been widely studied over the decade.

The major problems associated with plasma spraying process are that it is a line-of-sight process that produces non-uniform coating with heterogeneous structure and the high temperature involved alters the HAP and metal substrate phases. Hence, electrophoretic deposition of HAP on metal substrates has been used to overcome the above drawbacks and to achieve the uniform distribution of fine HAP deposits.

The advantages of this technique include high purity of layers formed, ease of obtaining the desired thickness and high layer adhesion to the substrate. Hence, it is advantages to attempt to develop thin layers of HAP on the surface of type 316L SS by electrophoretic deposition and to studying their electrochemical properties for applications as orthopaedic devices.

References:

1. Fundamentals Handbook - Material Science
2. Handbook of Materials for Medical devices
3. Austenitic Stainless Steels. Copyright © 2008 ASM International.
4. Stainless Steels for Design Engineers (#05231G)
5. Sydney H Avner: Introduction to physical metaluurgy.
6. Biomaterials by Sujata kumar
7. Introduction to Biomaterials in Orthopaedic Surgery. 2009 ASM International. (#05233G)
8. A brief rewiew: Biomaterial and their application; Amol Tathe, Mangesh Ghodke and Pratima Nikalje; International Journal of Pharmacy and Pharmaceutical Sciences, ISSN-0975-1491, Vol 2, Suppl 4, 2010
9. Biomaterials in Orthopaedic Surgery, 2009 ASM International (#05233G)
10. P. Atanda, A. fatudimu and O. Oluwole: Sensitisation study of normalized 316L stainless steel.
11. ASTM Standard E 8/E 8M – 08, Standard Test Methods for Tension Testing of Metallic Materials, March 2008, pp. 1-25.
12. Brendamari Rodríguez, Annette Romero, Omar Soto and Oswaldo de Varorna: Biomaterials for orthopedics.

A REVIEW OF THE INTERPRETATION OF DUALITY IN DIFFERENTIABLE AND NON-DIFFERENTIABLE MATHEMATICAL PROGRAMMING

Ram Naresh Singh*[1], Shailendra D. Deo[2] and Ajay Kumar Sharma*[2]

[1]Department of Mathematics,

Sardar Patel Mahavidyalaya Chandrapur - 442402 (MS), India

[2]Mahatma Gandhi College of Science,

Gadchandur, Dist - Chandrapur-442908(MS), India

*Corresponding author E-mail: drrbs79@gmail.com, ajaykumarsharma101010@gmail.com

Abstract:

The existence of an ideal solution to either the primal problem or the dual problem ensures the existence of an ideal solution to the other because these two projects are linked by the duality principle. If the primary issue is constrained minimization (or maximisation), then the dual issue is a constrained maximisation (or minimization) problem. The duality effects have shown to be very helpful in enhancing numerical techniques for addressing particular categories of optimisation problems. Because it offers appropriate halting guidelines for primary and secondary problems, The presence of duality theory in nonlinear programming problems makes it easier to create numerical algorithms. The inclusion of duality theory in nonlinear programming issues aids in the development of numerical algorithms. If the original issue is the dual of the dual, then a nonlinear programming problem and its dual are said to be symmetric. This postulation's major objective is to take into account optimality and duality in a range of mathematical programming issues, with a particular emphasis on non-differentiable nonlinear programming and variational issues, such as blended sort symmetric and self duality. Non-differentiable fractional minmax programming, continuous-time minmax programming, minimaxvariational problems, and continuous-time minmax programming.

Overview

In the late 1940s, mathematical programming achieved the status of a logical science unto itself, and ever since then, it has made enormous strides. It is currently regarded as one of the most dynamic and invigorating disciplines of contemporary mathematics, with numerous applications in a variety of contexts, including design, financial matters, and basic sciences. Mathematical programming problems frequently include finding the least-weight design of a structure that is bound to stress and deflection restrictions.

A mathematical programming problem has the following structure:

(MP): Maximize or minimize $f(x)$.

Depending on

$g_i(x) \leq 0, i = 1,2,3 \dots \dots, m$

$h_j(x) = 0, j = 1,2,3 \dots \dots, k,$

$x \in$ X

The function $f$and each of its f_j and h_j on the n-dimensional Euclidean space R_n and XR_n are real valued capabilities. This is referred to as the general mathematical programming issue. Fairness constraints are defined as $h_j(x) = 0, j = 1,2,3 \dots \dots, k$, whereas inequality constraints are defined as $g_i(x) = 0, i = 1,2,3 \dots \dots, m$.The integration of x X is a conceptual constraint. We define the previous issue as a differentiable programme if the objective and critical skills are differentiable. The aforementioned issue is referred to as a convex programming problem if X is a convex set and the objective and inequality constraints are both relative to capacity.

Duality in adaptable mathematical programming

If f: Rn R and hj: Rn R, where j = 1, 2,..., m, then think about the nonlinear programming problem:

Min $f(x)$(P) according to, $h_j(x) = 0, j = 1,2,3 \dots \dots . m$.

The Lagrangian dual for issue (P) in the case of Rm is defined as

This is,M inf (u) plus $Th(u)$ (LD)

If all of the functions f and $h_j(x) = 0, j = 1,2,3 \dots \dots . m$. are differentiable convex functions, then the problem (LD) is similar to the problem below:

$Maxf(x)$+ $Th(x)$ (WD)

Depending on $Vf(x)$ plus $Th(x)$ = 0, 0 and R_m,The Wolfe type of dual for the issue (P) is superior than everything else here. When Mangasaria gives an example, he suggests that some duality theorems might not hold true if the goal function or the constraint function is a summed up convex function. This inspired Mond and Weir to create a new dual for (P) as a result.

$Maxf(x)$+ (MWD)

Depending on

$f(x)$+ + λ $Th(x)$ = 0,λ $Th(x) \geq$ 0, λ $\geq$ 0, λ $\in R_m$,

Non-differentiable mathematical programming duality

The class of non-differentiable mathematical programming issues that Mond took into account was as follows:

Min f(x) + xTBx 1/2 (NP)

depending on $h_j(x) \leq 0, j = 1,2,3 \dots \dots . m$.

In this case, B is a n X n positive semi definite (symmetric) matrix, and f and $h_j(x), j = 1,2,3 \dots \dots . m$. are twice differentiable functions from Rn to R. The functions f and hj (j = l, 2,...,

m) are anticipated to be convex functions. They established a duality theorem between (NP) and the following problem.

(ND): $[f(u) + y\,Th(u)]\ \ maxf(u) + y\,Th(u)] - uT$

according to,

$yT + f(u) + h(u) + Bw = 0$,$wTBw\ 1, y \geq 0$.

A new dual program was introduced by Chandra, Craven, and Mond along the lines of Mond and Weir:

$Maxf(u) - [uTf(u) + yTh(u)]$ (NWD)

depending on $Bw = f(u) + yTh(u)$

$yTh(u) \geq 0$,$wTBw \leq 1$,y≥ 0.

Additionally, for any possible solution to (NP) and (NWD), settled duality theorems by assuming that $f(u)$ + $wTBw$ is pseudoconvex and that yTh is quasiconvex.

Later, Mond and Schechter replaced the square root term with the standard term and, in addition, defined the non-diffemtiable nonlinear programming problems as follows:

(NP)1 Min ||Sx||p $+f(x)$

according to, $h_j(x) \leq 0, j = 1,2,3 \ldots \ldots . m$.

Here, the convex functions f and h_j $(j = 1,2,3 \ldots \ldots . m)$ are twice differentiable from Rn to R. The problem is the dual for (NP)1:

$Maxf(u) + [\ yTh(u)$+ uTS Tv (NP)

according to $F(u) + Y\ Th(u) = \|v\|, q \leq 1, y \geq 0$.

where conjugate instances for p and q are.

Symmetric duality in distinctive mathematical programming

Take a look at a capacity f(x, y) that can be differentiated in the ranges Rn and Rm. The challenges that Dantzig et al. introduced are as follows:

$Min(f(x,y) - yTy\ f(x,y)$ (SP)

Depending on

$yf(x,y) \leq 0$,$(x,y) \geq 0.Maxf(x,y) - x\ Tf(x,y)$ (SD)

Depending on $Vxf(x,y) > 0$,$(x,y) > 0$.

Additionally, the existence of a typical optimal solution to the primal (SP) and (SD) was shown when (i) f is convex in x for every y and concave in y for every x, and (ii) f, which is twice differentiable, has the property that its matrix of second partials for y is negative definite at (x_0, y_0) and also provided the information below regarding symmetric dual programming issues:

$Min(f(x,y)$ (MSP) $yTy\ f(x,y)$

depending on $f(x,y) \leq 0, x \leq 0.\ Maxf(x,y)(MSD)x\ Tf(x,y)$

depending on $xf(x,y) \geq 0, y \geq 0$.

It should be noted that whereas in both (SP) and (SD) the constraints include x 0 and y 0, only x 0 is necessary in the primal and only y 0 is necessary in the dual.

Hypothetical duality in non-differentiable mathematical programming

Suppose $f(x,y)$ =be a true esteemed continuously differentiable function in the intervals x Rn and y Rn. Chandra and Husain created the following set of symmetric dual non-differentiable algorithms to demonstrate how duality is a result of the convexity-concavity assumption applied for the bit function $f(x,y)$ =:

$Min(f(x,y) = yTyf(x,y) + (xTBx)_{1/2}$

depending on Cw −yf(x,y) ≤ 0,wTCw ≤ 1,(x,y) > 0.

$Maxf(x,y) - x$Tx, $f(x,y)$- (yT Cy)1/2 (ND)

depending on $f(x,y) -$ Bz ≤ 0, zTB z ≤ 1,$(x,y) \geq 0$.

n x m and m x m positive semi-definite matrices, respectively, make up B and C.

Chandra, Craven, and Mond showed another combination of symmetric dual non-differentiable algorithms by reducing the convexity criteria on the bit function f(x,y) to pseudoconvexity.

The issues in are as follows:

$Min\ f(x,y)$+ (xTBx)1/2−yTCz (PS)

depending on Cz = yf(x,y) = 0.When yT[y f(x,y) - Cz] > 0,zTCz ≤ 1,x ≥ 0 .

Max $f(x,y) -$ (yT Cy)1/2 + xTBw (DS)

depending on Bw + $xf(x,y) \leq 0$, If xT[$xf(x,y)$+ B w] is zero, wTB w ≤ 1,y ≥ 0.

Programming with differentiable fractions

Assume that $f, -g$ and h_j (J = 1, 2, and m) are real, esteemed, differentiable convex functions defined on the open convex set X Rn. Consider the convex-concave fractional programming issue at this point. (FP): Depending on

h_j ≤ 0, (j = 1,2,…,m)

S = x X:hj (x) 0; j = 1, 2,..., m; g(x) > 0 for every x X; and if g(x) is not

Affine, then for all x X, f(x) ≤ 0.

There are two prominent duality models for (FP), and these have been extensively discussed in the work. These are a result of Schaible, Jagannathgan, and Bector. The issue is the Bector's double [17] for (FP).

according to,

The accompanying proportionate problem (EPF) of FP is essentially represented by this Wolfe double:

EFP: Minimum according to,

In (BD), one needs that $f(u) + yTh(u) \geq 0$, if g isn't relative, in order to make the target function pseudoconvex and ensure that duality theorems are true.

Programming that is not differentiable in fractions

Mond thought about the following fundamental issue according to,

$h(x) \leq 0.$,

In this scenario, G(x) > 0 for every possible x, B and D are n x n symmetric positive semidefinite frameworks, f, g, and h are differentiable functions from Rn into R, Rm separately. Mond [9] developed suitable duality theorems under convexity assumptions on f, g, and h, described a twofold problem to the (NFP) issue, and proved necessary and sufficient conditions for the existence of an ideal solution.

After some time, Zhang and Mend found a number of conditions that had to exist in order for the optimal configuration of (NFP) to exist. Additionally, as an application of these optimality requirements, the first and second request duals indicated below were derived individually, and relevant duality theorems were also produced.

Diverse problems

A variation problem is a particular example of an optimum control problem where the state function is subordinate to the control function.

In terms of numbers, the following variational problem exists:

Depending on

The separation operator D is given by $y =$Dx $\Rightarrow x(t) = x(a) + (s)ds$ except at discontinuities, and C (I, Rn) is the space of persistently differentiable functions x: I Rn. Where f:I x Ru x R n Rm and g:I x Ru x Rn Rm are persistently differentiable functions as for each of their contentions. Valentine establishes the necessary circumstances for the presence of an external force (VP).

Final summary

But in the current context, the heightened competitiveness and customer expectations frequently require the greatest solutions, not just practical ones. It has been found that even shortening the time it takes to improve can help a corporation save money by enhancing the design process. In this manner, the optimisation theory regulates selecting the best option among a few alternatives in terms of given capacity with the minimum amount of resources that is theoretically feasible. Mathematical programming problems are a new class of problems as a result. The most effective approaches for solving problems are frequently applied in operations research and are known as mathematical programming techniques.

References

1. S. K. Mishra (2010). Second Order Mixed Symmetric Duality in Non-Differentiable Multi-Objective Mathematical Programming, Journal of Applied Analysis, Volume 13, Issue 1, Pages 117–132
2. Yadvendra Singh, S. K. Mishra, and K. K. Lal (2017). Optimality and duality for nonsmooth semi-infinite multiobjective programming with support functions, Yugoslav Journal of Operations Research 27: Number 2, 205-218.
3. S. K. Mishra & M. Jaiswal (2015), Optimality Conditions and Duality for Semi-Infinite Mathematical Programming Problem with Equilibrium Constraints, Numerical Functional Analysis and Optimization, 36:4, 460–480, DOI: 10.1080/01630563.2015.1013552,
4. Shrivastav, Santosh Kumar (2012). Duality Gap in Nonlinear Programming, International Journal of Engineering Research & Technology (IJERT), Vol. 1 Issue 6, August - 2012, 1-13
5. Ivanov, V.I. (2013). Duality in nonlinear programming. Optim Lett 7, 1643–1658. https://doi.org/10.1007/s11590-012-0512-6
6. V. Jeyakumar and G.Y. Li (2014). Optimization, 63:5, 713–733, DOI: 10.1080/02331934.2012.690760
7. Dimitris Bertsimas, Frans J. C. T. de Ruiter (2016) Duality in Two-Stage Adaptive Linear Optimization: Faster Computation and Stronger Bounds. INFORMS Journal on Computing 28(3):500-511.http://dx.doi.org/10.1287/ijoc.2016.0689

CELLULAR COMMUNICATION IN BREAST CANCER

Juily R. Atitkar
Department of Biotechnology,
Mahatma Phule A.S.C. College, Panvel, Pin.: 410206
Corresponding author E-mail: juilyatitkar1995@gmail.com

Introduction:

"Cancer" is a condition with different abnormalities in the cell, where some of the body's cells begin to divide continuously and spread into surrounding tissues. Cancerous tumors are malignant, which means they can spread into, or invade nearby tissues. In this condition, as these tumors grow, some cancer cells can separate and migrate to distant places in the body through the body fluid such as blood or the lymph system and form new tumors far from the original tumor, referred to as "metastasis" (Rakoff-Nahoum, 2006). Cancer is a genetic disease caused by changes in gene and chromosome that control the cells function, especially cell growth and division. The genetic changes most frequently occur in proto-oncogene i.e. inactive form of oncogene which further may convert into active oncogene, tumour suppressor gene which will involve in reducing mutations and tumorigenesis and DNA repair genes (Rakoff-Nahoum, 2006). Before the cancer cells form in tissue, the cells undergo abnormal changes which are called as "hyperplasia" where cells increase in number but appear normal or dysplasia where cells appear abnormal under the microscope. Hyperplasia and dysplasia may or may not become cancer. Based on the site of origin of the cancer cells tumor terminology is characterized as a carcinoma (epithelial cells), sarcoma (bone and soft tissue), leukaemia (blood), lymphoma (lymph node), melanoma (melanocyte), brain and spinal cord (astrocyte) (Rakoff-Nahoum, 2006).

The normal mammary gland consists of a branching ductal structure embedded in a fatty, collagenous stroma. The ducts, ending at the nipple are progressively smaller in diameter. From the nipple they terminate in a group of alveoli which will form the main milk producing units during lactation. Mammary gland ducts comprise two type of cell: luminal cells line which form the ducts which is responsible for producing milk during lactation in response to cellular signalling molecule like hormone and which may develop the cell of origin of almost all breast tumors; surrounding the luminal cells is a layer of smooth muscle type myoepithelial (basal) cells which are involved in the milk ejection reflex. The increase in gap junction communication allows electrical coupling of the smooth muscle cells of the myometrium thus facilitating co-ordinated contraction during parturition (Pandya, 2011).

The staging system of disease characterize as "Tumor-Node-Metastasis" (TNM). Classification of malignant tumors is recommended by the Union for International Cancer Control (UICC), which is an anatomically based system that records the primary and regional nodal extent of the tumor and the absence or presence of metastases. Based on these parameters the determination of staging varies from stages 0 to IV (Banin Hirata, 2014).

The development of breast cancer involves a progression through series of intermediate stages, starting with ductal hyper proliferation, followed by subsequent progression to carcinoma in situ, invasive carcinoma, and finally into metastatic disease (Banin Hirata, 2014).

Tumor marker

"Tumor marker" is a biomarker that is detected in blood, urine, body fluid or body tissues that can be used to identify the presence of specific types of cancer. It is produced either by the tumor itself or by the host in the response to a tumor. The ideal tumor marker should be both specific and sensitive to detect small tumors to allow early diagnosis or help in screening. Few markers are specific for a single tumor. Different markers are produced by different tumors of the same tissue type. Hence understanding of the individual test characteristics and limitations of each marker is important for optimal use and accurate interpretation of results (Kabel, 2017). Tumor markers can be classified based on origin such as molecular, cellular, physiological, image based marker and biomolecule (Kaur and Armengol, 2022).

Protein biomarker has been used for 1st time for cancer diagnosis. Nucleic acid, protein detected in tissue of biopsy sample can be used for analysis of patient prognosis, response to therapy, cancer surveillance, monitoring body responses. Protein biomarkers include antigen, enzyme, hormone, variation in protein glycosylation. Glycans are polymer of monomeric unit of suger conjugated with protein which is also called as glycoprotein. Alteration in glycosylation of specific protein such as surface antigen can be used as tumor marker for eg. cancer antigen CA 15-3, CA 19-9, CA 27-29, CA 125, CA 549, β – hCG, AFM, carcinoembyonic antigen (CEA), CEACAMS, TAG -72, TG, Tn antigen. These markers can be used for characterization of tumor (Kaur and Armengol, 2022).

Cellular markers include change in morphology and function of cellular responses, DNA damage response, level of oxidative stress, cell death (apoptosis). Change in gene expression and protein synthesis can result in change in phenotypic feature. Blood cells can also be used as biomarker which includes white blood cells (WBCs), circulating tumor cells (CTCs), plasma, serum or EVs. EVs are small lipid bound vesicles release by cells play major role in intercellular communication. Depending on origin, they can be micro vesicles, endosome, apoptotic body. Based on which cargo is carrying vesicles are used for identification. Variation in expression of membrane bound vesicles, intra vesicular protein expression has been reported in body fluid collected from patient samples. Tumor markers are involved in prediction of tumor response and estimate effectiveness and toxicity of treatment. These biomarkers are also helpful for developing target specific therapy for treatment of patient (Kaur and Armengol, 2022).

Fibronectin

"Cellular fibronectin" (cFN) is one of the important tumour markers which are extensively studied. Cellular fibronectin (cFN) mediates a wide variety of cellular interactions with the extracellular matrix (ECM) and plays important roles in cell adhesion, migration, growth and differentiation. Cellular fibronectin exists as a dimer composed of two nearly identical ~250 kDa subunits linked covalently near their C-termini by a pair of disulphide bonds. Complete molecule is composed of series of domain which are EDA, EDB and IIICS. Cellular fibronectin is found as soluble and insoluble form. On the basis of its solubility, fibronectin are subdivided into two forms – soluble plasma fibronectin (pFN) and less-soluble cellular fibronectin (cFN). Soluble fibronectin is constituent of plasma and other body fluids where as another part of the insoluble extracellular matrix which is produce by cells in response to tissue injury and inflammation known as cellular fibronectin(cFN). Gene for fibronectin is present on chromosome no. 2. Expression of fibronectin is regulated by signalling molecule such as TGF- β, PDGF, IGF, glucose, glucocorticoid. Plasma fibronectin is synthesized predominantly in the liver by hepatocytes and shows a relatively simple splicing pattern. The alternatively spliced EDA and EDB domains are almost always absent from plasma fibronectin. Cellular fibronectin possess high affinity integrin binding site (RGD-arg-gly-asp), phosphorylation of integrin cytoplasmic domain has role in cell migration, differentiation and apoptosis .Along with it also has heparin and proteoglycan binding site (Pankov, 2002).

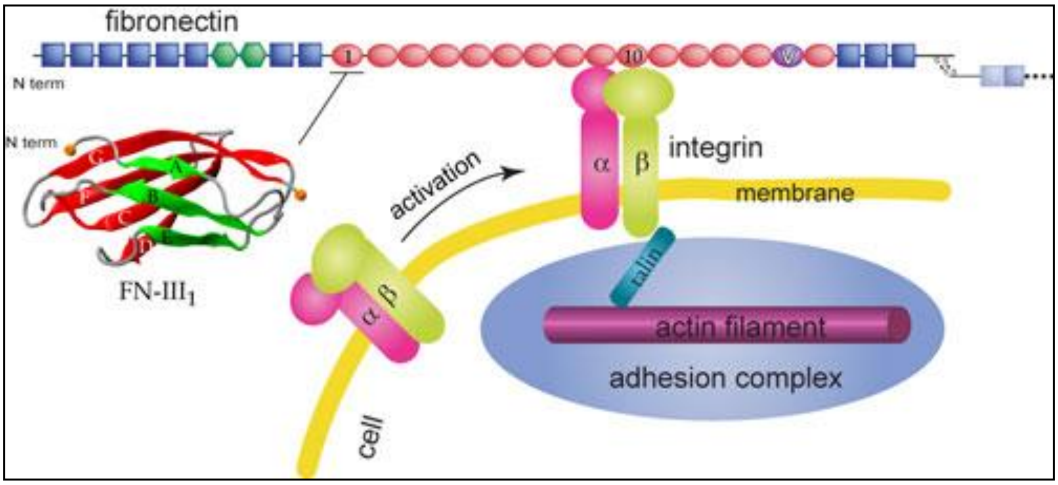


Figure 1: Schematic view of cellular fibronectin bound to integrin at the cell surface (Mu. Geo, 2006)

"Epithelial-mesenchymal transition" (EMT) of epithelial tumor is defined as having a transient mesenchymal appearance that is loss of epithelial characteristics and acquisition of a mesenchymal phenotype. EMT is associated with progression and invasion of cancer cells into the surrounding microenvironment and a chemo resistant phenotype. Cellular fibronectin is up-regulated during the EMT process in epithelial cancer cells, hence cellular fibronectin has been used as a mesenchymal marker for detection of EMT phenotype of cancer (Park, 2014).

Connexins

Intercellular communication plays critical role in transport of ions, metabolite, signalling molecule and secondary messengers. Cellular migration and cell growth is mediated by both channel dependent through gap junction and channel independent pathway. The intercellular communication is regulated by family of molecule called "Connexins". It is the basic unit of gap junctions which assembles and form channel across the cell membrane and mediates cellular signalling and transport. Connexins consist of a family of more than 20 proteins which basically cluster in homomeric or heteromeric groups of six to form a hemi channel also called "Connexion" in the cell membrane. Hemi channels allow for exchange of molecules between the cytosol and the extracellular space. Two hemi channels from adjacent cells can form together a complete gap junction channel that directly connects their respective cytosols. Gap junction channels are clustered in larger groups in the membrane forming so called "plaques". All connexin are tetra spanning membrane proteins, the N-terminus and the C-terminus of the proteins being located in the cytosol (Kameritsch, 2012).

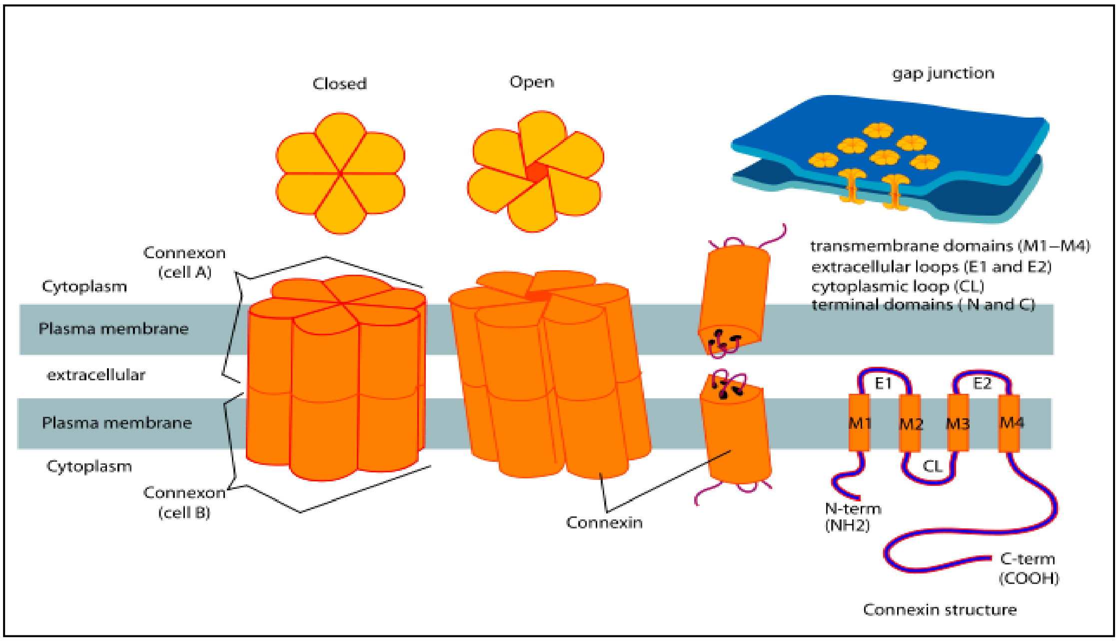


Figure 2: Connexion structure overview (Walber, 2005)

Intercellular cellular communication in tumor cells

Tumor formation in the breast involves changes in mammary epithelial cell interactions with the ECM. Loss of basement membrane integrity is a hallmark of cancer and occurs through changes in expression of ECM proteins or their receptors and through increased release of matrix metalloproteases and change in cell matrix composition (Ma, 2009).

Cancer cell-derived ECM proteins mediate the cancer cell-stromal cell crosstalk, which indicate that cancer cell-derived ECM molecules are critical regulators of the initiation of metastasis outgrowth through activating the stromal cells in the secondary organs. Hence inhibition of such signalling molecule which capable of altering expression in cancer cell which lead to repression of cell adhesion, migration, and invasion of cancer cell. In addition, cancer cell-derived ECM proteins (cellular fibronectin, collagen, and laminin) protect cancer cells from chemotherapy-induced apoptosis via activation of the PI3K/AKT pathway. Cell derived molecule through various signalling pathway involve in creating microenvironment for cancer progression. As tumour cells play critical role in ECM deposition and remodelling during cancer development and progression, it was showed that ECM molecule deposited by cancer cells promote cancer progression by enhancing cell survival and proliferation. However, it largely remains to be undetermined how cancer cell-derived ECM molecule is regulated and how those ECM proteins function in tumor microenvironment remodelling. Answering those questions is critical for developing potential cancer treatment (Xiong, 2016). As it is mentioned that change in cellular communication and tissue repair mechanism play important role in transformation, connexin and cellular fibronectin are two important molecule analysed for its expression in breast cancer cells.

Activation of p13k/AKT signalling pathway carried out by phosphorylation of tyrosine kinase receptor. This pathway induce by integrin binding through cellular fibronectin, it play role in cell cycle regulation. Signalling pathway is mostly altering during the tumorigenesis (Park, 2014).

The fibronectin matrix can also sequester growth factors and associated proteins, including bone morphogenetic protein-1, vascular endothelial growth factor (VEGF) and latent transforming growth factor (TGF)-β binding proteins (LTBP) to regulate cell signalling events (Bae, 2013). Cellular fibronectin stimulates proliferation of growth-arrested polarized mammary epithelial cells, disturbing the hollow acinar structure and promoting tumor-like behaviour (Park, 2014).

In the mammary gland, the stromal extracellular matrix (ECM) undergoes dramatic changes during development of cancer and in tumorigenesis, such as normal adult breast tissue is largely devoid of the ECM protein such as cellular fibronectin (cFN) whereas high cellular fibronectin levels have been detected in stroma of tumor cell (Park, 2014). In fact, cellular fibronectin levels in breast tumor tissues are positively correlated with tumor malignancy and negatively correlated with the survival rate of breast cancer patients, which suggest cellular fibronectin might play a role in cancer progression and severity. Cellular fibronectin transmits signal in ECM by binding to integrin receptors, which are heterodimer trans membrane proteins that link the ECM with the cytoskeleton and intracellular signalling pathways, which has role in cell diffentiation. (Park, 2014). Cellular fibronectin induces an EMT response which include up-regulation of the EMT markers FN, Snail, N-cadherin, vimentin, matrix metalloprotease (MMP2), α-smooth muscle Actin, and phospho-Smad2 as well as acquisition of cell migratory behaviour. Cellular fibronectin initiates EMT in normal cells which is enhances the effect of endogenous TGF- β. EMT marker expression is up-regulated in cells in presence fragment of fibronectin containing the integrin-binding domain, as integrin is key component in cell migration and adhesion. This indicate that during EMT, epithelial cell adhesion switches from cell-cell contacts to mainly cell-ECM interactions raising the possibility that cellular fibronectin may have a role in promoting this transition (Park, 2014). When gene expression of stromal and epithelial fibronectin was compared it was found that down-regulation of cytoplasmic ribosomal protein and up-regulation of mitochondrial ribosomal protein and change in ribosomal structure which results in tumour growth. Gene involves in cell signalling was dis-regulated and increase mitotic activity (Ma, 2009). To check fibronectin level primary tumour cells cytoplasmic staining showed higher fibronectin in metastasized tumour into lymph node than primary tumors which suggest that fibronectin could be synthesized by neoplastic cells which play role in metastasis. This could relate the positive correlation of stromal fibronectin with epithelial fibronectin (Bae, 2013).

Gap junction

"Gap junction intercellular communication" (GJIC) represents a central conduit of ions, essential metabolites, and second messengers, such as Ca2+, cAMP, cGMP, and IP3, between adjacent cells. Gap junction is unit assemble from hemi channel which involve in movement of NAD+ into and out of cells also regulate Ca2+ concentrations through the CD38 trans membrane glycoprotein. The regulation of hemi channels is through extracellular changes in ionic concentration, membrane depolarization, metabolic inhibition, and mechanical stimuli. Hemi channel's function can affect various cellular events, including calcium signalling, cell proliferation and death, as well as the normal functioning of cell. Initially it was suggested that intercellular transport was carried out through gap junction passively. However, studies have shown that gap junctions undergo complex regulation, and play an active role in intercellular signalling and communication. Gap junction channels, depending on cellular needs and conditions, alternate between "closed" and "open" conformations. These changes are regulated by various mechanisms including calcium concentration, pH, trans junctional potential, and protein phosphorylation (Dbouk, 2009). Protein phosphorylation results in change in gap junction at cell-cell interphase due to trafficking and degradation of membrane protein and by changing the efficiency of passage of molecule across the cell membrane. Gap junctions are also involved in cell death. This is due to the "bystander effect" in which gap junctions spread a death signal between dying cells and those adjoining them and this may be mediated by Ca2+ influx between the cells (Dbouk, 2009).

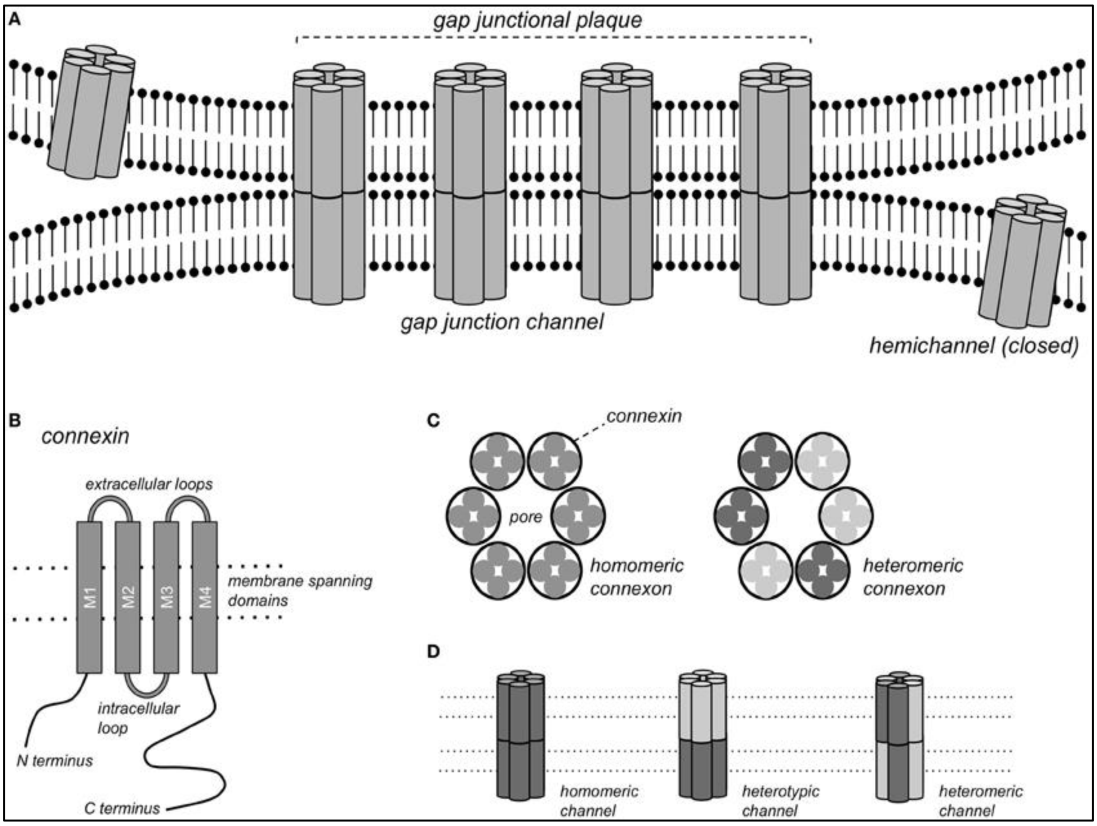


Figure 3: Gap junction channel. (A) Schematic representation of gap junction plaque, with each of the neighbouring cells contributing one hemi channel to form complete gap junction channel. (B) Structure of pattern of single connexin subunit with 4 domains traversing the membrane. (C) A hemichannel or connexin consists of a hexagonal arrangement of 6 cx subunit, either of the same cx isoform (homomeric connexin) or different isoforms (hetreromeric connexin) (D) gap junction channel formed by homomeric connexin of the same isoform (homomeric channel), by homomeric connexions of different isoform (heteromeric channel) (Sander verhule, 2013).

Connexin have been shown to play essential roles in normal development and differentiation of a variety of tissues. Connexin and their associated proteins are implicated in mammary gland differentiation, with their levels and phosphorylation status being modified throughout the differentiation process, as well as through modulation of their associated proteins. Connexin 43 is a predominant connexin protein that has been studied in a variety of human cancers including prostate, lung, liver, brain, and breast (Dbouk, 2009). Connexin 43 interacts with a range of cytoskeleton proteins, including zonula occludens-1 (ZO-1), ZO-2, and α- and β-tubulin, to regulate cell adhesion and migration (Dbouk, 2009). Increasing evidence indicates that Cx43 directly affects the proliferation and migration of keratinocytes and fibroblasts and inflammatory response through certain chemokine, cytokine and growth factor (Zhang, 2017), hence it is important to study about this molecule with respect to breast cancer.

Connexin expression can be regulated at various steps in the pathway from DNA to RNA to protein, i.e., transcriptional control, RNA processing control, RNA transport and localization control, translational control, mRNA degradation control, and protein activity control like DNA methylation, histone modifications, and microRNAs. (Oyamada, 2013). Expression of connexin 43 is based on alternative splicing between two exon and cell type specific. Up-regulation of connexin 43 is mediated through Ras mediated signalling pathway and putative consensus sequence in promoter region. Connexin 26 is present in between spanning two exons within GC box and TTAAAA box (Oyamada, 2013).

Human breast has shown presence of connexin 26 and connexin 43 in individual compartment of the epithelial cell. Connexin 43 is predominantly expressed in the basal myoepithelium of the duct whereas connexin 26 was observed in luminal epithelial cells of normal human breast cell. Connexin 43 has both pro and antimetastatic role which indicate that during the disease progression connexin 43 expression is either decreased or increased based on stage. Connexin 26 and connexin 43 expressed by carcinoma cell could alter GJIC between malignant and normal breast cell. Phosphorylation of connexin regulate assembly of connexin 43 and gap junction formation which mediate increase or decrease GJIC, this phosphorylation correlate with breast cancer stage (Phillips, *et al.,* 2017). Breast cancer cells have reduced connexin expression also impaired ability of transport, localize, assembly of connexin into a gap junction, which suggest that exogenous expression of connexin or up regulation of connexin may be a useful approach to restoration of gap junction and GJIC which possibly inhibit the tumour cell growth and metastatic potential. Connexin 43 expression was found to be up-regulated in invasive carcinoma in stromal compartment and down-regulated in cytoplasmic compartment (Phillips, *et al.,* 2017). Connexin 43 gene expression is under the control of steroid hormone, mostly it up regulated by estrogen and down regulated by progesterone through signal transduction (Oyamada, 2013). Epigenetic inactivation through hyper methylation of promoter region leads to silencing of connexin expression. Connexin expression and gap junction formation is enhanced by cyclic AMP (Oyamada, 2013). Connexin 26 down regulation in tumour cells which implicate connexin 26 may be a candidate tumour suppressor gene in the human breast (Dbouk, H. A., 2009). It was observed that the loss of connexin 43 in intercellular communication is an early event in malignancy, with the possibility of gap junction restoration in the event of metastasis. (Dbouk, 2009).

These molecular changes may contribute to tumour growth and can be used as cellular marker for analysis of disease progression. Further studies can be carried to check the whether molecular changes in the cellular compartment with respect to cell migration and communication in tumour which may contribute to change in stromal region and activation of other signalling molecule which lead to tumour growth and metastasis.

References:

1. Bae, Y. K., Kim, A., Kim, M. K., Choi, J. E., Kang, S. H., & Lee, S. J. (2013). Fibronectin expression in carcinoma cells correlates with tumor aggressiveness and poor clinical outcome in patients with invasive breast cancer. Human pathology, 44(10), 2028-2037.
2. Banin Hirata, B. K., Oda, J. M. M., Losi Guembarovski, R., Ariza, C. B., Oliveira, C. E. C. D., & Watanabe, M. A. E. (2014). Molecular markers for breast cancer: prediction on tumor behaviour. Disease markers, 2014.
3. Dbouk, H. A., Mroue, R. M., El-Sabban, M. E., & Talhouk, R. S. (2009). Connexins: a myriad of functions extending beyond assembly of gap junction channels. Cell Communication and signaling, 7(1), 4.
4. Kabel, A. M. (2017). Tumor markers of breast cancer: New prospectives. Journal of Oncological Sciences, 3(1), 5-11.
5. Kameritsch, P., Pogoda, K., & Pohl, U. (2012). Channel-independent influence of connexin 43 on cell migration. Biochimica et Biophysica Acta (BBA)-Biomembranes, 1818(8), 1993-2001.
6. Kaur, V.S, Armengol, G, (2022) Molecular biomarkers in cancer, biomolecules (12) 8: 1021, MDPI, PMC9331210.
7. Ma, X. J., Dahiya, S., Richardson, E., Erlander, M., & Sgroi, D. C. (2009). Gene expression profiling of the tumor microenvironment during breast cancer progression. Breast cancer research, 11(1), R7.
8. Pandya, S., & Moore, R. G. (2011). Breast development and anatomy. Clinical obstetrics and gynecology, 54(1), 91-95.
9. Pankov, R., & Yamada, K. M. (2002). Fibronectin at a glance. Journal of cell science, 115(20), 3861-3863.
10. Phillips, S. L., Williams, C. B., Zambrano, J. N., Williams, C. J., & Yeh, E. S. (2017). Connexin 43 in the development and progression of breast cancer: What's the connection? International journal of oncology, 51(4), 1005-1013.
11. Rakoff-Nahoum, S. (2006). Cancer issue: why cancer and inflammation? The Yale journal of biology and medicine, 79(3-4), 123.
12. Xiong, G. F., & Xu, R. (2016). Function of cancer cell-derived extracellular matrix in tumor progression. Journal of Cancer Metastasis and Treatment Volume, 2, 358.
13. Zhang, X. F., & Cui, X. (2017). Connexin 43: Key roles in the skin. Biomedical reports, 6(6), 605-611.

CUTTING-EDGE PROGRESS IN TANDEM SOLAR CELL INNOVATION

Kailash Nemade

Department of Physics,

Indira Mahavidyalaya, Kalamb Dist. Yavatmal 445401, Maharashtra, India

Corresponding author E-mail: krnemade@gmail.com

Abstract:

The demand for efficient and sustainable energy sources has spurred remarkable advancements in photovoltaic technology, particularly in the field of tandem solar cells. This research paper delves into the cutting-edge progress made in tandem solar cell innovation, exploring the latest developments, breakthroughs, and challenges in this rapidly evolving domain. Tandem solar cells, also known as multi-junction solar cells, offer the potential to surpass the efficiency limitations of traditional single-junction cells by stacking multiple sub-cells, each optimized for a specific portion of the solar spectrum. This chapter comprehensively reviews the state-of-the-art materials, design strategies, and fabrication techniques employed in tandem solar cell architectures. By critically evaluating the progress and prospects of tandem solar cell innovation, this research paper offers valuable insights into the future direction of photovoltaic technology. The synthesis of recent developments and emerging trends serves as a guide for researchers, engineers, and policymakers seeking to accelerate the deployment of high-efficiency solar energy solutions and drive the transition towards a sustainable energy landscape.

Keywords: Energy Conservation; Tandem Cell; Solar Cell

Introduction:

Energy conservation is a crucial and pressing need in today's world due to a myriad of reasons that encompass environmental, economic, and social dimensions. As global energy demand continues to rise, it is imperative that we adopt strategies to conserve energy and utilize it more efficiently. First and foremost, environmental concerns are at the forefront of the energy conservation movement. The excessive consumption of non-renewable energy sources, such as fossil fuels, leads to detrimental effects on the planet. Burning fossil fuels releases greenhouse gases into the atmosphere, contributing to climate change, air pollution, and the depletion of natural resources. By conserving energy, we can reduce our carbon footprint and mitigate the impact of climate change, thus preserving a habitable planet for future generations.

Economic factors also play a significant role in the necessity of energy conservation. Energy production and distribution require substantial financial investments, and high demand can strain energy infrastructure. By conserving energy, we can alleviate the pressure on energy grids, reduce the need for expensive infrastructure upgrades, and ultimately lower energy costs for consumers. Moreover, energy efficiency measures in industries and households can lead to substantial cost savings over time, enhancing economic stability and resilience [1].

From a social perspective, energy conservation contributes to a more sustainable and equitable society. Access to affordable energy is essential for human well-being, but it is not equally accessible to all. By conserving energy, we ensure that resources are used judiciously, extending access to energy services to marginalized communities and promoting social inclusivity. Furthermore, energy conservation fosters innovation and technological advancement. The drive to reduce energy consumption encourages research and development in renewable energy sources, energy-efficient technologies, and smart energy management systems. This, in turn, creates job opportunities and stimulates economic growth in the clean energy sector.

Advancements in photovoltaic technology across generations

Photovoltaic (PV) technology has evolved significantly over the years, progressing through different generations of solar cell technologies. These advancements have been driven by the pursuit of greater efficiency, cost-effectiveness, and environmental sustainability. The key developments in each generation of PV technology are as follows [2, 3],

1. First Generation (1G) - Crystalline Silicon (c-Si) technology

The first generation of PV technology mainly comprises crystalline silicon solar cells. These cells are made from slices of silicon crystals, typically in the form of monocrystalline or polycrystalline cells. While 1G technology laid the foundation for solar energy adoption, it has certain limitations, such as high production costs and lower efficiency compared to subsequent generations.

2. Second Generation (2G) - thin-film solar cells

2G PV technology introduced thin-film solar cells, which are made from semiconductor materials deposited in thin layers onto various substrates. This generation includes amorphous silicon (a-Si), cadmium telluride (CdTe), and copper indium gallium selenide (CIGS) solar cells. Thin-film technology offers advantages like lower manufacturing costs, flexibility, and the ability to be integrated into building materials. However, efficiency levels have generally been lower compared to crystalline silicon cells.

3. Third Generation (3G) - advanced PV concepts

The third generation encompasses various advanced PV concepts, such as multi-junction solar cells, concentrating photovoltaics (CPV), and organic solar cells. Multi-junction cells use multiple semiconductor layers to capture a broader range of wavelengths, boosting efficiency. CPV systems focus sunlight onto small, high-efficiency cells using lenses or mirrors. Organic solar cells aim to utilize organic materials to create flexible and lightweight solar panels, enabling novel applications. While these technologies hold promise, they often face challenges related to scalability and stability.

4. Fourth Generation (4G) - Tandem and perovskite solar cells

4G PV technology focuses on tandem solar cells and perovskite solar cells. Tandem cells combine different solar cell materials to capture a wider spectrum of light, thus enhancing efficiency. Perovskite solar cells have gained attention for their rapid efficiency improvements and potential for low-cost manufacturing. However, challenges related to stability and long-term performance still need to be addressed before widespread commercialization.

5. Fifth Generation (5G) - emerging concepts

The fifth generation represents emerging and innovative PV concepts that are still in the research and development phase. These include technologies like quantum dot solar cells, transparent solar panels, and solar windows. Quantum dots offer the potential to tune the absorption and emission properties of materials at the nanoscale, enabling high-efficiency and versatile solar cells. Transparent solar panels and solar windows aim to integrate solar harvesting into building materials and windows, enabling seamless energy generation. Figure 1 depicts various generations of photovoltaic technology.

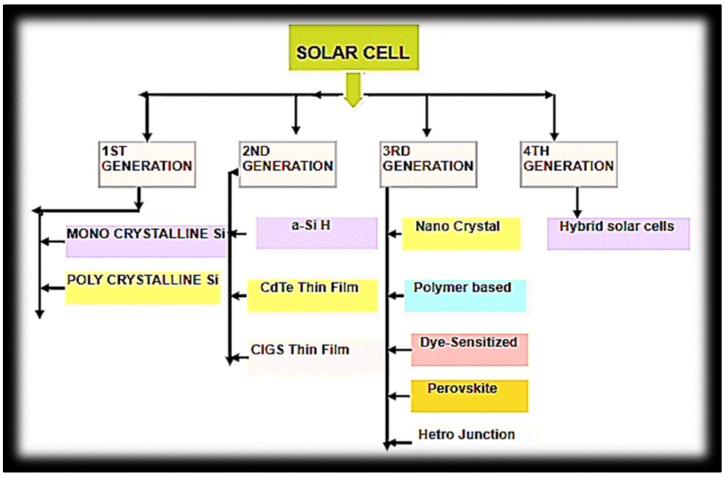


Figure 1: Various generations of Photovoltaic Technology

Tandem solar cells

Tandem solar cells, also known as multi-junction solar cells, are a type of photovoltaic device that combines multiple individual solar cell layers, each designed to capture a specific range of light wavelengths. By stacking these layers on top of each other, tandem solar cells can achieve higher efficiency compared to traditional single-junction solar cells.

The fundamental idea behind tandem solar cells is to optimize the absorption of sunlight by utilizing different semiconductor materials, each with a bandgap optimized for a specific portion of the solar spectrum. This allows tandem cells to convert a broader range of wavelengths into electricity, thereby increasing overall energy conversion efficiency [4].

The structure of a typical tandem solar cell involves multiple sub-cells stacked vertically, with each sub-cell designed to capture a specific portion of the solar spectrum. The sub-cells are connected in series to ensure that the current generated by each sub-cell adds up, contributing to the overall output current. Here's a basic overview of the structure:

1. Top sub-cell

This sub-cell is designed to capture high-energy photons (short wavelengths) from the solar spectrum. It has a bandgap optimized for blue and ultraviolet light. Materials commonly used for the top sub-cell include gallium arsenide (GaAs) or indium gallium nitride (InGaN).

2. Middle sub-cell(s)

In a tandem solar cell, there can be one or more middle sub-cells, each optimized for a specific wavelength range. These sub-cells are often made from different semiconductor materials, such as gallium indium phosphide (GaInP) or gallium arsenide phosphide (GaAsP), depending on the desired absorption characteristics.

3. Bottom sub-cell

The bottom sub-cell is designed to capture low-energy photons (long wavelengths) from the solar spectrum. It typically uses a material with a wider bandgap, such as silicon (Si) or copper indium gallium selenide (CIGS).

To ensure efficient energy conversion, each sub-cell is carefully designed to minimize losses and optimize the extraction of charge carriers (electrons and holes) generated by absorbed photons. Transparent conductive layers and antireflection coatings may also be incorporated to enhance light absorption and reduce surface reflections.

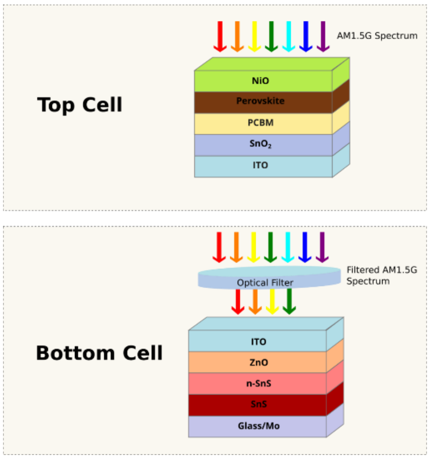


Figure 2: Pictorical architechture of tandem solar cells

One of the key challenges in tandem solar cell design is achieving proper current matching among the sub-cells. This involves balancing the current generated by each sub-cell to ensure that the overall current output is maximized. Advanced engineering techniques, such as tunnel junctions or interconnecting layers, are often employed to achieve optimal current matching and efficient charge transport between the sub-cells.

Tandem solar cells hold the potential to significantly improve the efficiency of solar energy conversion and play a crucial role in the advancement of renewable energy technologies. Ongoing research and development in tandem cell materials, design, and fabrication techniques aim to further enhance their performance and commercial viability. Figure 2 depicts pictorical architechture of tandem solar cells [5].

Future scope and challenges in tandem solar cells technology

Tandem solar cells have emerged as a promising avenue for achieving higher solar energy conversion efficiencies and overcoming the limitations of traditional single-junction solar cells. While significant progress has been made in recent years, there remains a wide scope for further advancements in tandem solar cell technology. This section discusses the prospects and challenges that lie ahead in harnessing the full potential of tandem solar cells.

Future scope:

Continued research into novel materials, including perovskites, quantum dots, and other emerging semiconductors, holds the promise of pushing tandem solar cell efficiencies beyond current records. The optimization of bandgap engineering and innovative sub-cell configurations will contribute to capturing a broader spectrum of sunlight and converting it into electricity with increased efficiency. Integrating tandem solar cells into various applications, such as wearable devices, building-integrated photovoltaics, and portable electronics, presents a significant opportunity. The development of flexible and transparent tandem cells could enable their incorporation into a wide range of surfaces and structures, expanding their deployment and energy generation potential.

Tandem solar cell technology has the potential to revolutionize space-based solar power generation. The unique characteristics of tandem cells, including high efficiency and adaptability to different wavelengths, make them an ideal candidate for powering satellites and spacecraft, offering a sustainable solution for energy-intensive space missions. Streamlining and cost-effective manufacturing techniques will play a crucial role in the scalability and commercial viability of tandem solar cells. Advances in deposition methods, roll-to-roll processing, and printing technologies can lead to efficient large-scale production, reducing manufacturing costs and increasing accessibility. Many advanced tandem solar cell materials, such as perovskites, face challenges related to stability and long-term performance. Researchers must focus on enhancing the stability of these materials to ensure the practical viability of tandem cells in real-world conditions, including exposure to environmental factors and temperature fluctuations. Achieving optimal current matching between sub-cells remains a challenge, especially as the number of sub-cells in tandem structures increases. The design complexity of tandem cells necessitates sophisticated engineering solutions to minimize energy losses and maximize overall efficiency. Transparent conductive materials that maintain high conductivity and transparency across the entire solar spectrum are essential for tandem solar cell performance. Developing new materials or improving existing ones to meet these requirements is an ongoing challenge. While tandem solar cell efficiencies have shown impressive gains in research settings, transitioning these technologies to commercial products at a competitive cost remains a hurdle. The development of cost-effective manufacturing processes and the establishment of a supply chain for tandem cell components are crucial steps towards widespread adoption [6].

Conclusions:

Through a meticulous analysis of recent studies, experimental data, and theoretical models, we present an in-depth overview of the advancements achieved in tandem solar cell efficiency, stability, and scalability. The utilization of novel semiconductor materials, such as perovskites and quantum dots, has opened new avenues for enhancing light absorption and charge carrier generation across the spectrum. Furthermore, the incorporation of innovative interfacial engineering and device architectures has contributed to improved charge transport and reduced recombination losses. The integration of tandem solar cells into various applications, such as building-integrated photovoltaics and portable electronics, underscores their potential to revolutionize renewable energy utilization. In conclusion, the future of tandem solar cell technology holds great promise for significantly advancing solar energy conversion efficiency and expanding its applications. Overcoming challenges related to stability, manufacturing, and integration will be essential for unlocking the full potential of tandem solar cells and accelerating their integration into our global energy landscape. Continued research, innovation, and collaboration across scientific, engineering, and industrial sectors will drive the realization of these transformative technologies.

Acknowledgements:

Author is very much thankful to Principal, Indira College, Kalamb Dist. Yavatmal for providing all necessary academic assistance during this work.

References:

1. Hui Li and Wei Zhang (2020). Perovskite Tandem Solar Cells: From Fundamentals to Commercial Deployment, Chem. Rev. 2020, 120, 18, 9835–9950.
2. Yuanhang Cheng, Liming Ding (2021). Perovskite/Si tandem solar cells: Fundamentals, advances, challenges, and novel applications, Sus. Mat., 2021, 1, 3, 324-344.
3. T. Leijtens, K.A. Bush, R. Prasanna, M.D. McGehee, (2018). Opportunities and challenges for tandem solar cells using metal halide perovskite semiconductors, Nature Energy, 2018, 3, 2, 828–838.
4. A.F.J. Boom, S. Ferro, M. Gelvez-Rueda, H. Zuilhof, B. Ehrler (2023). Toward improving triplet energy transfer from tetracene to silicon using a covalently bound tetracene seed layer, The Journal of Physical Chemistry Letters, 2023, 14, 19, 4454-4461.
5. H. Wang, J. He, H. Xiang, R. Ran, W. Zhou, W. Wang, Z. Shao (2023). Additive engineering for mixed lead–tin narrow-band-gap perovskite solar cells: recent advances and perspectives, Energy & Fuels, 2023, 37, 9, 6401-6423.
6. R. Jiang, Y. Luan, J. Li, F. Ye, S. Zhang, Z. Su, C. Jin, G. Tong, J. Tong, F. Huang, Y. Cheng, T. Bu (2023). Low-cost fabrication of efficient perovskite photovoltaics using low-purity lead iodide via powder engineering, ACS Applied Materials & Interfaces, 2023, 15, 17, 20958-20965.

PREDICTIVE MODELS FOR ESTIMATING THE SURVIVAL LENGTH OF GASTRIC CANCER PATIENTS USING THE MAMDANI FUZZY CONTROLLER: A COMPREHENSIVE REVIEW

Tabendra Nath Das

Department of Mathematics,

Dhakuakhana College, Lakhimpur, Assam, India

Corresponding author E-mail: tabendra2@gmail.com

Abstract:

Gastric cancer is a significant global health concern, ranking as one of the leading causes of cancer-related mortality. Prognostic prediction for gastric cancer patients is essential for treatment planning, patient counseling, and resource allocation. Traditional statistical models have been widely employed, but they often struggle with the uncertainty and imprecision inherent in clinical data. Fuzzy logic systems, particularly those based on the Mamdani fuzzy inference system, offer a promising alternative due to their ability to handle vague and linguistic information. This comprehensive review discusses the application of Mamdani fuzzy controllers in developing predictive models for estimating the survival length of gastric cancer patients. We analyze the theoretical foundations, design methodology, input and output variables, rule formulation, defuzzification strategies, validation metrics, comparative performance, advantages, limitations, and future prospects.

Keywords: Mamdani Fuzzy Controller, Gastric Cancer, Survival Length

1. Introduction:

Gastric cancer, or stomach cancer, refers to the uncontrolled growth of malignant cells in the lining of the stomach. It remains a leading cause of cancer deaths worldwide, particularly in East Asia and parts of Eastern Europe. Accurate prediction of survival length plays a pivotal role in clinical decision-making, allowing healthcare professionals to design personalized treatment protocols and to provide more accurate prognoses.

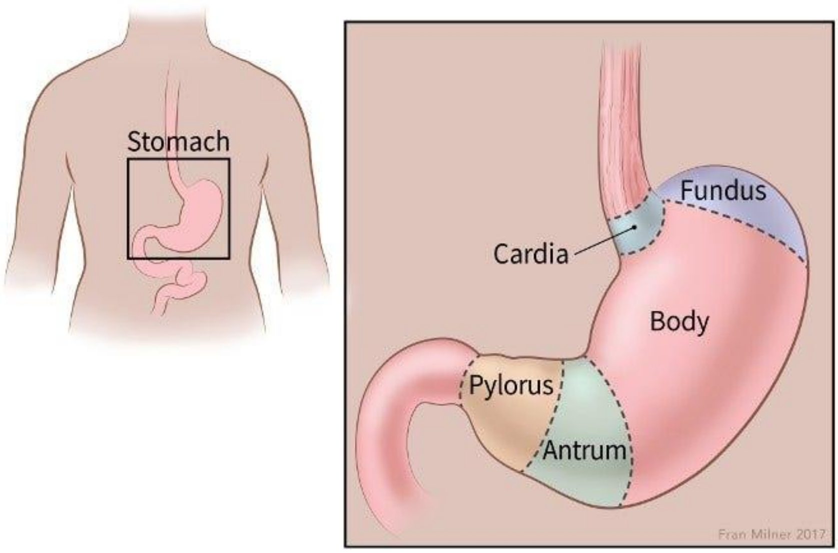
Survival prediction is inherently complex due to the multitude of factors that influence patient outcomes, including age, gender, tumor stage, histological type, lymph node involvement, metastasis, and the nature of therapeutic interventions. These variables are often imprecise, and their interactions are nonlinear, making it difficult for conventional statistical approaches to capture these complexities effectively.

Fuzzy logic, first introduced by Lotfi Zadeh in 1965, provides a means of modeling uncertainty and imprecision using linguistic variables. The Mamdani fuzzy inference system (FIS), developed by Ebrahim Mamdani in 1975, has become one of the most widely used fuzzy systems in various domains, including healthcare. Its ability to simulate human reasoning makes it particularly suitable for medical decision-making applications.

2. Background and rationale

Gastric cancer epidemiology and prognosis

Gastric cancer ranks fifth in incidence and fourth in cancer-related deaths globally. Its prognosis depends on the stage at diagnosis, with early-stage cancers having a much better survival rate than advanced-stage cases. Despite advancements in screening and therapy, the five-year survival rate remains below 30% in most countries, underscoring the need for better predictive tools.



Parts of the stomach

Limitations of traditional prognostic models

Traditional models, such as the Cox proportional hazards model, Kaplan-Meier survival analysis, and logistic regression, are based on the assumption of linear relationships and work best when the data is complete and neatly organized. These assumptions often do not hold in real-world clinical settings where data are noisy, incomplete, or expressed in qualitative terms.

Advantages of fuzzy logic in medical prognosis

Fuzzy logic accommodates uncertainty and imprecision by employing linguistic variables (e.g., "young age," "advanced tumor") and membership functions that map inputs to a range between 0 and 1. The Mamdani FIS uses a rule-based approach that resembles human decision-making, making it easier to interpret and justify clinical decisions.

3. Mamdani fuzzy inference system: theoretical framework

Components of mamdani FIS

Fuzzification: It is the process of turning exact input values into fuzzy sets using membership functions.

Rule base: Consists of if-then rules derived from clinical expertise or data.

Inference engine: Evaluates the rules and combines results to form fuzzy outputs.

Defuzzification: Converts fuzzy outputs back into crisp values.

Membership functions

Frequently used membership functions include shapes such as triangular, trapezoidal, and Gaussian. The choice of membership function impacts the model's sensitivity and interpretability.

Rule formulation

Rules in a Mamdani FIS are typically of the form:

IF (Tumor Stage is Advanced) AND (Age is Elderly) THEN (Survival Time is Short)

Rules can be derived from clinical guidelines, expert consultations, or data mining techniques.

Defuzzification methods

Popular methods include:

- Centroid Method
- Bisector Method
- Mean of Maximum (MoM)
- Smallest/Largest of Maximum

4. Designing a mamdani FIS for gastric cancer survival prediction

Input variables

- Typical inputs include:
- Age
- Gender
- Tumor Size
- Histological Type
- TNM Staging (Tumor, Node, Metastasis)
- Lymph Node Ratio
- Metastasis Presence
- Treatment Modality

Output variables

The output variable typically represents the estimated survival duration, which is classified as follows:

- Short (< 12 months)
- Medium (12–36 months)
- Long (> 36 months)

Knowledge acquisition and rule development

Rule development is a critical phase and can be performed through:

- Expert Panels
- Clinical Guidelines
- Machine Learning (e.g., decision trees, genetic algorithms)

5. Case studies and literature review

Several studies have implemented Mamdani FIS models for gastric cancer prognosis:

- **Study A**: Utilized 6 input variables and a rule base of 50 rules, achieving 85% prediction accuracy.
- **Study B**: Integrated Mamdani FIS with a genetic algorithm for rule optimization.
- **Study C**: Compared Mamdani and Sugeno models, concluding that Mamdani offered better interpretability.

6. Evaluation metrics and validation techniques

- Model performance is assessed using:
- Mean Absolute Error (MAE)
- Root Mean Square Error (RMSE)
- Accuracy
- Receiver Operating Characteristic (ROC) Curves
- Confusion Matrices
- Cross-validation and external dataset validation enhance model generalizability.

7. Comparative analysis with other AI models

- Mamdani FIS is often compared with:
- Neural Networks
- Support Vector Machines (SVM)
- Decision Trees
- Hybrid Models (e.g., Neuro-Fuzzy Systems)

While other models may outperform in raw accuracy, Mamdani FIS excels in transparency and interpretability.

8. Advantages and clinical relevance

- Handles imprecision and uncertainty
- Easy to understand and explain to clinicians
- Facilitates shared decision-making
- Can be integrated into clinical decision support systems (CDSS)

9. Limitations and challenges

- Rule explosion with multiple input variables

- Subjectivity in membership function selection
- Lack of standardized frameworks
- Computational complexity for large rule bases

Future directions

- Development of hybrid models (e.g., Fuzzy-Neural Networks)
- Use of deep learning for automated rule extraction
- Incorporation with big data platforms and electronic health record (EHR) systems
- Deployment in real-time clinical environments

Conclusion:

Mamdani fuzzy inference systems provide a robust framework for modeling the complex, uncertain, and nonlinear nature of survival prediction in gastric cancer patients. Their human-like reasoning and transparency make them suitable for clinical adoption. While challenges remain, ongoing research and integration with advanced computational methods hold the promise of improving prognostic tools in oncology.

References:

1. World Health Organization. (2020). *Global Cancer Observatory*. https://gco.iarc.fr/
2. Zadeh, L. A. (1965). "Fuzzy sets". *Information and Control*, 8(3), 338–353.
3. Mamdani, E. H. (1974). "Application of fuzzy algorithms for control of simple dynamic plant". *Proceedings of the Institution of Electrical Engineers*, 121(12), 1585–1588.
4. Ghosh, A. et al. (2019). "A hybrid Mamdani-GA system for prognosis of gastric cancer survival". *Expert Systems with Applications*, 123, 118–126.
5. Li, Y. & Wang, S. (2020). "Comparative study of Mamdani and Sugeno fuzzy models for cancer survival prediction". *Applied Soft Computing*, 92, 106290.
6. Jang, J.-S.R. (1993). "ANFIS: Adaptive-Network-Based Fuzzy Inference System". *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3), 665–685.
7. Chatterjee, S. et al. (2021). "Fuzzy logic and genetic algorithm-based prognosis model for gastric cancer patients". *Medical & Biological Engineering & Computing*, 59(4), 749–760.

Research Trends in Science and Technology Volume III (ISBN: 978-93-88901-74-1)

About Editors

Dr. Alok Ranjan Sahu serves as an Assistant Professor in Botany at Vikash Degree College, Bargarh, Odisha. He received M.Sc., M.Phil. (Life Science) and Ph.D. in Biotechnology from the School of Life Sciences, Sambalpur University, Odisha, India. He has published more than 45 research papers in National and International Journals and Books. He also has obtained three National Patents and filed one Patents to his credit. He is the sole author of two books and co-author of four books. He is a Fellow and Life member of Six reputed International and National Academic Societies. He is a Reviewer and Editorial Board Member of more than 30 numbers of National and International Journals, including Biology, Genes, Biomolecules, Sustainability, BABT, JAPS, JABB, etc. For his outstanding research contribution, he was awarded Research Excellence Award-2023, Academic Excellence Award (Botany/ Plant Biotechnology)-2022, Young Achiever Award-2022, Outstanding Scientist Award-2022, Best Faculty Award-2021, Young Plant Breeder Award-2021, Best Researcher Award-2020, Young Scientist Award-2014, Best Poster Presentation Award-2013. His current areas of research interest are Plant Molecular Biology & Genome mapping, Medicinal Plant Conservation, Natural products & therapeutics.

Mr. Basant Deshwal born on 24 June 1996 in Hirnoda village at Jaipur district of Rajasthan, India. He completed his bachelor's degree in agriculture from Agriculture University, Jodhpur in 2019.Following year, he cleared ICAR-JRF and joined Masters at Division of Nematology, IARI, New Delhi. After M.Sc., by 2022 he joined the same division for pursuing Ph.D. He has many research papers, book chapter and popular articles in reputed journals. He has also qualified the ICAR-NET.

Dr. Narayan Gaonkar is presently serving as Assistant Professor of Physics at University College of Science, a constituent college of Tumkur University, Tumkur, Karnataka. He obtained his Master's degree in Physics from Karnatak University, Dharwad in 2009 with gold medals and cleared CSIR-NET in 2010. He is awarded with Ph.D. degree from Tumkur University, Tumkur in 2023. His theoretical research work has resulted in 14 journal publications and a book chapter. He has attended more than 20 international/national conferences to present the research findings. He has jointly authored 03 books on undergraduate physics.

Dr. Sanjay Singh is presently working as assistant professor and head, department of Physics at Chintamani College of Arts and Science, Gondpipri, Dist. Chandrapur. He has more than Ten-year teaching experience. He has done experience at University College (KUK-Haryana) and Shyam Lal College, University of Delhi (DU). During the last twelve year of research experience, he has published 19 research papers in national and international reputed journals. He has presented many research papers in national and international seminar and conference.

Bhumi Publishing

ISBN 978-93-88901-74-1

9 789388 901741