IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Silicon Meets Cells: DNA As The Next Frontier Of Computing And Memory

Author: Aahana Jain (Student)
Institute: Greenwood High International School

I) ABSTRACT:

DNA computing, an emerging field at the intersection of molecular biology and computer science, harnesses the unique properties of DNA molecules for data storage and computational processes. This paper explores the fundamental mechanisms of DNA computing, including encoding binary data using nucleotide sequences (A, T, C, G), leveraging molecular reactions such as hybridization, ligation, PCR, and gel electrophoresis, and executing parallel computations to solve complex problems. Additionally, this paper explores DNA's advantages in memory capacity, energy efficiency, and parallelism. Further, it examines real-world applications, such as cryptography and solving Nondeterministic polynomial time problems like the Traveling Salesman Problem. It discusses the relationship between DNA computing and DNA data storage, emphasizing its potential to revolutionize information processing and long-term data preservation. As advancements in synthetic biology and nanotechnology continue, DNA computing presents a promising alternative to conventional computing, paving the way for future breakthroughs in computation and data management.

II) KEYWORDS:

DNA computing, Data storage, Polymerase Chain Reaction, Adleman's Experiment, Parallel processing, High density data storage, Cryptography, Nanotechnology

III) INTRODUCTION:

As the demand for faster, more efficient, and scalable computing grows, traditional silicon-based computers are approaching physical and computational limits. DNA computing, a revolutionary approach that utilizes DNA molecules for computation, offers a promising alternative by leveraging the natural properties of DNA, such as massive parallelism, high storage density, and energy efficiency. Introduced by Leonard Adleman in 1994 to solve the Hamiltonian Path Problem, DNA computing has since evolved into a field that explores new ways of processing information at the molecular level.

At its core, DNA computing relies on encoding data using nucleotide sequences (A, T, C, G) and utilizing biochemical reactions to perform logical and mathematical operations. Unlike conventional computers, which process data sequentially, DNA computing enables massive parallel processing, making it highly effective for solving complex problems, such as NP-hard[i] optimization tasks and cryptographic computations.

[i] NP stands for nondeterministic polynomial time – these problems are very hard to solve even for modern computers, especially as the size of the problem gets bigger, and have no known polynomial – time algorithms. A problem is NP- hard if it is at least as hard as the hardest problems in the class NP

One of the most compelling applications of DNA computing is its relationship with DNA data storage. With the exponential growth of digital data, traditional storage methods struggle to keep pace due to limitations in space, durability, and energy consumption. DNA, however, can store vast amounts of data in an ultra-compact and stable form, potentially preserving information for thousands of years. As research in molecular computing progresses, DNA-based technologies may redefine the future of information processing and storage.

IV) HISTORY OF DATA STORAGE:

From parchment scrolls to hard drives to cloud computing, data storage has been one of humanities relentless pursuits of preserving and accessing information. As technological advancements occurred, not only did storage capacity increase but the way in which data is recorded, accessed and shared also changed.

The effort at data storage, before computers were developed, were punch cards (developed by Basile Bouchon in 1725). In The late 1940s, magnetic core memory was developed, which became the primary way many early computers wrote, read and stored data. [1] After the Twister Magnetic memory came, in the late 1960s', Semiconductor Memory with Signetics's 8-bit RAM and floppy discs. [2] The introduction of the 1 KB Intel 1103 memory chip in 1971 marked the beginning of the end of the use of the magnetic core in computers and the start of semiconductor dynamic random-access memory (DRAM). [2] In 1980, came the first hard disc drive developed by Seagate. By the 1990s, computers were getting smaller and more lightweight; hard drives, CDs and DVDS, were becoming popular. Then in 2000 came the USB flash drive and in 2006 Amazon launched its Cloud based services. [2] From the late 2000s, the rise of smartphones was accompanied by the rise of cloud services including Dropbox, Microsoft Azure, Apple's iCloud and so on. Around the same time 1TB hard drives were also launched.

V) NEED FOR A NEW STORAGE MECHANISM:

In a world flooded with data, figuring out where and how to store data inexpensively and efficiently becomes a larger problem every day. All of the world's data including digital photos and tweets, financial and legal records, simulations underlying modern science and so on must go somewhere. All this information is stored in digital data centres, commonly thought of as the "cloud". [3] According to the latest estimate, around 402.75 million terabytes of data are created each day and 147 zettabytes of data exists as of 2024. [4] As of 2021, there were about 10 trillion gigabytes of digital data and much of it is stored in enormous facilities known as exabyte data centres, which can be the size of several football fields, cost around $1 billion to build and maintain, require large amounts of water for cooling, and emit large amounts of greenhouse gases into the atmosphere. [5] With the increasing use of IoT devices, real-time data processing and cloud-based storage, the global volume of data is projected to rise to 175 zettabytes by the end of 2025(1 zettabyte = 250 billion DVDs). [14][6] Much of this data is never used, called cold data, but still has to be stored. It consists of archives, old or deleted emails and legal records and so on.

The amount of data that would need to be stored would only increase, facilitated by technological advancement. Therefore, a major problem that must be addressed is how this data can be stored more efficiently. Many scientists believe that a solution lies in the molecule that contains our genetic information: DNA, which evolved to store massive quantities of information at very high density. Theoretically, a coffee mug full of DNA could store the entire world's data. [5]

The idea that individual molecules could be used for computation or storage dates back to 1959, when Richard Feynman presented his ideas on nanotechnology. However, DNA computing was only physically realised in 1994, when American computer scientist Leonard Adleman showed how molecules could be used to solve a computational problem. [7]

VI) WHY SHOULD WE USE DNA:

Living organisms exhibit a highly efficient and compact information storage system, i.e., DNA. Thus, DNA may serve to be an answer to the problem of data storage. [22] This is based on three key properties:

- Under proper conditions, DNA is known to retain data for thousands of years. For example, a 300,000-year-old sample of mitochondrial DNA from a bear was successfully sequenced. [14]
- DNA's energy of operation and storage is many times less that current electronic memories. The bear's retrieved DNA was stored in bone, thereby demonstrating this property. [22][14]
- An individual molecule of single stranded DNA is theoretically capable of storing information at a density of 6 bits for every 1nm of polymer, i.e., approximately 4.5×10^7 GB/g. [14]
- Moreover, the predictability of the pairing and stability of DNA make it ideal for encoding information and performing biochemical operations.

VII) OVERVIEW ABOUT DNA:

A) STRUCTURE:

Nucleic acids, in essence, Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are the organic materials present in all living organisms.

DNA is a group of molecules that is responsible for carrying and transmitting the hereditary materials or the genetic instructions from parents to offsprings.[9] The DNA molecule is a long molecule with a unique double stranded helical structure (which was discovered by the scientists James Dewey Watson and Francis Harry Compton Crick). The basic unit of a DNA molecule is a nucleotide. Each nucleotide has three components: a backbone made up of a pentose sugar (Deoxyribose) and a phosphate group, and a nitrogen-containing base attached to the sugar. There are 4 types of nitrogenous bases: Adenine(A), Guanine(G), Cytosine(c) and Thymine(T)[ii]. Adenine pairs with thymine with a double hydrogen bond and guanine pairs with cytosine with a triple hydrogen bond (refer to images 1 and 2). [9, 10]

Each strand consists of multiple nucleotides. The nitrogenous bases of each strand are complementary to the nitrogenous bases of the other strand (A complementary to T and G complementary to C), in order to maintain helical symmetry. These bases are responsible for storing genetic information. Most of the cell's DNA is stored in the nucleus and is called nuclear DNA. The DNA is present in the nucleus in a super-coiled manner because the DNA molecule is about 2m in length whereas the diameter of the nucleus is only about 6µm. [9, 10]

[ii] A and G are largely classified as purines whereas T and C are classified as Pyrimidines (both purines and pyrimidines are heterocyclic aromatic organic compounds). Purines are 2-carbon nitrogen ring bases, whereas pyrimidines are one-carbon nitrogen ring bases. Therefore, purines are larger due to their two-ring structure.

Image 1: *The structure of DNA*[10]

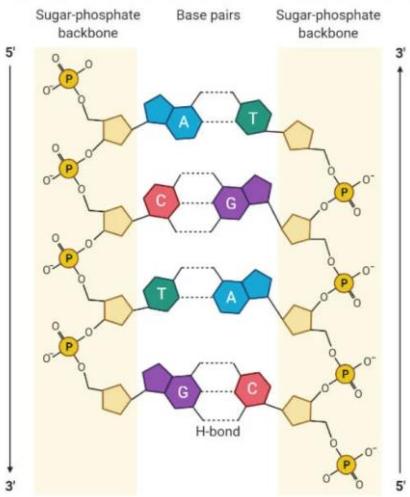
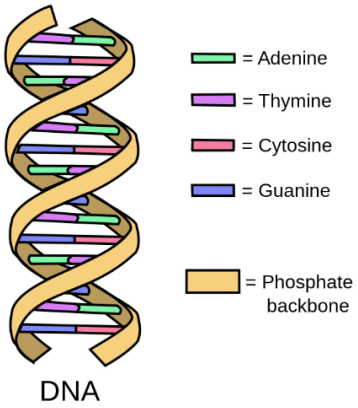


Image 2*: Simplified structure*[37]



B) PROPERTIES OF DNA:

- DNA is made of two helical chains that intertwine with each other to form a double helix (dsDNA).
- The helical chains run anti-parallel to each other and are connected to each other through the hydrogen bonding of the nitrogenous bases.
- The amount of adenine is equal to the amount of thymine and the amount of guanine is equal to the amount of cytosine due to the complementary base pairing. [16]
- The double-helical structure of DNA is highly regular, each turn of the helix measures approximately 10 base pairs. [10]

C) DNA REPLICATION:

DNA replication occurs in the S phase or synthesis phase of the cell cycle[iii]. It the process by which a DNA molecule is copied to produce two identical DNA molecules. [11]

DNA replication occurs majorly in three steps: first the opening or unwinding of the double helix, next the priming[iv] of the template strand and lastly the assembly of a new DNA segment. [12]

During separation of the two strands, the DNA double helix uncoils at a specific location called the origin. This uncoiling process is initiated by the initiator protein. Next, the helicase protein attaches to and breaks apart the hydrogen bonds between complementary bases, thereby creating to free single strands of DNA (ssDNA). These free strands serve as templates for the formation of the new DNA strands. Simultaneously, another enzyme called the primase attaches to each of the free strands and creates a foundation at which replication can begin (refer to images 3 and 4). Primase functions by synthesizing short RNA sequences that are complementary to a single-stranded piece of DNA (the original piece serves as the template). [13] This short stretch of nucleotides is called the primer. Once the primer is in place, another enzyme called DNA polymerase wraps itself around that strand and attaches new nucleotides to the exposed nitrogenous bases (refer to image 5). The polymerase relies on the pool of free-floating nucleotides surrounding the strand in order to build the new strand. A,T and G,C always pair with each other due to their structures, resulting in a phenomenon known as complementary base pairing (CBP). CBP ensures that the newly formed strand is the anti-sequence of the original template strand.

Inage 3: *Unwinding of the double helix through helicase (yellow)*[12]

Image 4*: Primase (red) assembles a primer*[12]

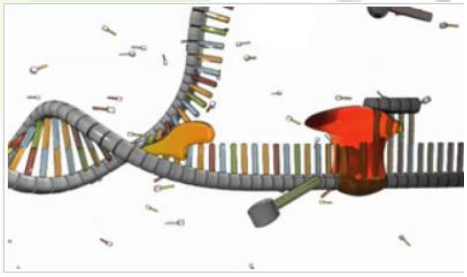


Image 5: *DNA Polymerase, starting at the primer sequence, attaches to the original strand and begins assembling a new complementary stand* [12]

[iii] The cell cycle is the series of events that takes place when a cell grows and divides into two daughter cells.

[iv] The primer is a short nucleic acid sequence (or RNA strand) that provides a starting point for DNA synthesis

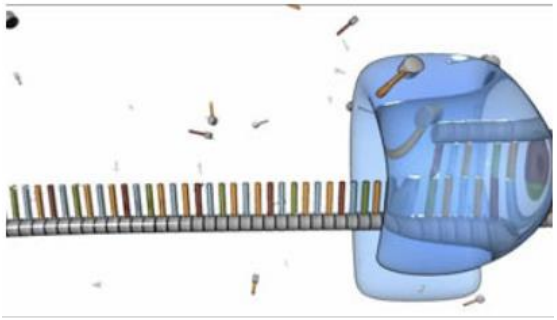
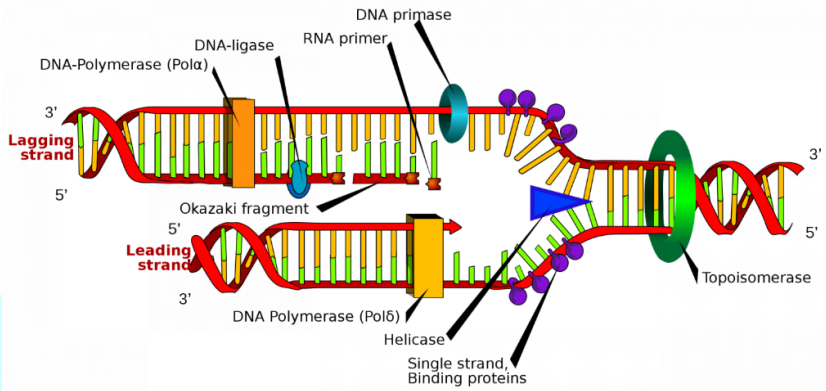


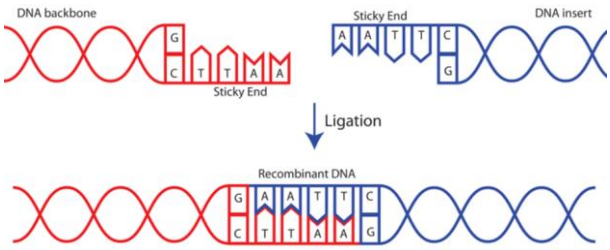
***Image 6**: DNA Replication* [30]



VIII) DNA COMPUTING MECHANISMS:

- **Hybridization:** The concept of hybridization is based on Chargaff's rules.[v] [15] It is the process in which two complementary single strands of DNA or RNA bond together to form a double stranded molecule. [17] This process is facilitated by the formation of hydrogen bonds between the nucleotide bases.
  It is used in computation for data matching and to identify complementary sequences, enabling logical operations like searching for a specific solution.
  It can also be used in data storage to selectively isolate strands of data.
- **Ligation:** Ligation refers to the joining of two DNA fragments through the formation of a phosphodiester bond[vi]. An enzyme known as a ligase catalyses the ligation reaction. In the cell, ligases repair single and double strand breaks that occur during DNA replication. [18]
  It is used in computing as it helps assemble DNA fragments into longer chains that represent solutions or intermediate steps in computation.

***Image 7**: Ligation* [36]

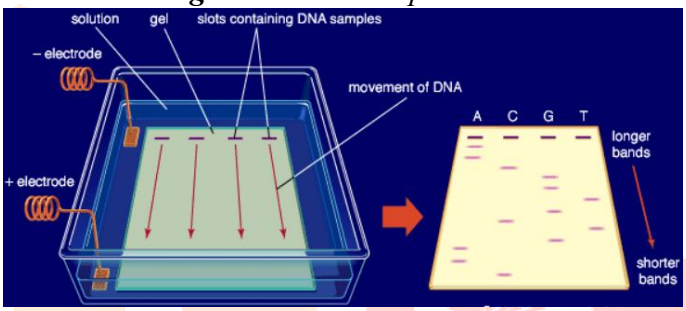


[v] These rules were given by scientist Erwin Chargaff. It states that the amount of purine is always equal to the amount of pyrimidine in a DNA molecule. This means that the amount of adenine is equal to the amount of thymine and the amount of cytosine is equal to the amount of guanine. [16]

[vi] It is a covalent linkage between the phosphate of one nucleotide and the hydroxyl (OH) group attached to the 3′ carbon of the deoxyribose sugar in an adjacent nucleotide.[19] It is essentially a covalent bond that connects nucleotides in a DNA strand, thus forming the sugar-phosphate backbone of DNA.

- **Polymerase Chain Reaction (PCR):** PCR is a laboratory technique for rapidly producing (amplifying) millions to billions of copies of a specific segment of DNA, which can then be studied in greater detail. PCR involves using short synthetic DNA fragments called primers to select a segment of the genome to be amplified, and then multiple rounds of DNA synthesis to amplify that segment. [20]
  It is used in computing to amplify the possible solutions or outcomes for analysis and is also used in algorithms to focus on specific subsets of data. In data storage, it ensures that the desired sequences are present in sufficient quantities for accurate reading of information.
- **Gel Electrophoresis:** It is a technique used to separate DNA fragments (or other macromolecules, such as RNA and proteins) based on their size and charge. Electrophoresis involves running a current through a gel[vii] containing the molecules of interest. Based on their size and charge, the molecules will travel through the gel in different directions or at different speeds, allowing them to be separated from one another. [21]
  It is used in computing in order to identify and isolate DNA strands that represent the correct solution to a particular problem.

***Image 8**: Gel Electrophoresis:*[35]



IX) HOW DNA COMPUTING STARTED:

ADLEMAN'S EXPERIMENT:

The performing of computations using biological molecules rather than silicon chips is known as DNA computing. [7]

The concept of DNA computing was introduced by the USC (University of Southern California) professor Leonard Adleman in the November 1994 Science article, "Molecular Computations of Solutions to Combinatorial Problems". Adleman showed that DNA could be used to store data and even perform computations in a massively parallel fashion (a situation where two or more things are happening at the same time).[8]

Adleman stumbled upon the idea of DNA computing when he noticed how the mechanism of DNA computation was remarkably similar to an early theoretical computer developed by Alan Turing. During replication, DNA polymerase slides along the single DNA strand and similarly in one version of the Turing machine, a mechanism traversed two tapes, reading instructions and giving a result. [8]

Using the four bases of DNA (adenine, thymine, cytosine, and guanine), Adleman encoded a NP-complete problem[viii] known as the Hamiltonian Path problem[ix] into strands of DNA and utilized biological properties of DNA to find the answer. [7][8] Adleman's experiment involved finding a route through a network of towns (labelled '1' to '7') connected by one-way roads. The problem specified that the route must start and end at

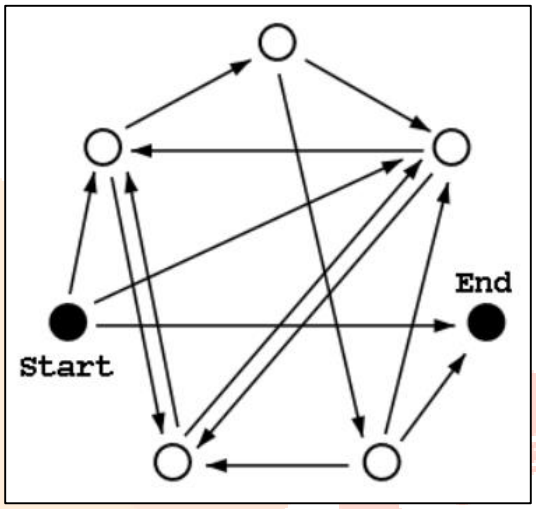
[vii] Gels for DNA separation are often made out of a polysaccharide called agarose, which comes as dry, powdered flakes. At the molecular level, the gel is a matrix of agarose molecules that are held together by hydrogen bonds and form tiny pores.[21]

[viii] A class of problems for which no efficient solution algorithm has been found.

[ix] The Hamiltonian Path problem asks whether a given graph contains a path that visits every vertex exactly once.

specific towns and visit each town only once. Adleman used the Watson-Crick complementarity property of DNA in order to solve this (so the sequence AGCT would stick perfectly to TCGA).[x] Adleman designed short DNA strands to represent towns and roads, with road strands linking town strands to form sequences representing routes. Though most sequences represented incorrect answers to the problem, Adleman used a large enough pool of DNA to ensure that the correct answer would be in his initial plot of strands. The problem then, was to extract this unique solution. This was done using several steps. First, he greatly amplified only those sequences that started and ended at the right towns, using the polymerase chain reaction method[xi]. He then sorted the strands by length, to ensure that only the strands of correct length were retained, using gel electrophoresis. Finally, he repeatedly used affinity purification to pull out those strands that contained every town only once. The strands Adleman was left with were then sequenced to reveal the solution to the problem. [7]

Image 8: *The Hamiltonian Path Problem* [8]



The process of gel electrophoresis involves DNA being forced through a gel matrix (which acts as a molecular sieve), where it encounters more resistance than smaller strands. By applying an electric current to the gel bloc, the negatively charged DNA molecules (negatively charged due to phosphate backbone) are forced through the gel. Smaller fragments move faster through the gel's tiny pores, while larger fragments move more slowly, causing them to separate based on size. Fragments of the same size cluster and form bands that can be detected with ultra-violet light (after being stained by a fluorescent dye). Fragments of the same length cluster together, and the band corresponding to the proper path-length can simply be cut out. [8]

X) HOW DNA STORES AND PROCESSES INFORMATION:

A computer file is essentially a very large number. For example, to store text, each letter or character is encoded with a particular number and these numbers are then assembled in a sequence. To store an image, the amount of red, green and blue (RGB) in each pixel is quantified and these numbers are put one after another. [23]

In conventional computers, data is stored in the binary format, ie, through 0s and 1s. One of the simplest ways of storing binary data through DNA is to convert the 0s and 1s to A, T, G, C. This is essentially the mapping of binary data into quaternary data.

[x] Watson–Crick pairing rules follow two rules of complementarity:
(a) size complementarity (large purines (A and G) pair with small pyrimidines (T and C))
(b) hydrogen-bonding complementarity (hydrogen-bond acceptors, A, pair with hydrogen-bond donors, T).

[xi] Selective polymerase chain reaction is a variant of the standard PCR technique that focuses on amplifying only specific DNA sequences based on carefully chosen primers.

The process of storing information in data involves:

1. **Encoding** – the electricity consumed by computers can be either high or low voltage and can therefore represent the two binary digits, 1 and 0. These two digits are used by the computer to store any file (the binary system is a machine level language that computers can understand directly). On the other hand, DNA is a sequence of four bases A, T, G, C, which can be interpreted as 0, 1, 2, 3. The advantage of this quaternary system is that it uses only half as many digits as the binary system to store data. Therefore, to store a file in DNA, the binary numbers must be converted to quaternary. [23]
2. **Synthesis** – The creation of the DNA molecule in order to store the file is called synthesis. Due to advancements in molecular biology researchers are able to synthesize DNA. During this step, the digital file is first divided into many smaller segments. Then, each segment is chemically synthesized into DNA, and multiple copies of each segment are made. This redundancy allows researchers to compare the copies to ensure that the information is accurately recorded and to identify and correct any errors that might occur during the synthesis process.
3. **Storage –** Many external factors risk degrading DNA. It is, therefore, important to isolate the DNA from oxygen, humidity, light or heat, which could make it unstable, break up the molecules or introduce errors. For this dehydrated DNA is synthesised and stored in an airtight metal capsule. [23]
4. **Sequencing –** technologies have been developed in order to read the information stored in DNA. This is called sequencing. To read DNA, a sequencing machine is used that determines the order of bases. This is then decoded back into the original binary data by applying the same conversion used in encoding.

XI) DIFFERENCES BETWEEN TRADITIONAL AND DNA-BASED COMPUTATION:

- Traditional computing uses binary logic and electrical signals in order to process information through transistors and logic gates. However, DNA computing uses biological molecules to perform parallel processing through biochemical reactions like hybridization and ligation.
- Traditionally, data is stored on silicon- based devices, such as hard drives, that have limited data density. However, DNA molecules have an exponentially higher data storage density.
- DNA performs parallel operations while conventional, silicon-based computers usually handle operations sequentially. A modern CPU follows a repetitive “fetch and execute cycle”, wherein it fetches an instruction and the appropriate data from main memory and executes it. This process is repeated billions of times per second. In general, increasing performance of silicon computing means faster clock cycles, placing emphasis on the speed of the CPU and not on the size of the memory. Oppositely, the power of DNA computing comes from its memory capacity and parallel processing. [25]
- Traditional computing requires significant amounts of power for processing, cooling and data storage. However, DNA computing requires minimal energy, as chemical reactions drive computations without electricity.
- Traditional computing is highly reliable with built-in error correction and precise logic operations. However, DNA computing is prone to sequencing errors and degradation, requiring complex redundancy and error-correction techniques.

XII) ADVANTAGES AND LIMITATIONS:

ADVANTAGES:

- Parallel processing – (a large number of reactions can occur simultaneously) As such, this has the potential of speeding up large, but otherwise solvable, polynomial time problems requiring relatively few operations. DNA can perform countless calculations in parallel. While classical computing quickly reaches a limit of how many parallel computations can be made, DNA computing has almost no limit. This makes it extremely fast and incredibly powerful for scenarios like machine learning. [27]
- Performance rate - Performing millions of operations simultaneously allows the performance rate of DNA strands to increase exponentially. Adleman’s experiment was executed at 1,014 operations per

second, a rate of 100 Teraflops (100 trillion floating point operations per second). The world's fastest supercomputer runs at just 35.8 Teraflops. [25]

- Large storage capability – DNA molecules require just 1 cubic nanometre of space to store one bit of data, compared to traditional storage media, such as videotapes, that require 10^12 cubic nanometres of space to store the same amount of data. In other words, a single cubic centimetre of DNA holds more information than a trillion CDs. [25]
- DNA data storage and computing required very minimal or no power in comparison to traditional silicon-based devices which require large amounts of power. [26]

LIMITATIONS:

- Accuracy - The sequencing process can introduce multiple errors, so it is necessary to store multiple copies of the same strand to cross check information and avoid errors. Further, DNA synthesis is liable to errors, such as mismatching pairs, and is highly dependent on the accuracy of the enzymes involved. [25]
- Speed of synthesis or sequencing of DNA molecules – A single nuclear base takes two minutes to write. The technology used is also very expensive. [23] With current technologies, DNA data storage is expected to cost 800 million USD per one TB of data. [14]
- Requires large amounts of resources in terms of memory - Although DNA can store a trillion times more information than current storage media, the way in which the information is processed necessitates a massive amount of DNA if large-scale problems are to be solved. [25]
- Resource intensive - Each stage of parallel operations requires hours or days, with extensive human or mechanical intervention between steps. Since a set of DNA strands is tailored to a specific problem, a new set would have to be made for each new problem. [25]

XIII) APPLICATIONS:

- Cryptography and data security – Cryptography is used to prevent any unintentional party from reading or manipulating any confidential information by converting it from the sender to the receiver and vice versa. It brings ideal security to the e-commerce industry and the Internet. [28] DNA-based cryptography works largely like classic cryptography, using a private and public key. But, because DNA cryptography is incredibly fast, the keys can be massive. [27] DNA computing is efficacious for cryptography for authentic, uninterrupted communication given the extensive parallelism as well as extraordinary storage volume of a DNA molecule. [28]
    - Steganography: DNA sequences can be used to encode hidden messages that can only be decrypted using a specific key (like a complementary DNA strand).
    - One-time pad encryption: DNA strands can generate extremely large, random keys for secure data transmission.
    - DNA-based digital signatures: Unique DNA sequences can serve as authentication mechanisms, ensuring data integrity and preventing tampering.

- Problem-solving – DNA computing excels in solving complex computational problems, especially NP-hard problems that require testing multiple possibilities simultaneously. It can be used to solve complex graph theory problems, such as finding the shortest path, Hamiltonian paths, and network optimizations. Further, DNA molecules can represent possible solutions to Sudoku puzzles, and chemical reactions can eliminate incorrect possibilities through selective amplification and degradation.

- Medical and biological applications – It can be used for biological data analysis, particularly in genomics and medical research. For example, it can help identify genetic mutations responsible for diseases such as cancer by rapidly analysing large datasets of DNA sequences.

- Cold data storage systems – By leveraging the fact that DNA has a high storage density, is durable, requires minimal energy to maintain, is sustainable and can easily be replicated, DNA can be used for the long-term archival of critical data.
- DNA based logic gates – These are molecular versions of electronic logic gates (AND, OR, NOT), built from specially designed DNA strands. It operates using complementary base paining and strand displacement. This acts as a foundation for computing inside cells or DNA-based devices and underscore DNA's potential in nanoscale computing devices. [33]
- DNA nanotechnology - It is a broad field that uses DNA as a material to build nanoscale structures and devices. DNA origami is a powerful technique within this field that enables the construction of molecular machines and nanorobots. It utilizes DNA for the synthesis of nanoparticles. In this technique, long strands of DNA are folded into a complex scaffold of staple strands having around 200 – 300 nucleotides. Essentially, long single stranded DNA molecules are folded into specific 2D or 3D shapes using shorter DNA strands. This leads to the formation of a complex structure that has characteristic features because of its nanoscale dimensions. Theoretically, DNA origami has the potential to contribute immensely to fields such as drug delivery and cancer therapy. [29]

XIV) RECENT DEVELOPMENTS:

1. **A Primordial DNA Store and compute engine:**

In 2023, researchers from North Carolina State University and John Hopkins university have demonstrated a technology using DNA that is capable of both data storage and computing functions, including repeatedly storing, retrieving, computing, erasing, and rewriting data.

This technology was made possible as scientists created a soft polymer material called soft dendritic colloids (SDCs) that start at a microscale and then branch off from each other in a hierarchical way to create a network of nanoscale fibres. This fibrillar morphology gives it a very large surface area to volume ratio, which allowed the scientists to deposit the DNA among the nanofibrils without sacrificing the data density of DNA. They have binding capacities of over 10^4 TB cm^{-3}. This material was chosen for its compatibility with biological materials, ease of fabrication and low cost. Moreover, it was found that the dendricolloid material helped to preserve the DNA when it was deposited on it. The ability to distinguish DNA information from the nanofibers its stored on allowed the scientists to perform many of the same functions that can be performed with electronic devices. DNA information could be copied directly from the material's surface without harming the DNA and targeted pieces of DNA could be erased and then rewritten to the same surface. Nanopore sequencing allowed the scientists to read the data in RNA after copying it from DNA on the material's surface.

Through this technology, scientists demonstrated that multiple distinct image files could be individually or completely erased and that simplifies 3×3 chess and sudoku problems could be computed and solved. [31, 32]

The process took place in the following steps [31]:

- The SDCs were prepared using cellulose acetate, cellulose and agarose. DNA is immobilized on this high surface area substate and RNA is transcribed from it non-destructively.
- Multiple RNA promoters for non-destructive transcription – multiple synthetic T7 RNA promoters were incorporated into the design. This allows non-destructive readout by transcribing RNA directly from the DNA without altering the original strand. DNA information was copied onto RNA using IVT (in-vitro transcription).
- Once the RNA transcripts were formed, multiple enzymes were used in a programmable manner. In essence, the using the RNA transcripts the system performed enzymatic computations, such as solving 3×3 chess and sudoku puzzles.
- Many copies of a single DNA sequence could be adsorbed onto the SDCs and RNA could be transcribed from them. However, real data files consist of many different DNA strands with

uneven distribution, even during synthesis, and can be further distorted by processes such as PCR and Illumina sequencing. This leads to loss of strands and higher costs of data access and decoding.

- The DNA-SDC matrix is highly stable and can be frozen for long term storage, with a projected half-life of 2 million years at -18°C.

2. A group of scientists at Shanghai Jiao Tong university in China developed a new class of DNA based computing systems called DNA integrated circuits (DICs). They took inspiration from field-programmable gate arrays or FPGAs, the basis of all computing devices, and created DNA-based programmable gate arrays (DPGAs) using short single-stranded DNA molecules as information carriers. These functioned similar to electrons or photons in integrated circuits, enabling logic gates to be reconfigured dynamically.
   Their system successfully implemented a three-layer logic architecture capable of running over 100 billion different circuits. They also demonstrated the DICs ability to solve quadratic equations and detect disease-associated biomarkers by correctly identifying all 23 samples tested in 2 hours. [34]

XV) CONCLUSION:

DNA computing and data storage mark a revolutionary shift in the way we approach information processing and data preservation. By leveraging the extraordinary biochemical properties of DNA, such as its high data density, massive parallelism and self-replication, we open up possibilities beyond the limits of current silicon-based technologies. Though this field at the intersection of biology and computer science is still in its early stages, DNA-based technology promises groundbreaking applications in areas such as cryptography and nanotechnology. As research advances, the challenges of scalability, speed and error correction will continue to be addressed, pushing the boundaries of what is computationally and biologically possible. DNA, the molecule of life, may soon become the molecule of the future's information revolution.

XVI) ACKNOWLEDGEMENT:

I would like to express my sincere gratitude to my parents for their valuable guidance, insightful feedback and unwavering support throughout my journey of writing this paper. I would also like to thank my advisor Mr Prabhat Kumar Tiwari for his constant support.

XVII) REFERENCES:

1. https://www.dataversity.net/brief-history-data-storage/#:~:text=Punch%20cards%20were%20the%20first,textile%20looms%20and%20player%20pianos.
2. https://www.computerhistory.org/timeline/memory-storage/
3. https://www.scientificamerican.com/article/dna-the-ultimate-data-storage-solution/
4. https://explodingtopics.com/blog/data-generated-per-day
5. https://news.mit.edu/2021/dna-data-storage-0610
6. https://rivery.io/blog/big-data-statistics-how-much-data-is-there-in-the-world/#:~:text=By%202025%2C%20the%20global%20volume,%2C%20and%20cloud%2Dbased%20storage.
7. https://www.britannica.com/technology/DNA-computing
8. https://cs.stanford.edu/people/eroberts/courses/soco/projects/2003-04/dna-computing/history.htm
9. https://byjus.com/biology/dna-structure/

10. https://microbenotes.com/dna-deoxyribonucleic-acid/
11. https://www.genome.gov/genetics-glossary/DNA-Replication#:~:text=Definition,with%20its%20own%20complete%20genome.
12. https://www.nature.com/scitable/topicpage/cells-can-replicate-their-dna-precisely-6524830/#:~:text=How%20is%20DNA%20replicated%3F,specific%20location%20called%20the%20origin.
13. https://www.nature.com/scitable/definition/primase-306/#:~:text=Primase%20functions%20by%20synthesizing%20short,before%20DNA%20replication%20can%20occur.
14. https://pubs.acs.org/doi/10.1021/acsnano.2c06748
15. https://study.com/learn/lesson/dna-hybridization-technique-examples.html
16. https://www.toppr.com/ask/question/state-chargaffs-rule/
17. https://www.genome.gov/genetics-glossary/hybridization#:~:text=Definition&text=Hybridization%2C%20as%20related%20to%20genomics,the%20two%20single%2Dstranded%20molecules.
18. https://www.jove.com/v/5069/dna-ligation-reactions-principle-procedure-and-applications#:~:text=Overview,that%20occur%20during%20DNA%20replication.
19. https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/phosphodiester-bond#:~:text=The%20phosphodiester%20bond%20is%20a,%2Dphosphate%20backbone%E2%80%9D%20of%20DNA.
20. https://www.genome.gov/genetics-glossary/Polymerase-Chain-Reaction-PCR#:~:text=Polymerase%20chain%20reaction%20(abbreviated%20PCR,be%20studied%20in%20greater%20detail.
21. https://www.khanacademy.org/science/ap-biology/gene-expression-and-regulation/biotechnology/a/gel-electrophoresis#:~:text=Gel%20electrophoresis-,Gel%20electrophoresis%20is%20a%20technique%20used%20to%20separate%20DNA%20fragments,containing%20the%20molecules%20of%20interest.
22. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-023-05264-6?utm_source=chatgpt.com
23. https://www.youtube.com/watch?v=ZuB3yB7QWCk
24. https://www.sciencedirect.com/science/article/abs/pii/S0065245822000730
25. https://cs.stanford.edu/people/eroberts/courses/soco/projects/2003-04/dna-computing/evaluation.htm
26. https://www.sciencedirect.com/science/article/abs/pii/S0065245822000729#:~:text=DNA%20computing%20has%20many%20advantages,DNA%20chip%2C%20and%20medical%20application.
27. https://www.pluralsight.com/resources/blog/cloud/what-is-dna-computing-technology#:~:text=DNA%20can%20perform%20countless%20calculations,for%20scenarios%20like%20machine%20learning.
28. https://www.sciencedirect.com/science/article/abs/pii/S0065245822000730
29. https://www.sciencedirect.com/topics/neuroscience/dna-origami
30. https://slcc.pressbooks.pub/collegebiology1/chapter/dna-replication/
31. https://www.nature.com/articles/s41565-024-01771-6.epdf?sharing_token=AUSCdQ2JfSO8Me4lX41rGtRgN0jAjWel9jnR3ZoTv0Oyni3oZGiwLZFH4rYm11oXJ5LbPxsfeCGpNQbXyE5vBY_Uz_vg7XWG7X2nKw0gUtvFx-tqfAAxVO42ASijAss-HwHQ2nytEeYtSwX8Nym7SRFXqgThhjxrAxvQpWuA0Xa6HGdhVKNZnWHtxIbZpw_jFXuJOGKPKrW53tKQVH-gsg%3D%3D&tracking_referrer=cen.acs.org
32. https://news.ncsu.edu/2024/08/functional-dna-computing/#:~:text=Here%20we%20present%20a%20DNA,computational%20advantages%20of%20nucleic%20acids.
33. https://www.nature.com/articles/s41598-023-42383-9#Sec6

34. https://www.nature.com/articles/s41586-023-06484-9.epdf?sharing_token=jZ3WsvFNFP-tcA8S9JnkAdRgN0jAjWel9jnR3ZoTv0M2baFtgHnqr44prd171DW-D3vG7dPYY5WJQnW_wx9Q8vEWdA6c1WX92k24L-3HbgUdBWnqFKE_kxkkDfwNqXu9mkPEEKs_gYUDu_4CYKB4brPZVfgLjbbrtn_R0-3EPR8MFlAM8E45XOYTgICThCO34fmZCvlXoknljRTJpVXZn741H_C_b5AU_iHilfB1-lI%3D&tracking_referrer=www.genengnews.com
35. https://www.britannica.com/science/DNA-sequencing
36. https://www.addgene.org/protocols/dna-ligation/
37. https://courses.lumenlearning.com/ivytech-wmopen-nmbiology/chapter/storing-genetic-information/