El Cezeri
Journal of Science and Engineering
ISSN: 2148-3736

Research Article

DNA Data Storage: A Novel Approach to High Density, Long Term Digital Storage

Pallab Banerjee[1a], Mohammad Hashim[1b], Dipra Mitra[1c], Kamal Upreti[2d], Rituraj Jain[3e], Ashwani Kumar[4f], Mohit Kumar[1g]

[1] Department of Computer Science and Engineering, Amity University Jharkhand, Ranchi, India
[2] Department of Computer Science, Christ University, Delhi NCR Campus, Ghaziabad, India
[3] Department of Information Technology, Marwadi University, Rajkot, Gujarat, India
[4] School of Computer Science Engineering Technology, Bennett University, Greater Noida, UP, India

kamalupreti1989@gmail.com.

DOI : 10.31202/ecjse.1663692
Received: 24.03.2025 Accepted: 03.08.2025
How to cite this article:
Pallab Banerjee , Mohammad Hashim , Dipra Mitra , Mohit Kumar , Rituraj Jain , Ashwani Kumar, "DNA Data Storage: A Novel Approach to High Density, Long Term Digital Storage", El-Cezeri Journal of Science and Engineering, Vol: 13, Iss: 1, (2026), pp.(59-70).
ORCID: [a]0000-0003-4264-2979, [b]0009-0002-0764-9163, [c]0000-0001-7609-9965, [d]0000-0003-0665-530X, [e]0000-0002-5532-1245. [f]0000-0002-2100-900X, [g]0000-0002-9004-5856

Abstract With global data volume expected to reach 175 zettabytes by 2025, existing data storage systems face growing issues due to limited durability, high costs, and susceptibility to data loss. DNA, a naturally occurring biomolecule with extremely high storage density and extraordinary stability, has emerged as a promising alternative for long-term digital data preservation. Major challenges remain, including the high cost of DNA synthesis and the slow rates of encoding and data retrieval. Recent advances in DNA nanotechnology, molecular computation, nanopore sequencing, and hybrid silicon-DNA systems are helping to overcome these difficulties. Sequence-based techniques provide exceptional density, while structural DNA approaches allow for dynamic rewriting and reusability. Ongoing research and innovation indicate that DNA-based storage is becoming more feasible in terms of efficiency, scalability, and cost-effectiveness, making it a plausible contender for meeting future data preservation demands.

Keywords: Data density, data retrieval, DNA data storage, DNA nanostructures, hybrid storage systems.

1. INTRODUCTION

Due to exponential data growth, data storage availability is currently a problem. By 2025, the world's data storage needs are expected to exceed 175 zettabytes, and by the end of the decade, they might rise even higher. Specialized applications that call for storage solutions with higher densities and longer lifespans are therefore becoming more and more in demand. The majority of conventional storage technologies, such as optical and magnetic devices, have a restricted storage density and usually have a 50-years data retention limit [1]. Frequent transfers to more modern storage media are frequently required due to rapid data production, which is expensive and time-consuming [2]. Furthermore, the shortcomings of the current storage paradigms such as their poor data density and high energy consumption make the need for better alternatives more obvious.

1.1 DNA as a Storage Medium

Because of its remarkable density and durability Deoxyribonucleic acid (DNA), the molecule that contains all genetic information in living things, is a promising choice for data storage. DNA has demonstrated the ability to encode, transport and preserve the information with minimum energy loss over thousands of years despite the fact that current practical implementations are limited in comparison to established technologies [3]. As an illustration of the medium's exceptional stability, scientists recently retrieved readable DNA from the bone of a 300,000-year-old bear. DNA has an unparalleled informational density in addition to its resilience [4]. Compared to traditional storage media, theoretical models predict a storage potential of approximately 6 bits per nanometer or up to 45 petabytes per gram [5].

$$\text{Density} = \frac{n_{\text{bits}}}{m_{\text{DNA}}} \tag{1}$$

where n_{bits} = number of encoded bits, m_{DNA} = mass of DNA(gm).

1.2 Recent Developments

Lately, there has been a lot of progress in the storage of data in DNA molecules, to form bio digital information systems. It includes converting digital data into a string of DNA sequences, synthesizing new strand of oligonucleotides, preservation of DNA under appropriate conditions and finally, the scanning of the DNA after storage through sequencing [6]. Among the excellent biotechnological instruments employed in this process some of them are Chemical and enzymatic DNA synthesis, PCR for amplification and high throughput DNA sequencing technologies [7]. Because these technologies were first developed in biological aspect, they have greatly contributed to developing the DNA data storage technology [8].

1.3 Challenges

However, DNA data storage is still unable to provide sufficient density of encoding the data as to outcompete several barriers that make it less than ordinary storage systems. Factors affecting the commercial optoelectronics market include high costs arising from DNA synthesis, where large DNA sequences undergo complex synthesis and encoding that must be divided into smaller sections because of the costs, and slow data readout through sequencing. Most currently utilized DNA sequencing technologies, such fluorescence-based and other optical approaches, are expensive and time-consuming, and they frequently need highly qualified individuals to run the machinery and interpret the findings. Also, for the DNA solution, it costs about 800 million USD to archive a terabyte data and only about 15 USD per terabyte using tapes. About the current limitations one of the possible solutions is that instead of storing information in the specific sequence of DNA we can use its three-dimensional properties. One of the recently introduced techniques is DNA nanotechnology, which provides some physical bases for storage: DNA strand is able to form any two- and three-dimensional structure with the help of base-pairing. These structures can be used to encode data, readout with imaging or nanopore measurements, which can potentially decrease the use and costs associated with sequencing. Moreover, as the outline structure can be reprogrammed, the data can be erased and rewritten without other synthesis, which establishes this method as more flexible. The nucleotide sequence and the molecule's structural organization are the two primary approaches to DNA data storage that are examined in this work. It offers a comprehensive evaluation of both systems' benefits and drawbacks. Hence, the research also describes the applicability of DNA data storage that may be useful in the future in areas such as archiving, cryptography, bio barcoding, and DNA computing. However, there are still many challenges that need to be addressed, DNA based storage seems to have great potential in the future developments based on the fact that DNA is very dense and its storage requires low energy input.

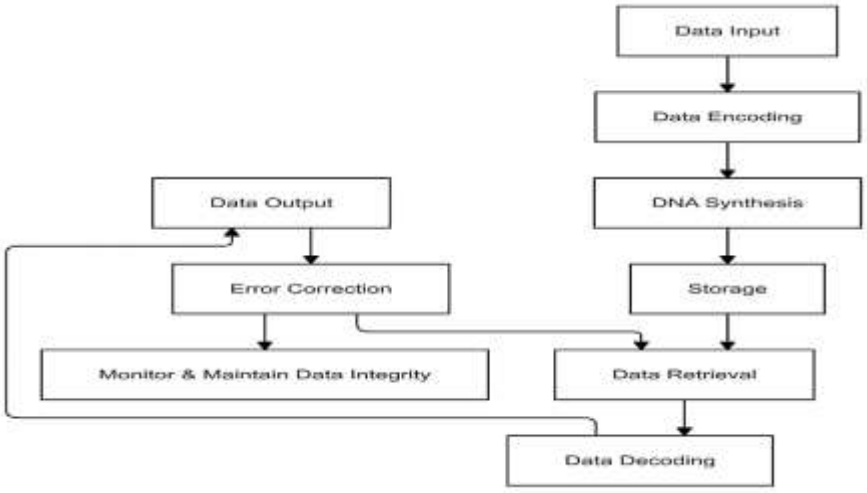

Figure1. Process Flow Diagram.

2. LITERATURE REVIEW

Due to the inability of current technologies to handle the anticipated 175 zettabytes of data worldwide by 2025, DNA has attracted a lot of interest as a possible storage medium [9]. Compared to conventional magnetic and optical storage options DNA provides unmatched data density and stability [10]. Authors in [11] showed the tremendous potential of DNA storage with a theoretical capacity of 2.2 petabytes per gram. It is crucial to remember that different studies' assumptions about DNA storage capacity vary. While some theoretical models have estimated up to 45 PB/g under ideal, lossless conditions, the 2.2 PB/g value represents experimental results that take error correction and practical restrictions into account. A comparison that highlights the variations in context and assumptions is presented in Table 1 in order to resolve this disparity.

Authors of [12] developed this concept and enhanced the DNA's potential as a high-density storage medium by integrating digital information into it. In response to reliability concerns, [13] created the DNA Fountain approach to increase retrieval accuracy and mistake correction. Some strategies to maintain data integrity while expediting the storage and retrieval processes

were proposed by [14]. Despite these developments it is still challenging to produce long and accurate DNA sequences. Enzymatic synthesis shows potential but comes at a greater cost, estimated at $800 million per terabyte, whereas current chemical approaches are inefficient [15]. Although nanopore sequencing has promise for real-time, economical readouts, sequencing methods like as Illumina are still costly [16]. Studies have shown that DNA can store data for thousands of years making it perfect for archival storage under the right circumstances [17]. Structured DNA nanotechnology goes beyond nucleotide sequences by allowing DNA to be programmed into two- and three-dimensional forms for information storage, improving processing and access capabilities [18]. Significant obstacles must be overcome, nevertheless, such as exorbitant prices, sluggish writing rates, and ineffective search engines [19]. New opportunities include the use of AI to optimize encoding and decoding procedures DNA based hybrid memory devices, and the combination of silicon and DNA storage for enhanced performance [20]. New uses might be unlocked by future developments that include biological data storage within living things [21]. The increasing interests in DNA storage is a reflection of its ability to satisfy the expanding need for dependable high density data storage in a variety of fields.

Table 1. Comparison of Reported DNA Storage Capacities

Source / Study	Storage Density	Context / Assumptions
[10]	2.2 PB/g	Practical implementation with error correction and redundancy.
[11]	1.5 PB/g	Huffman encoding with redundancy; real-world use-case.
Theoretical Estimate	45 PB/g	Idealized capacity assuming perfect synthesis, no redundancy, and optimal encoding.

3. METHODOLOGY

3.1 Sequence-Based DNA Data Storage

3.1.1 Data Writing to Encoding

Binary Conversion: For DNA storage, data (texts, pictures, etc.) is encoded into binary (0s and 1s). After that, the binary data is converted into DNA sequences, which normally include between 60 and 200 nucleotides (nt). This length is ideal for minimizing mistakes in chemical synthesis

The Binary to DNA Mapping Formula:

Binary Pair Mapping:

$$00 \rightarrow A,\ 01 \rightarrow T,\ 10 \rightarrow C, \text{ and } 11 \rightarrow G$$

These four DNA bases A, T, C, and G are created from the binary data.

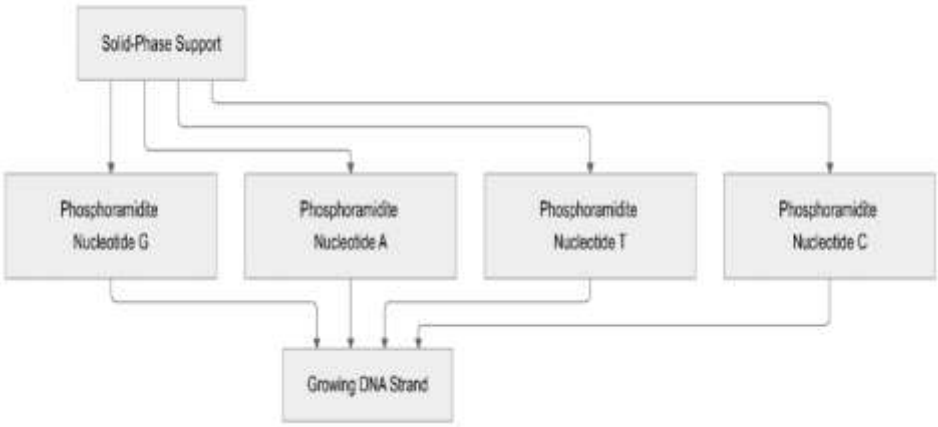

Figure 2: Phosphoramidite Chemistry for DNA Synthesis

DNA Synthesis Method: Phosphoramidite chemistry is used to create DNA strands. Nucleotides are gradually incorporated into the DNA strand by this mechanism. Synthesis Formula:

$$\text{DNAStrand} = \sum_{i=1}^{n} \text{Nucleotide}_i \quad (2)$$

Every Nucleotide_i represents the DNA bases.

Nevertheless, enzyme-based techniques are also taken into consideration (e.g., TdT for antisense oligonucleotide synthesis). These techniques are less efficient in terms of data density, but they could save money and time

During synthesis, speed and error rates are issues for both phosphoramidite and enzyme-based approaches.
Error Detection and Redundancy Formula:

$$\text{ErrorRate} = \frac{\text{Error}}{\text{TotalBases}} \tag{3}$$

Errors may be avoided via the use of redundancy, both logical and physical. While logical redundancy employs error-correcting algorithms to fix mistakes made during data retrieval, physical redundancy entails making several copies of the sequence.

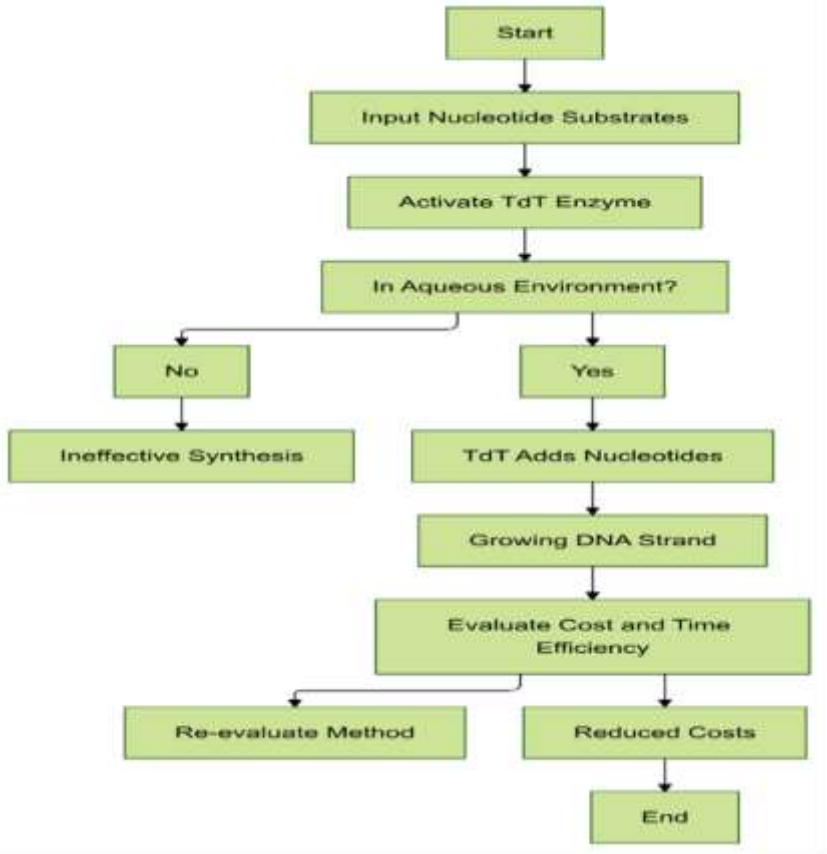

Figure 3: Flowchart illustrating the enzyme-based synthesis of antisense oligonucleotides using Terminal Deoxynucleotidyl Transferase (TdT).

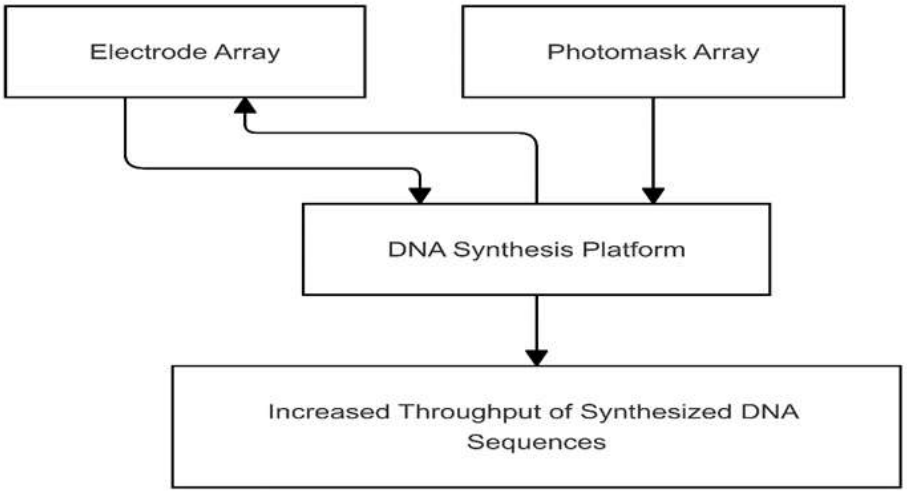

Figure 4: Diagram showing how to overcome the drawbacks of conventional synthesis techniques.

Advanced Techniques: To boost synthesis throughput, new technologies including electrode arrays and photomask arrays are being developed. These techniques enable high-density DNA storage and parallelize the synthesis process.

Using parallel synthesis methods,throughput formula is:

$$\text{SynthesisRate} \times \text{ParallelChannels} = \text{Throughput} \qquad (4)$$

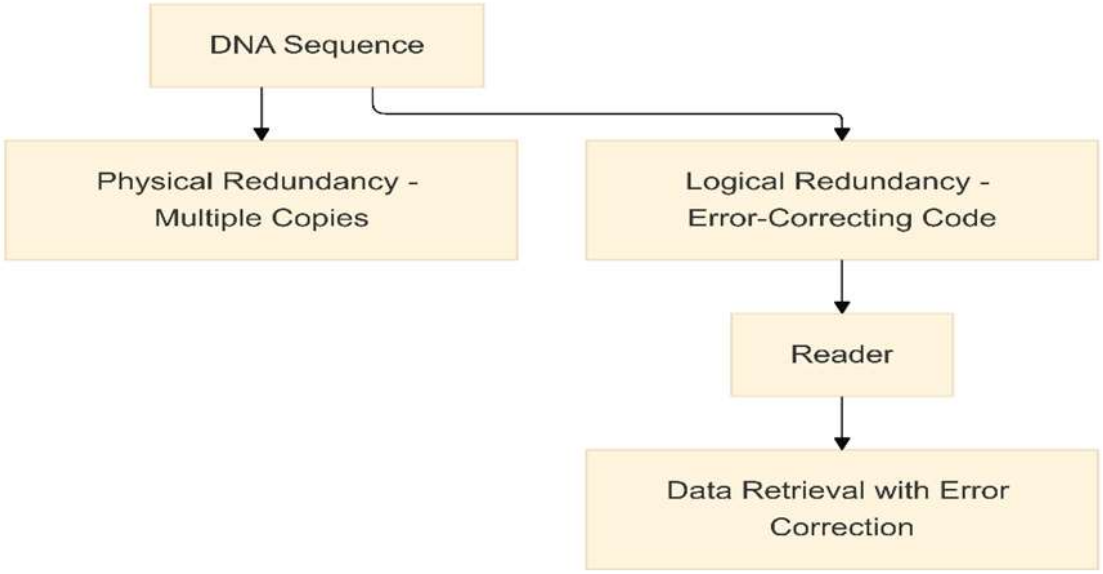

Figure 5: Strategies for Redundancy in DNA Storage Systems

Real-World Example: DNA Encoding of a 1KB Text File

Step 1: Binary Conversion

Each character in a text file is represented by its ASCII value and then converted to binary.

For a 1KB file:

- 1 KB = 1024 bytes
- 1 byte = 8 bits ⇒ 1024 bytes = 8192 bits
- Grouped into pairs: 8192 bits ÷ 2 = 4096 binary pairs

Binary to DNA Base Mapping:

Table 2: Binary to DNA Base Mapping

Binary Pair	DNA Base
00	A
01	T
10	C
11	G

Example:

Character ‘A’ in ASCII = 65 ⇒ Binary: 01000001 Group into pairs: 01 00 00 01 ⇒ DNA Sequence: T A A T

Step 2: DNA Sequence Construction

The resulting 4096 DNA bases are grouped into oligonucleotide sequences of 150–200 bases for synthesis. This chunking allows for manageable strand lengths and reduces synthesis errors.

Step 3: Time and Cost Estimation

Table 3: Estimated time and cost for DNA encoding & sequencing (1 KB file)

Process	Rate / Cost per Base	Cost for 4096 Bases
DNA Synthesis	$0.10 – $0.50	$409 – $2048
DNA Sequencing	$0.02 – $0.10	$82 – $410
Encoding Time	Instant (<1s for 1 KB)	–
Decoding Time	Several hours	–

Pseudocode: Binary to DNA Encoding

Algorithm 1: Binary to DNA Encoding

```
Require: binarySequence {A string of binary digits (e.g., "01001100")}
Ensure: dnaSequence {Encoded DNA sequence (e.g., "TGA")}
dnaSequence ← " "
for i = 0 to length(binarySequence) - 1 step 2 do
   bitPair ← binarySequence[i : i + 2]
   if bitPair = "00" then
      dnaSequence ← dnaSequence + "A"
   else if bitPair = "01" then
      dnaSequence ← dnaSequence + "T"
   else if bitPair = "10" then
      dnaSequence ← dnaSequence + "C"
   else if bitPair = "11" then
      dnaSequence ← dnaSequence + "G"
   end if
end for
   return dnaSequence
```

3.1.2 Problems with Storage and Degradation

DNA has a far shorter half-life under ambient circumstances, even though it may live hundreds of thousands of years in regulated settings. The stability of DNA in aqueous or dry conditions varies from a few months to many years. Because of this, physical storage is essential to preserving the DNA utilized in data encoding over the long term. Strand breakage, UV-induced cross-linking, oxidation, hydrolysis, and mechanical stress are some of the processes that contribute to DNA degradation, which is impacted by environmental conditions. The most common reason for data storage failure among them is hydrolysis. Current storage techniques concentrate on protecting DNA from oxygen and moisture, either at the microscopic level (by encapsulating it in silica or salt matrices) or at the macroscopic level (by drying and sealing it in inert surroundings) in order to prevent deterioration.

Due to the low DNA loading ratios with storage carriers, one significant issue is the loss of storage density, which can cause DNA storage systems' capacity to decrease by up to 1,000 times. Complexity is further increased by the expense and duration of DNA preservation, particularly when considering the PCR-restriction of oligonucleotide pools, which restricts the quantity of data that can be kept in a single pool. These issues, however, are less important for long-term storage than they are for temporary fixes. With a hypothetical pool size of 5.5 terabytes, we looked at the storage densities and half-lives of several macroscopic & microscopic storage techniques in Table 4.

Table 4: Comparison of DNA Storage Approaches

Storage Technique	Longevity (at 10 °C) / years	DNA Concentration / PB/g	Effective Data Density / PB/g
Macroscopic			
In Aqueous	18	0.0049%	0.83
Dried Form	7	100%	18,000
Bone Structure	1700	0.05%	8.5
DNA Shell	>100,000	0.00002%	0.0035
Microscopic			
Trehalose Matrix	160	0.2%	3.0
Silica Nanoparticles	520	3.8%	580
Polymer Matrix	110	0.13%	18
Salt Matrix	730	22%	3800
Calcium Phosphate	580	17%	3040
Silk Protein Matrix	NA	0.00045%	0.0047

Furthermore, the introduction of longer oligonucleotides and improvements in third-generation sequencing have made DNA degradation more difficult since longer sequences degrade more quickly, leading to higher mistake rates. However, sequence length has little effect on microscopic and macroscopic storage techniques. Even though the optimal density for storing DNA is near to its natural structure, actual storage practices sometimes result in a decrease in density, which occasionally outweighs the

advantages of sophisticated encoding techniques. Denser storage devices should be the focus of future research to balance preservation stability with encoding density.

3.1.3 DNA Storage Subpool Access Mechanisms

Without completely sequenced the pool, DNA storage systems must permit selective access to subsets of DNA. To extract certain sub pools, a precise addressing method is necessary. Physical separation and PCR-based addressing are the two main techniques for gaining access to DNA sub pools.

PCR Based Addressing: This method enriches specific DNA sub pools using specified primers. PCR suffers from data loss during re-amplification, particularly as pool size increases, yet it can store up to 10^10 sequences per reaction. The larger the DNA pool, the less effective and efficient it is.

PCR Amplification Efficiency Formula:

$$A = A_0 \times 2^n \tag{5}$$

where:

- A = Final number of copies
- A_0 = Initial number of copies
- n = Number of cycles

PCR Amplification Loss Formula:

$$L = (1 - \text{Efficiencyfactor})^{\text{cycle}} \tag{6}$$

Physical Separation This technique extracts specific sequences by employing magnetic beads and primers tagged with biotin. While using DNA barcodes in silica particles increases scalability, it also raises questions about how long-lasting the barcodes will be.

Success Rate:

$$S = \frac{A_t}{A}, \quad A_t = \text{Numberoftargetedsequences} \tag{7}$$

3.1.4 Data Retrieval Techniques

Retrieval techniques are crucial to DNA's commercial viability as archival storage. One popular strategy for DNA sequencing is Illumina's Sequencing by Synthesis (SBS) technology. It offers precise data extraction, which is essential for storing DNA. Oxford Nanopore Technologies' Nanopore Sequencing is another intriguing technique. Larger molecular alphabets and real-time data capture are made possible by its ability to read DNA as it passes through a nanopore using electrical impulses. However, compared to SBS (0.5%), nanopore sequencing has a greater error rate (10%), requiring extra processing. With developments in solid-state nanopores, nanopore sequencing has the potential to increase sequencing speed and efficiency even though its data retrieval capacity is smaller (1.67 MB compared to 200 MB with SBS).

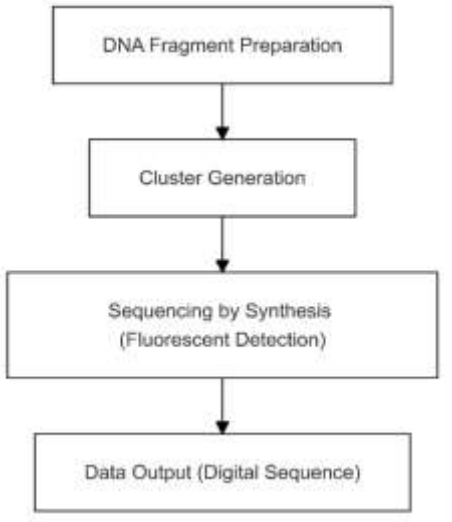

Figure 6: Diagram illustrating Illumina's Sequencing by Synthesis.

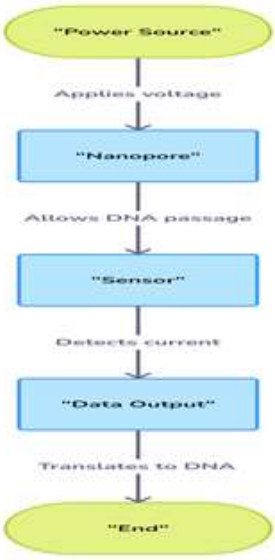

Figure 7: Flowchart illustrating the process of DNA sequencing.

3.1.5 Error Correction and Decoding

Because of faults in synthesis, storage, amplification, and sequencing, DNA data storage is prone to errors. The most common forms of these mistakes are deletions, substitutions, and infrequently, insertions. Error-free decoding is hampered by the gradual loss of sequence information. Redundancy is required when DNA deteriorates, yet this lowers storage density.Error correction codes, such as Reed-Solomon codes, are essential for reducing mistakes that occur when retrieving data. Accurate data recovery is made possible by efficient error correction, which guarantees little redundancy. This is accomplished by combining outer codes to repair sequence loss and inner error correction codes for nucleotide errors. The best encoding and clustering techniques for DNA storage, enhancing recovery from noisy readings, and tackling issues like insertion/deletion channels are the main topics of recent study. The best methods for storing DNA for an extended period are still being researched, though.

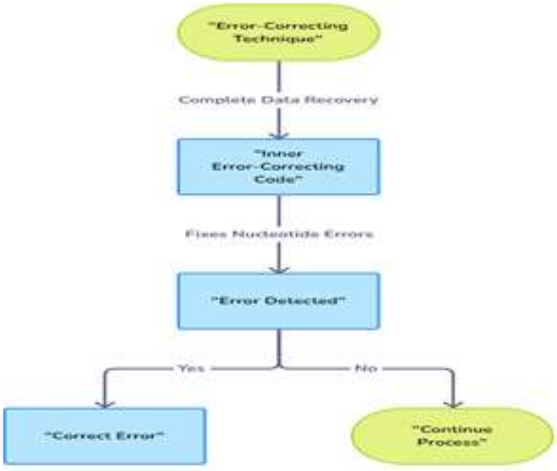

Figure 8: Flowchart emphasizing the order of error-correcting algorithms.

3.1.6 Limitations

1. Cost-Related Concerns:
    - Currently, DNA storage is costly, encoding one terabyte of data costs over $800 million, while standard tape storage only costs $16.
    - Yet, improvements in error-correcting codes and synthesis methods may drastically reduce costs; some projections indicate that prices may fall to $0.01 per terabyte.

- An affordable substitute that might drastically lower synthesis costs is enzymatic DNA synthesis.

2. Problems with Process Time Scales:
   - The present write rates for DNA data storage are in kilobytes per second region, which is slower than traditional systems.
   - Over the next ten years, read/write speeds must increase by orders of magnitude for DNA storage to compete with commercial systems.
   - To solve these performance concerns, research on synthetic polymers is still under progress.

3.2 Structured-Based DNA Data Storage

3.2.1 Synthetic DNA Sequence vs DNA Nanotechnology

An inventive way to get around the drawbacks of conventional DNA data storage is using DNA nanotechnology. Data storage may be made simpler by taking use of DNA nanostructures' self-assembling capabilities, which lessens the need for intricate synthesis and next-generation sequencing. The fundamental idea is that DNA strands may self-assemble into 2D or 3D forms, providing easy production and exact control. For nanoscale data storage, DNA nanotechnology uses three primary methods: wireframe structures, DNA origami, and DNA tile assembly. The most often used of these is DNA origami. Instead of directly storing data in the base sequences, these nanostructures enable data to be stored in their 3D topologies.

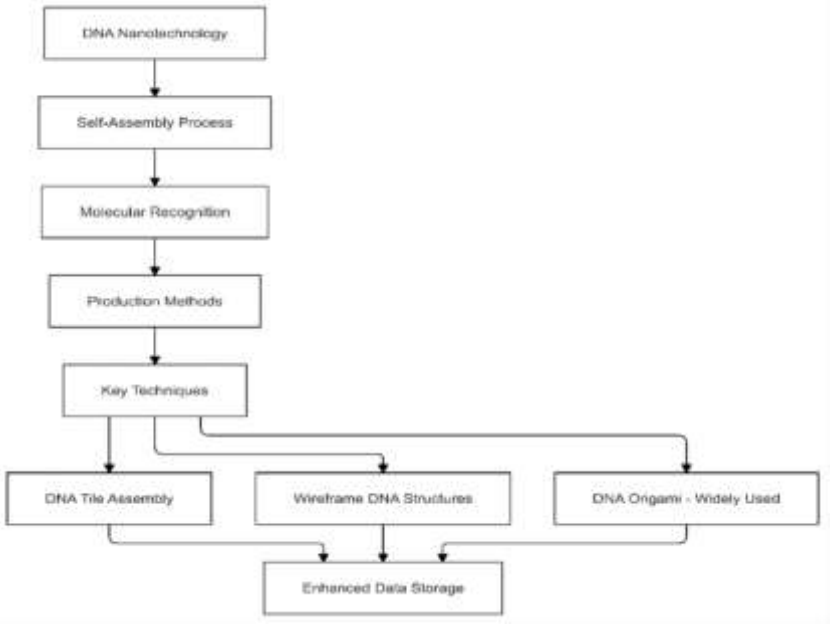

Figure 9: Flowchart showing data storage techniques using DNA nanotechnology.

3.2.2 Data Storage Using DNA Nanostructure

DNA nanostructure data storage uses both special writing and reading procedures, in contrast to conventional sequence-based systems. While DNA origami provides a solution with massive scaffold strands folded into exact forms using short "staple" strands, conventional DNA synthesis is sluggish and expensive. This makes it possible to encode data into these 2D and 3D forms.

Important Methods for Describing These Structures:

- Gel Electrophoresis: Due to its bulk nature, gel electrophoresis has a limited data capacity and consumes more DNA, but it is useful for recognizing DNA nanostructures based on size and shape.
- Fluorescence: While single-molecule and bulk fluorescence improves data retrieval capabilities, they have issues with spatial addressability.
- Atomic Force Microscopy (AFM): AFM helps in binary information writing, such as the "DNA Braille" method, and visualizing three-dimensional DNA origami constructions.
- Electron Microscopy (EM): High-resolution imaging of hybrid DNA structures is possible with electron microscopy (EM), and throughput is increased by recent developments in liquid cell and cryoEM.
- Nanopores: By transforming DNA structure into electrical signals, nanopore technology provides a revolutionary way to read DNA nanostructures, increasing readout speed and accuracy.

3.2.3 Molecular Computation via DNA Nanotechnology

The possibility of reliable, long-term data preservation is presented by DNA data storage. Hybrid electronic-biomolecular computing systems may result from the combination of DNA encoding with conventional electronic systems. The computing capabilities of DNA have been significantly enhanced by the development of bigger, programmable DNA tiles, which now allow computations on par with those carried out by conventional computers. Recent developments that combine deterministic logic with indeterministic computing, such as the development of cellular automata using DNA, demonstrate DNA's ability to perform intricate computational tasks. To sum up, DNA nanotechnology is a revolutionary platform for molecular computation and data storage. Researchers are opening the door to more effective data processing and storage systems by utilizing DNA's special qualities and cutting-edge methods.

Even if the aforementioned methods such as DNA Braille, DNA-based cellular automata, and programmable DNA tiles represent encouraging developments in the field of DNA nanotechnology, many of them are still in the conceptual or simulation stage. Although theoretical frameworks and current literature serve as the foundation for these models they have not yet undergone complete validation in real-world laboratory tests or large-scale deployments. Such methods are not meant to be established or economically viable technologies rather their inclusion in this study is meant to demonstrate the possible paths of DNA-based computation and storage. Experimental findings and the empirical support for these methods continue to be a crucial area of future study.

4. RESULTS

A comparison between conventional storage systems and DNA data storage is provided in this section. The capacity, speed, cost, and lifespan variations between various technologies are shown in Table 5, and the trends in storage efficiency over time are shown in Figure 10.

Table 5: DNA Data Storage vs.Traditional Storage

Feature	DNA Data Storage	Traditional Storage
Storage Space	Up to 200 MB (potential for TB)	Varies (GB to TB)
Density	Extremely high	Moderate
Cost	High and decreasing	Low
Speed	Slow	Fast
Durability	Long-lasting	Varies
Accessibility	Limited	Widely Accessible
Error Rates	High	Low

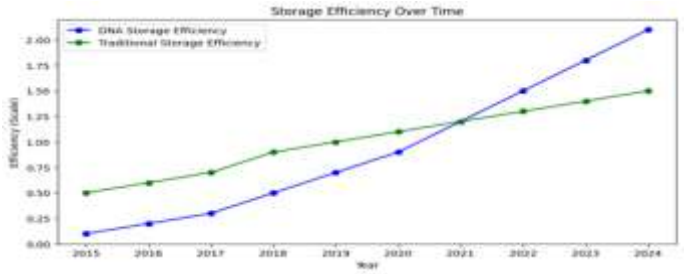

Figure 10: Trends in Storage Efficiency Over Time

With an emphasis on cost, read/write speed, and longevity, the tables and graphs that follow provide a thorough summary of the advantages and disadvantages of each approach.

Table 6: Cost Comparison of Storage Technologies

Storage Technology	Cost per GB	Estimated Cost (1 TB)
DNA Data Storage	$1 Billion(estimated)	$1 Trillion
SSD	$0.8 - $1.2	$50 - $150
HDD	$0.3 - $0.5	$20 - $50

Table 7: Lifespan and Durability Comparison

Storage Technology	Lifespan	Durability	Env. Stability
DNA Data Storage	1000+	100,000	High
SSD	5-10	3,000-100,000	Moderate
HDD	3-5	1,000	Low

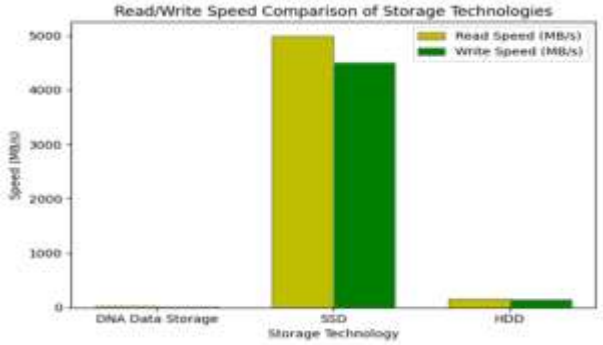

Figure 11: Read/Write Speed Comparison

The graph that follows Figure 12. compares modern approaches and cutting-edge tactics for storing DNA data, emphasizing important elements including cost, speed, scalability, data integrity, and innovation potential.

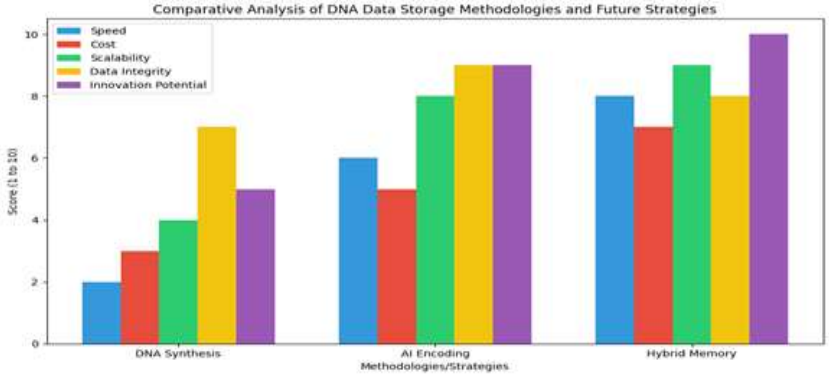

Figure 12: Comparative analysis of current methodologies and futuristic strategies in DNA data storage

Since accurate sequencing impacts data integrity and retrieval, it is essential for the dependability of DNA data storage. Figure 13 illustrates how advances in sequencing contribute to the viability of DNA storage in comparison to conventional systems by comparing the accuracy of three key sequencing technologies: Sanger Sequencing, Next-Generation Sequencing (NGS), and Nanopore Sequencing.

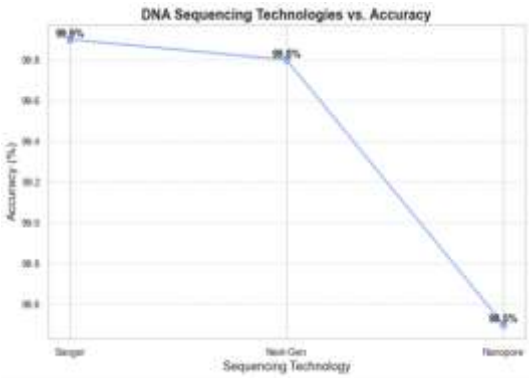

Figure 13: Sequencing Accuracy Comparison

5. CONCLUSION

The promise of DNA data storage, highly dense and durable, has been slightly pushed away from traditional storage owing to slow writing speeds and exorbitant costs. Loading a DNA chip, at least for now, takes almost 24 hours for a 200MB capacity. Researchers are now working to take this up to terabyte levels by improving its encoding, developing cost-effective means, etc.

Real-world Applications: DNA data storage offers great potential in real-world situations despite its present drawbacks particularly in domains where security, density and long-term retention are crucial:

- Long-term Digital Archiving: Perfect for conserving historical and cultural information, including government documents, old manuscripts, and museum collections.
- Military and Defense: Enables highly secure, tamper-resistant and long-duration storage of classified or mission-critical information.
- Medical Data Repositories: Suited for compact and stable storage of large genomic datasets and long-term health records.
- Space Missions: Useful for storing large volumes of data in spacecrafts due to DNA's low weight, stability and radiation resistance.

The outlook seems bright for DNA and archival storage; making it more dynamic in that it would allow updates, deletions, and automation will only drive its realization further. Breakthroughs in chemical production and sequencing technologies may further speed up such development while minimizing errors during the process. Wrong solid-state nanopores and optical techniques appear to be almost there in terms of reading with a degree of accuracy. Such hurdles notwithstanding, innovations in AI based coding schemes that hybrid memory systems might bring can turn DNA storage upside down. Proper cross-team synergy between technology and science could soon make DNA storage genuinely revolutionary.

Authors' Contributions

In this study, authors contributed equally to the study.

Competing Interests

The authors declare that they have no conflict of interest.

Funding Statement

No fund was received for this work.

References

[1] Lai Q, Hua H. Secure medical image encryption scheme for Healthcare IoT using novel hyperchaotic map and DNA cubes. Expert Systems With Applications 2024;264:125854. https://doi.org/10.1016/j.eswa.2024.125854.
[2] Liu B, Wang F, Fan C, Li Q. Data readout techniques for DNA-Based information storage. Advanced Materials 2025;37:e2412926. https://doi.org/10.1002/adma.202412926.
[3] Mitra A, Mitra P, Mahadani P, Trivedi S, Banerjee D, Das M. Application of character based DNA barcode: a novel approach towards identification of fruit fly (Diptera: Tephritidae) species from cucurbit crops. BMC Genomics 2025;26:70. https://doi.org/10.1186/s12864-025-11261-1.
[4] Xu X, Wang W, Ping Z. Biotechnological tools boost the functional diversity of DNA-based data storage systems. Computational and Structural Biotechnology Journal 2025;27:624–30. https://doi.org/10.1016/j.csbj.2025.02.002.
[5] Wang S, Mao X, Wang F, Zuo X, Fan C. Data storage using DNA. Advanced Materials 2023;36:e2307499. https://doi.org/10.1002/adma.202307499.
[6] Yang S, Bögels BWA, Wang F, Xu C, Dou H, Mann S, et al. DNA as a universal chemical substrate for computing and data storage. Nature Reviews Chemistry 2024;8:179–94. https://doi.org/10.1038/s41570-024-00576-4.
[7] Yu M, Tang X, Li Z, Wang W, Wang S, Li M, et al. High-throughput DNA synthesis for data storage. Chemical Society Reviews 2024;53:4463–89. https://doi.org/10.1039/d3cs00469d.
[8] Mao C, Wang S, Li J, Feng Z, Zhang T, Wang R, et al. Metal–Organic frameworks in microfluidics enable fast Encapsulation/Extraction of DNA for automated and integrated data storage. ACS Nano 2023;17:2840–50. https://doi.org/10.1021/acsnano.2c11241.
[9] Doricchi A, Platnich CM, Gimpel A, Horn F, Earle M, German Lanzavecchia, et al. Emerging Approaches to DNA data Storage: Challenges and Prospects. ACS Nano 2022;16:17552–71. https://doi.org/10.1021/acsnano.2c06748.
[10] Masanet E, Shehabi A, Lei N, Smith S, Koomey J. Recalibrating global data center energy-use estimates. Science 2020;367:984–6. https://doi.org/10.1126/science.aba3758.
[11] Organick L, Ang SD, Chen Y-J, Lopez R, Yekhanin S, Makarychev K, et al. Random access in large-scale DNA data storage. Nature Biotechnology 2018;36:242–8. https://doi.org/10.1038/nbt.4079.
[12] Erlich Y, Zielinski D. DNA Fountain enables a robust and efficient storage architecture. Science 2017;355:950–4. https://doi.org/10.1126/science.aaj2038.
[13] Seeman NC, Sleiman HF. DNA nanotechnology. Nature Reviews Materials 2017;3. https://doi.org/10.1038/natrevmats.2017.68.
[14] Grass RN, Heckel R, Puddu M, Paunescu D, Stark WJ. Robust Chemical Preservation of Digital Information on DNA in Silica with Error-Correcting Codes. Angewandte Chemie International Edition 2015;54:2552–5. https://doi.org/10.1002/anie.201411378.
[15] Yazdi SMHT, Gabrys R, Milenkovic O. Portable and Error-Free DNA-Based data storage. Scientific Reports 2017;7:5011. https://doi.org/10.1038/s41598-017-05188-1.
[16] Kosuri S, Church GM. Large-scale de novo DNA synthesis: technologies and applications. Nature Methods 2014;11:499–507. https://doi.org/10.1038/nmeth.2918.
[17] Goldman N, Bertone P, Chen S, Dessimoz C, LeProust EM, Sipos B, et al. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. Nature 2013;494:77–80. https://doi.org/10.1038/nature11875.
[18] Church GM, Gao Y, Kosuri S. Next-Generation Digital Information Storage in DNA. Science 2012;337:1628. https://doi.org/10.1126/science.1226355.
[19] Mardis ER. The impact of next-generation sequencing technology on genetics. Trends in Genetics 2008;24:133–41. https://doi.org/10.1016/j.tig.2007.12.007.
[20] Kasianowicz JJ, Robertson JWF, Chan ER, Reiner JE, Stanford VM. Nanoscopic porous sensors. Annual Review of Analytical Chemistry 2008;1:737–66. https://doi.org/10.1146/annurev.anchem.1.031207.112818.
[21] Rothemund PWK. Folding DNA to create nanoscale shapes and patterns. Nature 2006;440:297–302. https://doi.org/10.1038/nature04586.