Journal of Theoretical and Applied Information Technology
15th August 2021. Vol.99. No 15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SYNTHETIC DNA AS A SOLUTION TO THE BIG DATA STORAGE PROBLEM

[1]MANAR SAIS, [2]NAJAT RAFALIA, [3]JAAFAR ABOUCHABAKA

[1,2,3]Laboratory of Computer Science Research, Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco

E-mail: [1]manar.sais@uit.ac.ma, [2]arafalia@yahoo.com, [3]abouchabaka3@yahoo.fr

ABSTRACT

In the last few years, we have witnessed unprecedented growth in data and gigantic amounts of data are being produced every day. By 2020, the amount of information we want to store it will be around 44 trillion gigabytes. on the one hand the data volumes continue to grow at an even higher speed, However, our traditional databases are limited in the storage and processing of this large and complex data and we do not have a reliable physical storage medium that can withstand the weather. On the other hand, the term Big Data is now the new natural resource and current analysis architectures face much greater challenges in terms of scalability, rapid ingestion, performance, processing and storage efficiency. In order to cope with these massive and exponentially increasing amounts of heterogeneous data generated more and more quickly, many researchers believe that they have found the solution to this problem, either develop and add an intelligent touch to the available technology, or have discovered solution effective in the field of chemistry, DNA for example. DNA molecules as information carriers have many strengths over traditional storage media. Its high storage density, large capacity, long term stability, potentially low maintenance costs, and other excellent features make it an ideal alternative for information storage, and it is expected to provide wide practicality in the future. In this article, we present a DNA storage technology designed to optimize the storage of huge amounts of data or what is called Big Data. We study its storage capacity for large amounts of data, its operating principle and its added value compared to available technologies.
Keywords: *Storage; Big data; DNA, Datasets;*

1. INTRODUCTION

Data is the dominant player in today's society, and the "digital universe" (all the world's digital data) is expected to reach more than 16 ZBs in 2017 What is shocking is that even though we plan to improve storage technology, the exponential growth rate can easily exceed our storage capacity. A significant fraction of this data is in the form of archives, for example, Facebook recently set up a comprehensive data center dedicated to 1 exabyte of refrigerated storage [1] Over the years, we have stored data on disc players and tapes. Almost all of the information we consume depends on the data collected from our transactions, and our data even determines what news is shown to us. However, as the demand for data storage has increased exponentially, today's data storage technology strives to keep pace with its time.

In order to meet the storage demand companies and researchers continue to optimize current technologies, others have proposed fundamentally new methods of data storage. One of the most recent methods is encoding information in DNA, a process called gene data storage. DNA is nature's hard drive and the permanent recording of genetic information written in chemical language. There are only four letters in the DNA alphabet, and these four nucleotides are usually abbreviated to A, C, G and T. When arranged in different ways, these letters set different instructions for our cells. The entire manual of instructions for our existence exists in 3 billion of these letters that make up the human genome. All this information is stored in every cell in our body. DNA is millions of times smaller than a computer's hard drive. Compared to current storage systems, genetic data storage offers a much higher density of information (how much

data can be stored per unit of space) as well as a longevity (how long can the data be stored without decomposition).

Davis [2] is mentioned that the idea of storing data in synthetic strands (sequences) of DNA has existed since 1988 and that research on DNA-based data storage has developed rapidly in recent years with the advancement of DNA synthesis and sequencing technology. DNA storage has several competitive advantages such as long shelf life, extremely high density and low energy consumption.

The main aim of the paper is to present synthetic DNA as a new data storage technology and a lasting solution to storage problems, and to achieve these objectives our article is structured as follows; we first briefly present DNA storage technology as a future solution to the big data storage problem. The second part is intended to describe the important steps in the storage process that allow the conversion of digital data into genetic data, namely synthesis and sequencing operations. The third part is devoted to the benefits of DNA storage technology, which demonstrates the incredible storage capacity of thousands of gigabytes in a totally reliable way, and can be used as a solution to store as much data as possible in the smallest space. Not to mention the major challenge that is holding back the popularization of DNA technology, namely the high cost of DNA synthesis and the problem of rapid access. The last part presents an experiment that aims to convert the data into a DNA series, and decode series of DNA (FASTA files) into original text files, and end with a comparison of the storage density of synthesized DNA with traditional storage media.

2. DNA AS STORAGE MEDIUM: OVERVIEW AND RESEARCH

The emerging scientific field of DNA has the potential to host annual global digital information weighing only four grams, and has a unique information storage potential, as a large amount of information can be recorded (synthesized) and read (sequencing) at a moderate cost, and rapid decline (in theory, DNA can store up to 455 EB/g3 [3].

One of the most fundamental articles in the history of biology in nature is published in 1974 by Waston and Crick [4], revealing the structure of DNA molecules as a carrier of genetic information. Since then, people have realized that biological genetic information is stored in a linear sequence of four bases in DNA. In 1988, the artist Davis first tried to build a real DNA storage [2]. He converted the pixel information of a "Microvenus" image into a 0-1 sequence arranged in a 5 × 7 matrix, where 1 represents a dark pixel and 0 represents a bright pixel. This information is then encoded in the form of a DNA molecule of 28 base pairs (bp) and inserted into E. coli. After restoration by DNA sequencing, the original image was successfully restored. After Clelland proposes to use methods based on "micro-points of DNA" (such as steganography) to store information in DNA molecules [5]. Two years later, a proposal was made in the same way that the coding of amino acid sequences in the DNA Bancroft proposed a method of direct coding of English letters using DNA bases [6]. In just a decade researcher have proposed the concept of storing specific information in DNA (Dong et al., 2020). However, the concept was not realized because DNA synthesis and sequencing technology is still in its infancy.

Recently, researchers managed to store up to 659KB of data in a DNA molecule, while the maximum amount of data previously stored was less than 1KB. The molecule can store more information, reaching 739KB. It should be noted that the stored data contains not only text, but also images, sounds, PDF files, etc., confirming that DNA can store several types of data.

3. THE FUTURE SOLUTION TO THE DATA STORAGE PROBLEM: SYNTHETIC DNA

The volume of computer data storage is a challenge thanks to the intensive use of social networks and cloud computing, there is a paradigm shift in the volume of data produced, but the persistence of data over time is another. Because all the storage technologies we have developed today have limited life cycles and are sensitive. It is estimated that by 2021 [25], 35 Zettabytes of digital information will be generated. This highlights a major concern to store and maintain the rapid growth of data that forces researchers to experiment with other storage methods and design a new architecture to store a large amount of data sustainably such as DNA [7]. The idea of biological storage of DNA information comes into play to address these problems and this demand continues and considers it an ideal storage medium of choice for the next generation because of their durability and high density of information [24].

Journal of Theoretical and Applied Information Technology
15th August 2021. Vol.99. No 15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 1: DNA-Based Data Storage-The Big Data Storage Solution Of The Future

3.1. The DNA Molecule

DNA is the natural molecule of storage of information which encodes our genetic information. Owing to his huge sequence, it is compressed in an invigorating task, it is therefore resolution preferred to treat big quantities of data [7] DNA uses four foundations A (Adenine), C (Cytosine), G (Guanine) and T (Thymine) to stock genetic information. As a result, the sequence of DNA is only a combination of four foundations (In, C, G, T). The sequences of DNA transport information on the expression of various organisms [8]. The stocking of DNA is similar to the stocking of numerical CD, who uses the ground and mines represented by 0 and 1 in footprint in spiral to stock information. The potential of DNA as hard disk is represented well in [9].

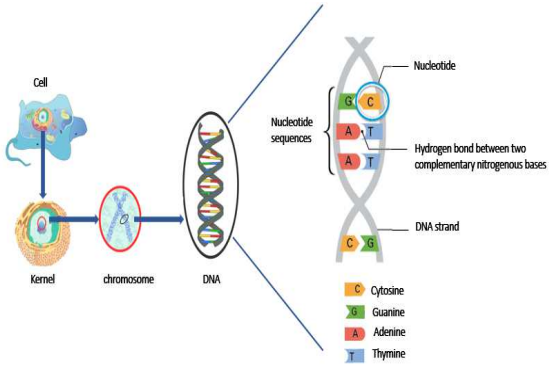

Figure 2: The DNA molecule

3.2. Structure of DNA Molecule

The DNA macromolecule is composed of two strands that are twisted around a common axis into a shape called a double helix, which resembles a twisted scale. the bars of the scale are composed of nitrogen base pairs (base pairs) and the sides of the scale are composed of an alternation of sugar molecules and phosphate groups [29].

Each strand of DNA is a polynucleotide composed of units called nucleotides. The important components of each nucleotide are nitrogen bases, 5-carbon sugars, and phosphate groups. The sugar in DNA nucleotides is called deoxyribose-DNA is an abbreviation for deoxyribonucleic acid. The sugar of a nucleotide is covalently bound to the phosphate group of the next nucleotide to form the sugar-phosphate skeleton of the DNA chain.

Each nucleotide is named according to its nitrogen base. These nitrogenous bases can be purines, such as adenine (A) and guanine (G), or pyrimidines, such as cytosine (C) and thymine (T), and have the chemical properties of a base [30].

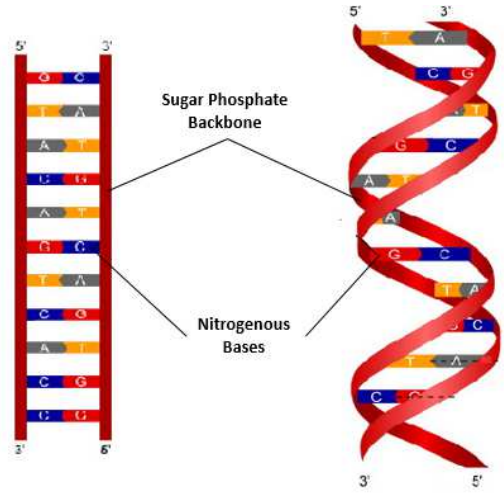

Figure 3: DNA double Helix structure

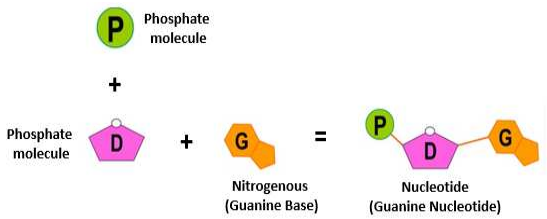

Figure 4: Guanine Nucleotide structure

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.3. Steps in Storing Data on Synthesized DNA

In this section, we present the basic steps to store and retrieve digital data to and from DNA storage.

The process of storing digital data in DNA is usually stored in binary form. This means that the data we store is zero and zero. The hard drive data is 1 and 0, but the DNA is made up of four basic components instead of binary files. The four bases are adenine, thymine, cytosine and guanine, and they work in the same way as A, T, C and G. The data can be stored in pairs instead of storing 0 or 1 separately [10].

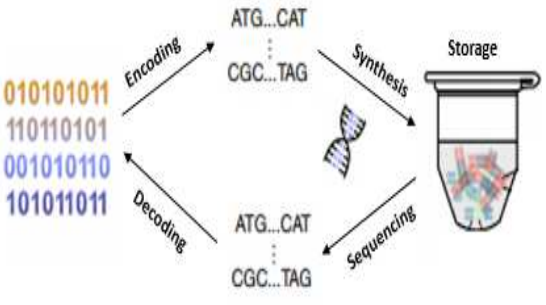

Figure 5: Basic steps of DNA storage.

3.3.1. Encoding data to DNA Sequences

The first step in storing DNA data is coding, which uses computer algorithms to convert binary digital information into DNA sequences [11]. Each base pair can be represented by 2 bits, which gives 4 different possibilities and can be mapped to 16 combinations of DNA base pairs (e.g. 00 → AT, 01 → GC, 10 → TA, and 11 → CG) [7]. One byte (or 8 bits) can represent 4 DNA base pairs. DNA-encoded data can be used for encryption or long-term storage. The number of bits per base pair is called the encoding density. Due to biochemical limitations and internal indexing overhead (i.e. offset and length), most existing work [12] has coding densities less than 2 [13]. Depending on the purpose, DNA can be integrated into a non-coding DNA (ncDNA) or a DNA-coding protein (pc-DNA), or synthetic DNA.

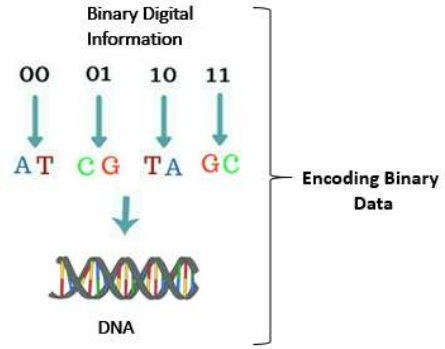

Figure 6: Encoding binary data into DNA sequences

3.3.2. Decoding DNA Data

Corresponding to coding, decoding is the final step in storing DNA using another computer algorithm that has the opposite function to coding to convert DNA sequences into original binary digital data.

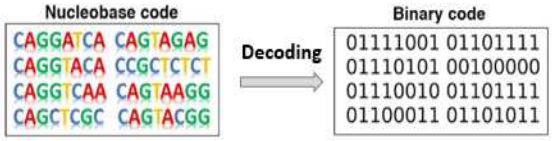

Figure 7: Decoding DNA data

3.3.3. DNA synthesis (writing)

After coding, the generated DNA sequence must be written into the DNA molecule, usually strands of 100 to 200 NT of strand length are synthesized. Several DNA strand synthesis techniques for information storage have been developed, which can chemically synthesize any single-stranded DNA sequence, one nucleotide by a nucleotide [13]. The code that accompanies this protocol accepts a text file read as lines or a Fasta file output by the sequencer, and decodes the information into a stored digital file.

3.3.4. DNA sequencing (reading)

We assume that the data is synthetic and stored in a single pool/isolated DNA tube, which means that millions of different DNA strands are mixed into a single tube. To read the target DNA strand (a strand of DNA synthesized using the same pair of primers), the first step is to extract the droplets from the DNA tube to amplify the target DNA strand by PCR. During the PCR (Polymerase Chain Reaction), we must enter a specific pair of primers and use the pair of primers to copy the DNA strand. In general, the PCR process requires several cycles to accumulate enough target DNA strands. After that, the DNA strand samples are sent to the sequencer for sequencing.

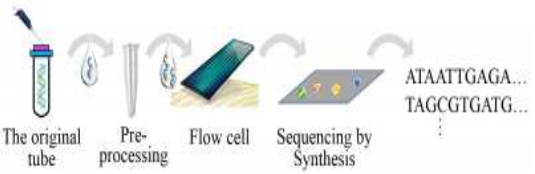

Figure 8: Detailed sequencing process.

These sequenced DNA strands are then decoded into raw numerical data. When decoding, error correction codes (ECCs) are essential to ensure the integrity of the data, as the sequencing process is also prone to errors. Finally, the internal index of the DNA strand will help identify the target DNA strand. Generally, in ECC, the higher the redundancy, the greater the tolerance to failures. In the study of [13] there are two methods to increase redundancy for error correction purposes. One involves adding ECC to the DNA strand and the other storing redundant data on the DNA strand such as a RAID (Redundant Array of Independent Disks) system. Both have the same goal of recovering DNA strand errors.

3.3.5. Storage capacity of the synthesized DNA

In search of a more suitable storage system, various researchers are interested in chemistry and biological molecules. They showed that theoretically, every cubic millimeter of DNA can store 50 billion bytes (10^{18}) of data. The number and scale of this number exceed human understanding [26].

The entire data center created by IBM in 2011 has a data storage capacity of about 100 petabytes (PB). However, DNA as a data storage medium and because of its high density can store a large amount of data in a small size. In theory, each gram of DNA can store about 200 Po of data, almost double the complete data from the IBM data center set. In other words, all the information stored in the world can be stored in a few kilograms of DNA, which is equivalent to a box compared to the requirement of millions of large data storage centers for traditional media [16].

3.3.6. Preservation

The preservation of DNA has aroused great interest among scientists in various fields of research, from ancient biological remains to the field of information [31]. In light of the different DNA conservation requirements (e.g. storage time, storage conditions), scientists have proposed different methods to maintain a high degree of integrity, accuracy, reproducibility and reliability of DNA storage sequences. Therefore, methods to extend the life of DNA are needed. Strategies for the long-term preservation of DNA data can be divided into three categories:

-Dry state (DNA stored in solutions is subject to degradation)
-robust chemical conservation
-Conservation in vivo.

1) Dry state preservation

A variety of solutions and techniques can be used to preserve DNA under optimal conditions. These technologies can retain DNA for a long time, but they seem expensive and cumbersome, so they are not suitable for the long-term preservation of DNA data [32]. On the contrary, dry storage at room temperature is currently an interesting option. Use plates such as commercial GenPlates and polyvinyl alcohol (PVA) plates to store DNA in the dry state and at room temperature. GenPlates are proprietary format plates for storing and transporting biological samples at room temperature with a high density at 384 wells that contain 6mm discs of FTA™ paper molded into a hemispherical shape. FTA is a microporous cellulose filter paper that is chemically treated to inactivate bacteria and viruses or reduce oxidation or withstand high temperatures [33]. These plates enable physical storage of large amounts of information sharing the same addressing scheme through spatial isolation between containers.

Figure 9: A six-region GenPlate for storing DNA

2) Robust chemical conservation

The robust conservation technique is similar to the natural method of preserving DNA sequences and its potential has been confirmed by the fossil. The principle of the robust chemical preservation strategy is the encapsulation of DNA with silica. In particular, researchers propose that DNA can be stored in any form by applying encapsulated DNA to 3D printing technology, but for silica, DNA must be encapsulated in magnetic nanoparticles to achieve a high storage capacity.

3) Conservation in vivo

Compared to the two methods mentioned above, the preservation of DNA assembled in vitro in cells has another advantage that DNA assembled in a plasmid can be naturally amplified during cell replication. Therefore, as long as the cells are alive, the DNA can be preserved [32].

4. THE ADVANTAGES AND CHALLENGES OF DNA DATA STORAGE MEDIUM

Innovations in the digital world are being achieved at a high rate with emerging technological trends. Information that is present in digital form must be stored for a long time with a high density [14] Scientists and researchers have tried to develop a different technology to solve low storage problems and therefore, synthetic DNA is the new technology for data storage [15].

The storage of genetic data has gained much more attention for storage because of its versatility, its long-running data storage capacity and its ability to keep data intact and unchanged. The data-coded DNA medium is capable of long-term storage due to high durability. DNA can last for thousands of years in cold, dry and dark places. Even in a worse environment, the half-life of DNA can go up to a hundred years [16].

DNA media can also protect data to a greater extent than traditional digital storage media. Although the amount of new data is increasing exponentially, most of it is archived for long-term preservation. This cold data will not be retrieved immediately or used frequently. Therefore, storing them on DNA media is simple, convenient and free [16]. Another advantage is that DNA is highly preserved. Natural DNA can replicate accurately with great efficiency and always with the basic principle of pairing (A with T, C with G). In this way, DNA support can strongly maintain data fidelity for a long time.

Preservation of DNA molecules away from undesirable factors (for example, by encapsulating into silica beads) can provide long-term stability for data retention. Appropriate environmental conditions and additives must be taken into account to maintain a longer retention time until the theoretical limit is reached. In addition, encapsulation retention avoids direct access to data files. For example, DNA encapsulated in silica cannot be directly amplified and recovered by PCR unless it is released from the beads. Therefore, the optimized retention method must balance long-term stability and data accessibility [28].

However, at present, the following disadvantages prevent widespread DNA implementation:

The main drawbacks of synthetic DNA storage are the high cost and long access time. Although the current cost of storing data in the form of DNA is high, it can be expected that with the development of the technology, this cost will decrease rapidly. Currently, it takes hours to capture and retrieve DNA data, which is not practical for most real-time applications. Scientists are working hard to reduce this visitation time [17].

The second disadvantage of DNA storage is that reading and writing DNA are fairly slow processes compared to other data storage forms, and may not be appropriate when information is needed quickly. DNA storage can work best for archiving purposes. Storing genetic data is also more expensive than storing it on hard drives. The more you want your DNA to exist, the higher the cost.

5. SYNTHETIC DNA VS TRADITIONAL STORAGE MEDIAS

As mentioned in the figure, the density of DNA information storage is several orders of magnitude higher than any other known storage technology, In the initial work, the apparent density of DNA molecules was approximated to pure water density, resulting in an information density of 4,606 × 1017 bytes/mm. In comparison, the density of information storage of conventional media, such as flash drives, optical strips, and hard drives, is in the order of 109 bytes/mm. For decades, tens of kilograms of DNA can meet the world's storage needs [23].

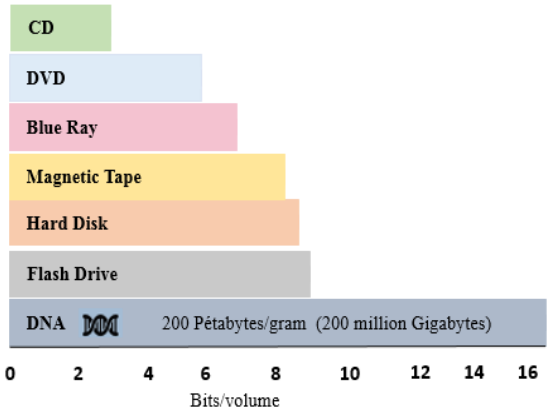

Figure 10: Information storage density of DNA versus traditional media

Traditional data storage media, including magnetic (tape, floppy disk, hard disk), optical (CD, DVD, Blu-ray), and flash memory (memory card, SSD drive), degrade over time, and their long-term use is restricted. DNA molecules are one of the most powerful biomolecules found in nature and have a longer shelf life without data mitigation. Due to its complementarity and ability to self-assemble during the tertiary structure formation, the DNA strands are arbitrarily folded into a polygonal digital grid. In addition, DNA is non-volatile and uses little energy to function in living cells [27].

6. CONVERTING DATA INTO DNA SEQUENCES

In this section we present an experiment that aims to convert a stored file in memory into a DNA sequence using the python language and a Hamming code for mutation resistance and data corruption.

The program can read any ASCII text file, and encode it in a DNA sequence that is aligned with 13 bits of data in a (12.8) Hamming Code. That is, 5 bits of parity for every 8 bits of original data. The output is in FASTA format. The same program can be used to decode a sequence of coded DNA, with a file encoded in FASTA format, to return to its original content.

The program makes use of the Python multiprocessing library:

✓ Bitstring

Bitstring is a pure Python module designed to help make creating and analysing binary data as simple and natural as possible. They can be built from whole (big endian and little endian), hexadecimal, octal, binary, chain or file. They can be sliced, connected, inverted, inserted, crushed, etc. Has a simple function or slice symbol. You can also read, search for, replace and browse them from files, similar to files or streams [18].

✓ Bitarray

This module provides a type of object that efficiently represents an array of Booleans. Bitarrays are sequence types and behave very similarly to regular lists. Eight bits are represented by one byte in a contiguous block of memory. The user can choose between two representations: little-endian and big-endian [19].

✓ Binascii

The binascii module contains methods to convert between binary and various binary representations encoded in ASCII. Normally, you won't use these functions directly, but you will use wrapper modules like base64, or binhex instead. The binascii module contains faster low-level functions written in C which are used by high-level modules.

✓ Multiprocessing

multiprocessing is a package that supports build processes using an API similar to the threading module. The package offers both local and remote concurrency, effectively bypassing the global interpreter lock by using subprocesses instead of threads. For this reason, the multiprocessor module allows the programmer to fully exploit several processors on a given machine. It works on both Unix and Windows [20].

6.1. Hamming Code

In the late 1940s, Richard Hamming realized that the further development of computers required greater reliability, particularly the ability to detect and correct errors. At that time, parity was used to detect errors, but this was not possible. It has created Hamming codes, which ensures that no information transmitted/stored is corrupted or affected by errors on a single bit. The input is an error-free information of k-bits long that is sent to encode it. Encode it and then apply the theoretical hamming, calculate the parity bits, and attach them to the information data received, to form a n-bit code word. Processed information that contains additional parity bits is now ready for storage [21].

Each check is now a sum only on the bits in the selected positions. In the simplest case,

message words of length (2 - k - 1) bits, where k is any whole, must be sent with control k bits, so that each code word (message bits plus bits of control) contains (2 - 1) bits. The positions of the code word are numbered from left to right. The first bit of control is in position 1, and is a parity check on positions that have a 1 as the least significant bit of their binary representations (i.e. positions 1, 3, 5, 7, ...). The second bit of control is in position 2, and is a parity control on positions that have a 1 as the second bit the least significant of their binary representations (i.e. positions 2, 3, 6, 7, ...), and soon. If no parity check fails, the code word is supposed to be correct. In addition, if a bit of the code word is misrepresented, the error is in the location whose binary representation is equal to the failed parity control model. Hamming codes are still widely used in computer science, puzzles and turbo codes [22].

6.2. Files in FASTA Format

The FASTA (or Pearson) format is a text file format used to store biological sequences of a nucleic or protein nature. These sequences are represented by a series of letters encoding nucleic acids or amino acids according to the IUPAC nomenclature. Each sequence can be preceded by a name and comments. This format originally came from the FASTA suite of programs but, because of its widespread use, has become a de facto standard in bioinformatics.

The simplicity of the FASTA format makes manipulating and reading (or syntactic analysis) of sequences easy by using word processing tools and script languages such as Python, R, Ruby or Perl.

The description line is distinguished from the sequence data by a greater-than (">") symbol in the first column (X representing nucleic or amino acids):

>Identifying Comment

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXX

6.3. Experience and Result

In this section, we present the results of the experiment which aims at converting text files into DNA series and decoding files in FASTA form into text files. The program contains two operations, encoding and decoding.

- **Specify encoding or decoding**

✓ **--encode / -e:** Convert a text file to DNA, using Hamming code with (12.8), the output file is in FASTA format.
✓ **--decode /-d:** Converts a 12-bit aligned FASTA DNA file to its original format, the input file is the same file format as the output of --encode and the output file is the original text file.
✓ **--corrupt / -c:** Corrupt a FASTA file at a desired rate to simulate mutation over time.

- **Specify input file**

The first argument is the input text file. For the encoding operation (--encode), this is a normal text file. For the decoding operation (--decode) this is a file in the Hamming encoded FASTA format.

- **Encoding**

Big test file in txt format

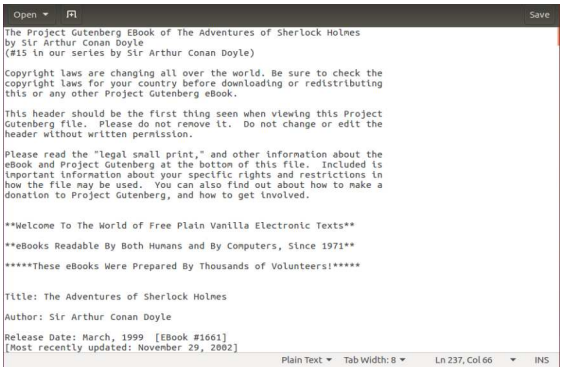

Figure 11: The text file to convert

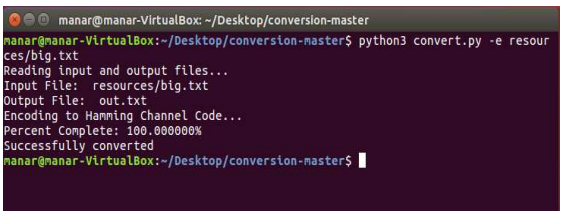

Figure 12: Converting the text file to FASTA file

Journal of Theoretical and Applied Information Technology
15th August 2021. Vol.99. No 15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 13: The file in FASTA format after the conversion

- **Decoding**

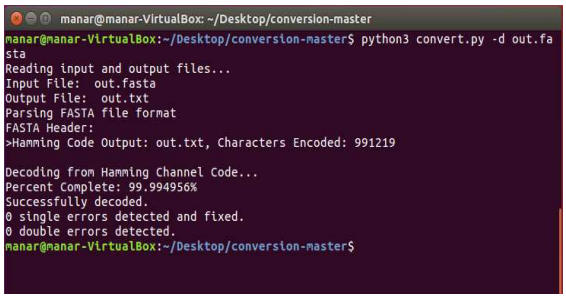

Figure 14: Converting the FASTA file to the text file

7. CONCLUSION

Today, most of today's digital data is mainly stored on magnetic and optical media. In the explosive age of digital data, digital data is generated every day and growing exponentially. These traditional media cannot meet the urgent need for massive digital data storage. Synthetic DNA has the benefits of high density, high replication efficiency, long-term durability, and long-term stability, and is an excellent choice to become an urgent long-term storage medium for valuable digital data. Deoxyribonucleic acid (DNA) is expected to become a potential new data storage medium. For the new DNA data storage, files or any readable data will be converted into binary files and then encoded into DNA consisting of sequences of adenine (A), cytosine (C), guanine (G), and thymine (T).

DNA molecules are certainly robust enough to remain intact for at least several thousand years, despite being stored dry, cold and dark, while no high data migration costs occur as with current mass technologies. Other strong arguments for using DNA molecules to archive numerical data are the possible storage density of around 5 PB per gram, the simplicity of copying DNA molecules at any time, and that there is in fact no risk of format obsolescence. In this article, we discussed the technology of DNA storage, namely its biologicals compositions, the steps of encoding and decoding information in synthetic DNAs, its large storage capacity compared to traditional technologies, the advantages, and challenges of this model of biological storage, concluding with a test which allows encoding textual data in series and vice versa in order to store and preserve them in synthesized DNAs.

Discussion Topics

We have arrived at the end of this work, but we cannot conclude without giving a glimpse of the future and new research ideas. Researchers in the fields of chemistry, biochemistry and computer science at the University of Ghent in Belgium have developed a revolutionary new technology for data storage, inspired by the idea of genetic storage of information. Powdered data storage could become a viable alternative and it will be a more environmentally friendly technology.

Acknowledgements

This work was supported in part by the National Center for Scientific and Technological Research (CNRST) and this within the program of the research grants initiated by the Ministry of National Education, Higher Education, Management Training and Scientific Research.

REFERENCES:

[1] https://www.datacenterknowledge.com/archives/2013/01/18/facebook-builds-new-data-centers-for-cold-storage,Facebook Builds Exabyte Data Centers for Cold Storage | Data Center Knowledge ,18/03/202.
[2] Davis Joe, “Microvenus”, Art Journal,Vol. 55, No. 1, March 1, 1996, pp. 70-74.
[3] Azat Akhmetov, Andrew D. Ellington and Edward M.Marcotte, “A highly parallel strategy for storage of digital information in living cells, BMC Biotechnology”, vol. 18, No. 1, October 17, 2018, pp. 64.
[4] J. D.Watson, F. H.Crick, “Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid”, Nature, Vol. 248, No. 5451, April 26, 1974, pp. 765.

[5] C. T.Clelland, V., Risca, C.Bancroft, "Hiding messages in DNA microdots", Nature, Vol. 399, No. 6736, June 10, 1999, pp. 533-534, Doi. 10.1038/21092

[6] C.Bancroft, T.Bowler, B.Bloom and C. T.Clelland, "Long-term storage of information in DNA", Science (New York, N.Y.), Vol. 293, No. 5536, September 07, 2001, Doi. 10.1126/science.293.5536.1763c.

[7] Dixita Limbachiya, Manish K. Gupta, "Natural Data Storage: A Review on sending Information from now to then via Nature", arXiv:1505.04890 [cs, math], may 19,2015.

[8] Yesenia Cevallos, Luis Tello-Oquendo, Deysi Inca, Nicolay Samaniego, Ivone Santillán, Amin Zadeh Shirazi and Guillermo A. Gomez, "On the efficient digital code representation in DNA-based data storage", Proceedings of the 7th ACM International Conference on Nanoscale Computing and Communication, Article No. 18, September 23, 2020, New York, NY, USA, pp. 1-7, Doi. 10.1145/3411295.3411314.

[9] R. Deaton, M.Garzon, R. C.Murphy, J. A.Rose,D. R.Franceschetti and S. E.Stevens, "Reliability and Efficiency of a DNA-Based Computation", Physical Review Letters,January 12,1998, Vol. 80, No. 2, pp. 417-420, Doi. 10.1103/PhysRevLett.80.417.

[10] Malithi Chamalka, Hmdb Herath, Dushan Herath and Malithi Madujani,"DNA Digital Data Storage", August 19, 2020.

[11] https://vonguru.fr/2019/03/27/microsoft-devoile-le-futur-du-stockage-de-donnees-informatiques-ladn/, "Microsoft dévoile le futur du stockage de données informatiques : l'AND", 22/12/2020

[12] James Bornholt, Randolph Lopez, Douglas M.Carmean, Luis Ceze, Georg Seelig and Karin Strauss,"A DNA-Based Archival Storage System", Proceedings of the Twenty-First International Conference on Architectural Support for Programming Languages and Operating Systems, March 25, 2016, Atlanta Georgia USA, pp. 637-649, Doi. 10.1145/2872362.2872397.

[13] Bingzhe Li, Nae Young Song, Li Ou and David H. C.Du, "Can We Store the Whole World's Data in {DNA} Storage?", 12th {USENIX} Workshop on Hot Topics in Storage and File Systems (HotStorage 20),2020.

[14] Fatima Akram, Ikram ul Haq, Haider Ali and Aiman Tahir Laghari, "Trends to store digital data in DNA: an overview", Molecular Biology Reports, Vol. 45, No. 5, October 1, 2018, pp. 1479-1490.

[15] Shyna Sharma, Shruti Pathak, Taranjot Singh Kathuria, Tarun Sharma and Malvinder Singh Bali, "Big Data Storage using Synthetic DNA", International Journal of Computer Trends & Technology, Vol. 67, No. 4, April 25, 2019, pp. 128-130, Doi. 10.14445/22312803/IJCTT-V67I4P125.

[16] Lichun Sun, Jun He, Jing Luo and David H Coy, "DNA and the Digital Data Storage", Health Science Journal, Vol. 13, No. 3, 219, pp. 8.

[17] https://medium.com/hackernoon/dna-data-storage-d0f0e93513b, Medium, 15/03/2021

[18] Scott Griffiths, https://github.com/scott-griffiths/bitstring, 18/03/2021.

[19] https://github.com/ilanschnell/bitarray,bitarray: efficient arrays of booleans -- C extension, 18/03/2021.

[20] https://docs.python.org/fr/3/library/multiprocessing.html, multiprocessing — Parallélisme par processus — Documentation Python 3.9.2, 19/03/2021.

[21] Caleb Hillier, V.Balyan, "Error Detection and Correction On-Board Nanosatellites Using Hamming Codes", Journal of Electrical and Computer Engineering, Vol. 2019, February 10, 2019, pp. 1-15, Doi. 10.1155/2019/3905094.

[22] Arash Ahmadpour, A. Ahadpour Shal, "A Novel Formulation of Hamming Code", May 1,2009,Doi.10.1109/ECTICON.2009.5137169.

[23] Yiming Dong, Fajia Sun, Zhi Ping, Qi Ouyang and Long Qian, "DNA storage: research landscape and future prospects", National Science Review, Vol.7, No. 6, June 01, 2020, pp. 1092-1107, Doi. 10.1093/nsr/nwaa007.

[24] Manar Sais, Najat Rafalia and Jaafar Abouchabaka," The Future of Big Data Storage with Synthetic DNA", 36th International Business Information Management Association (IBIMA), ISBN: 978-0-9998551-5-7, November 4-5, 2020, Granada, Spain, p 13612-13618.

[25] Manar Sais, Najat Rafalia and Jaafar Abouchabaka, "Intelligent Management of Data Storage in Big Data: Relational VS NOSQL, HDFS VS REHDFS", 36th International Business Information Management Association (IBIMA), ISBN:

978-0-9998551-5-7, 4-5 November 2020, Granada, Spain, p 11928-11944.

[26] https://www.spiria.com/fr/blogue/megadonnees/adn-prochain-media-de-stockage-pour-le-big-data/, L'ADN, prochain média de stockage pour le "Big Data"?, 01/05/2021

[27] Darshan Panda, Kutubuddin Ali Molla, Mirza Jainul Baig, Alaka Swain, Deeptirekha Behera, Manaswini Dash, "DNA as a digital information storage device: hope or hype?", 3 Biotech, Vol. 8, No. 5, May, 2018, pp.239, Doi. 10.1007/s13205-018-1246-7

[28] Chengtao Xu, Chao Zhao, Biao Ma, Hong Liu, "Uncertainties in synthetic DNA-based data storage", Nucleic Acids Research, Vol. 49, No. 10, June 4, 2021, pp. 5451-5469, Doi. 10.1093/nar/gkab230

[29] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter, "The Structure and Function of DNA", Molecular Biology of the Cell. 4th edition, 2002.

[30] https://courses.lumenlearning.com/wmopen-nmbiology1/chapter/storing-genetic-information/, Storing Genetic Information | Biology for Non-Majors I, 03/07/2021

[31] Xin Tan, Liqin Ge, Tianzhu Zhang, Zuhong Lu, "Preservation of DNA for data storage", Russian Chemical Reviews, Vol. 90, No. 2, March 3, 2021, pp. 280, Doi. 10.1070/RCR4994.

[32] Yaya Hao, Qian Li, Chunhai Fan, Fei Wang, "Data Storage Based on DNA", Small Structures, Vol. 2, No. 2, 2021, pp. 2000046, Doi. 10.1002/sstr.202000046.

[33] Anjali G. Kansagara, Heather E. McMahon, Michael E. Hogan, "Dry-state, room-temperature storage of DNA and RNA", Nature Methods, Vol. 5, No. 9, 2008, Doi. 10.1038/nmeth.f.219.