International Journal of Emerging Technology and Innovative Engineering
Volume 5, Issue 2, February 2019 (ISSN: 2394 – 6598)

A SURVEY ON RECENT TRENDS IN DIGITAL DATA STORAGE ON DNA

S. SathyaBama

Assistant Professor, Sri Krishna College of Technology, Coimbatore

ABSTRACT

DNA (deoxyribonueclicacid) which is said to be the blue print of human life, is recognized as a powerful storage medium for storing digital information. The binary information of the digital data can be mapped to the four building nucleotides adenine, guanine, cytosine, and thymine. A very large data can be stored within a small of part of DNA.The storage is made by the process of encoding the digital information to the nucleotides. The encoded information is synthesized to get a sequence. The sequence obtained is further decoded to retrieve the data.Using the DNA as a storage medium has several advantages such as tremendously reducing the storage space and preserving the data for very long time. At the same time it has a constraint that synthesizing a DNA is a very expensive process. This survey elaborates recent new trends adopted by several scientists of different institutes to store and retrieve data on DNA Strand.

KEYWORDS

DNA Sequencing, Synthesizer, Digital data

INTRODUCTION

Earlier days we used several traditional storage mediums such as papers, files etc,. These traditional media required very large storage space and maintenance cost.

Figure.1 DNA Strand

Later we moved on to digital storage mediums such as zip drives, floppies, and compact disks and so on. A $3^{1/2}$ floppy disk can store 1.44 mega bytes of data, but the smart phone era totally changed this situation by increasing the storage by 44,000 times.[9]-[14]

Now we can store the contents of 150 smart phones that are 10,000 giga bytes of data within an extremely smaller mediumof storage DNA.

In the living organisms the nucleotides are encoded upon the cells for their behaviours.

The same way digital information is encoded andmapped onto the DNA codes for storing the information. The stored information can later retrieveby means of decoding the encoded data by means of some algorithms.

TRADITIONAL STORAGE MECHANISMS

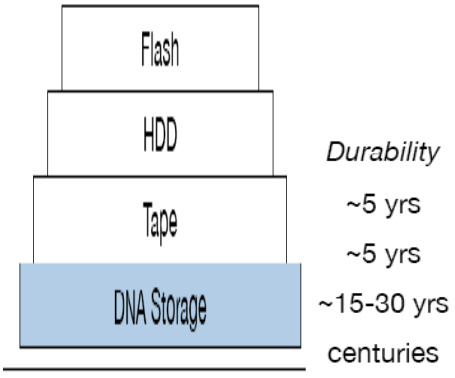

Figure.2 storage mechanisms and their durability

Traditional storage mechanisms include tapes, compact disks, floppies, magnetic tapes.

DNA STRUCTURE VS BINARY DIGITS

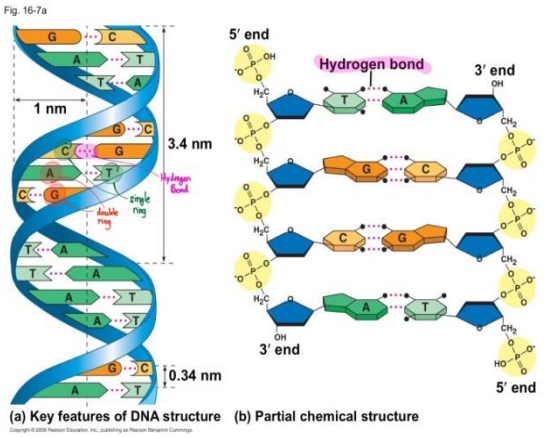

Figure.3 Structure of DNA Strand and its bonding

The complementary partners in a DNA base pair are always made up of one purine and one pyrimidine, never two purines or two pyrimidines. Purines are the nucleotides that have the double-ring structures. Pyrimidines are the nucleotides that have the single-ring structures. An easy way to think of it is that the longer names (pyrimidines) have the smaller rings, and the shorter name (purines) have the bigger rings.

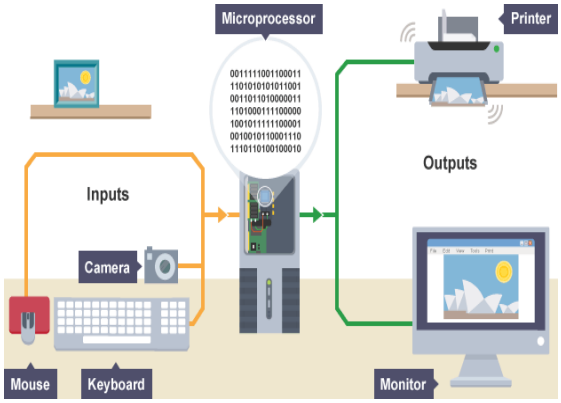

Figure.4 Data in the form of binary digits

Every data that is being fed into the computer system is stored in the form of binary values called 0's and 1's. These binary values will be converted again into its original form by the microprocessor. This mechanism is explained in figure.4.

DNA SYNTHESIS

A single-strand DNA sequencecan be synthesized chemically. The first step of DNA replication is the breaking up of hydrogen bonds between the two anti parallel strands DNA molecule. The separation of bonds between adenine and thymine is quite easy because there exist only two hydrogen bonds whereas the separation of guanine and cytosine is difficult because of its three-hydrogen bond. With the help of the breaking enzyme called helicase the bonds are broken. The strand after the replication process is called replication fork. The replication process takes place followed by the elongation process. The newly formed double stranded DNA molecule contain one old strand and a new one, hence it is said to be a semi conservative process. When the polymerase reaches the end of the strand the process terminates.

MECHANISM OF STORING DIGITAL INFORMATION WITHIN THE DNA MOLECULE

Figure.5 pictorially depicts the storage and retrieval mechanism of data in DNA strand. The input is fed into the DNA Synthesizer and get stored into the DNA storage library. Using the PCR thermo cycler it gets again retrieved by means of DNA Sequencer.

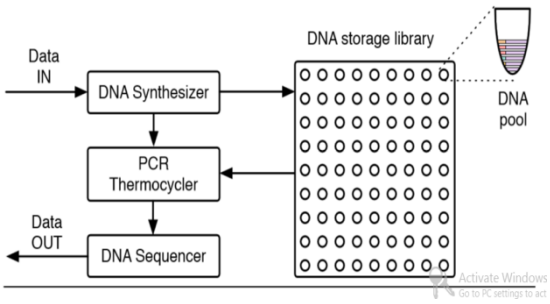

Figure.5 General Method of storing digital data on DNA strand.

DNA SYNTHESIZER

The DNA synthesizer will encode the input data.

DNA POOL

The pool contains a collection of DNA molecule that can be mapped on to the storage library.

DNA SEQUENCER

The DNA sequencer will sequences the data and converts them back to digital data.

ENCODING MECHANISM

The binary data is converted into a ternary data with the help of Huffman encoding [1]. This ternary data is further converted into nucleotide code.

MICROSOFT'S REPORT 2016

Microsoft with its researchers and scientists has reported a big leap forward for DNA data storage, DNA could be a better way to store data for the long term than the magnetic tape companies rely on today[3]. Microsoft has stored 200MB of data and demonstrated that on July 7 2016.

MICROSOFT'S REPORT 2017

March 2, 2017,Columbia University School of Engineering and Applied Science has proposed an algorithm designed for streaming video on a cell phone can unlock DNA's nearly full storage potential by squeezing more information into its four base nucleotides, say researchers[3].They demonstrate that this technology is also extremely reliable.

MICROSOFT AND TWIST BIO- SCIENCE REPORT 2018

Researchers at Microsoft in collaboration with the U.S Biotech Company named Twist biotech have succeeded in storing 400 megabytes of data. This project named "Random Access in Large scale DNA Storage"has overcome the disadvantage of the previous method proposed in 2016 [7]. The researchers and the scientists used durable synthetic DNA strand for data storage. The conventional technique in the previous research, retrieving the stored data in the DNA molecule requires sequencing the entire pool of DNA. This approach encoded about 35 different files of 400 MB into 13 million strands DNA molecules and retrieves the files individually with zero error. Figure 5 depicts the entire phase of their research.

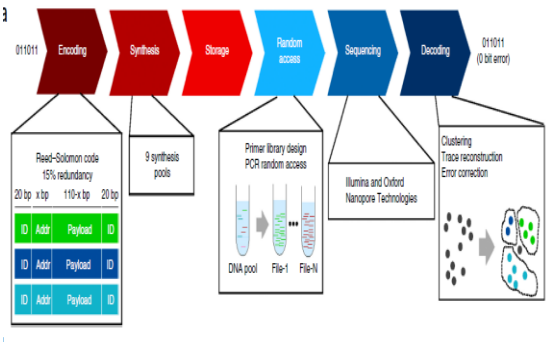

Figure.6 phases of "Random Access in large scale DNA"

NEW TECHNIQUE OF WATERFORD RESEARCHERS 2018

On 20th February 2018 Irish scientists at Waterford institute announced a new technique of storing and recovering data in DNA strand using bacterial molecules. Their survey shows that human in 2025 the proliferation of data will become 160 zettabytes. The researchers encrypt an easy message - during this case "Hello World" - into the plasmids and store them during a strain of the E. coli bacterium referred to as Novablue that's cornered during a specific location that becomes the archive storage location [8].

Another variety of E. coli bacterium, HB101, that is mobile, is then discharged and travels to the Novablue. Once it meets it, the plasmids containing the information are transferred from the Novablue to the HB101 through an association method called conjugation.

The HB101 then swims to a tool capable of extracting the plasmids and reading the info they store. The movement of the bacterium and therefore the conjugation is controlled and created potential by the utilization and placement of 2 completely different antibiotics, antibacterial and antibiotic, among the archive storage and retrieval space.

Novablue is immune to Achromycin, whereas HB101 is in a position to resist antibiotic.

In order to complete its swim across the archive enclosure, the HB101 should thus initial conjugate with the Novablue so as to select up its resistance to Achromycin.

In the method of doing this, the HB101 picks up the info storing plasmids from the Novablue. There are a unit variety of challenges to be overcome before this proof of principle may well be used habitually and firmly.

CONCLUSION

In this article we discussed various methods of storing digital data onto DNA. The recent achievements of Microsoft have also been discussed. Even though DNA storage has greater advantages of storing data, the cost remains a barrier. Scientists are working on several projects for minimizing the cost of artificial DNA synthesis.

In the future DNA storage will surely make a great impact on big data storage.

REFERENCES

[1] S. Shrivastava and R. Badlani, “Data storage in DNA,” International Journal of Electrical Energy, vol. 2, no. 2, pp. 119–124, 2014. View at Publisher · View at Google Scholar

[2] A DNA-Based Archival Storage System James Bornholm, Randolph Lopez, Douglas M. Carmean, Luis Ceze, Georg Seelig Karin Strauss, University of Washington Microsoft Research

[3] ScienceDaily web portal.

[4]Forward Error Correction for DNA Data Storage. Meinolf Blawat, Klaus Gaedke, Ingo Hütter, Xiao-Ming Chen, Brian Turczyk, Samuel Inverso, Benjamin W. Pruitt, George M. Church Technicolor Research & Innovation, Hannover, Germany Harvard Medical School, Wyss-Institute, Boston, US. Elsevier

[5]Research Journal of Pharmaceutical, Biological and Chemical Sciences Genetic Code as Binary.UdayBhaskar, AdeeshNagpal, Himanshu Paul, and M Yamuna VIT University, Vellore, Tamilnadu, India.

[6]www.rte.ie/news/ireland/2018/0219/941956-dna-data/

[7]Lee Organick1, Siena Dumas Ang2,et.alRandom access in large-scale DNA data storage”2018 Nature America, Inc., part of Springer Nature.

[8]www.rte.ie/news/ireland/2018/0219/941956-dna-data/

[9] Rajkumar, Dr M. Newlin, M. Sruthi, and Dr V. Venkatesa Kumar. "IoT based smart system for controlling Co2 emission." Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol 2.2 (2017): 284.

[10] Devi, C. Akalya, D. KarthikaRenuka, and S. Soundarya. "A Survey Based on Human Emotion Identification Using Machine Learning and Deep Learning." Journal of Computational and Theoretical Nanoscience 15.5 (2018): 1662-1665.

[11] Raja, J. Beschi, S. Chenthur Pandian, and J. Pamina. "Certificate revocation mechanism in mobile ADHOC grid architecture." Int. J. Comput. Sci. Trends Technol 5 (2017): 125-130.

[12] Raja, J. Beschi, and V. Vetriselvi. "Mobile Ad Hoc Grid Architecture Based On Mobility of Nodes." International Journal of Innovative Research in Computer and Communication Engineering 2 (2014): 49-55.

[13] Lydia, E. Laxmi, et al. "Correlating NoSQL Databases With a Relational Database: Performance and Space." International Journal of Pure and Applied Mathematics 118.7: 235-244.

[14] Raja, J. Beschi, and K. Vivek Rabinson. "Iaas for Private and Public Cloud using Openstack." International Journal of Engineering Research 5.04 (2016).