Surge of Big Data and Emergence of DNA Storage System

Ilhaan Rahim
Department of Computer Engineering
GH Raisoni Institute of Engineering and Technology, Nagpur, Maharashtra, India

Abstract— There is no denying of the fact that data has become the most valuable asset of the 21stC and it's safe to say that "Those who knows the data owns the future", according to the most up to date estimation the total amount of data created, captured, copied and consumed globally is close to around 64.23 zettabytes in 2020 [1]. This projection had estimated to grow to more than 180 zettabytes in 2025. Now we ask ourselves, will we be able to suffice the growing unprecedented need of data storage with the technologies we have right now? The answer, is "No". Silicon has serious limitations and limited data storage which can't keep up with the growing need of data, data storage density and capacity of silicon chips is also limited. The solution to this inevitable problem lies within us. The DNA itself is the greatest storage saving device on earth because after decades, centuries of research and science we have been able to develop SD cards, HDD's and SSD's, compare to DNA they are nothing. As the Harvard paper points out, the theoretical maximum data storage of DNA is 455 Exabyte's per gram [2]. Less then 150 grams of DNA can store entire data present on this planet. This paper aims to highlight the extension and limitations of DNA digital data storage system and emphasizes the on going research in this field.

Keywords: Big data, DNA storage, Digital data storage

I. INTRODUCTION

Deoxyribonucleic acid or DNA is biological information storage system of all the living beings that exist on earth. DNA is usually discussed when talking about biological systems but DNA is an appealing option for storing digital information's. DNA stores information of biological organisms in the form of A's, T's, G's, and C's. The human genome has 6.4 billion letters precisely sequenced in a chemical alphabet similar to the sequential code of 0's and 1's as found in a computer. DNA at it's theoretical maximum capacity (455 Exabyte's per gram) can encode about 2 bits per nucleotide. An entire data centre built by Facebook in 2012 has about 1 Exabyte of cold storage capacity. However, due to high density, DNA acting as a data-storage medium can store a large amount of data at an astonishingly small size.

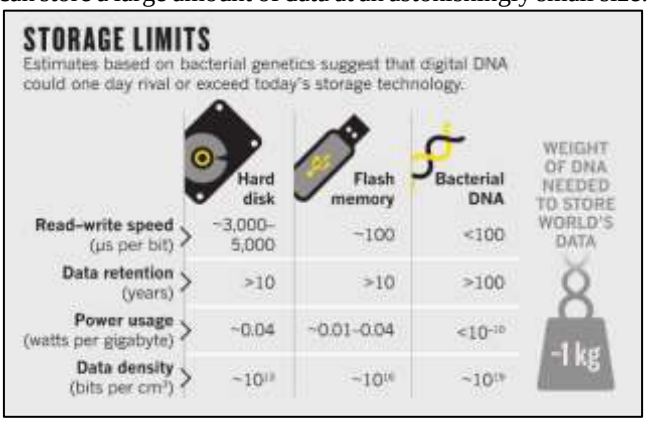

Fig. 1: DNA Data Storage Limits [4]

A single gram of DNA at it's theoretical maximum limit can store about 455 Exabyte's of data, almost 450 times more than that of the entire Facebook data centre. In other words, all the information recorded all over the world can be stored in several kilograms of DNAs, or at the back of your car compared with the requirement of millions of large data storage centres for traditional media [3]. With DNA, a whole human genome can fit into a cell that is invisible to the naked eye. The sheer density of information storage, DNA could be orders of magnitude beyond silicon — perfect for long-term archiving.

This is one of the reason scientist were able retrieve DNA from the fossils of animals who were extinct millions of year age. Because information retention of DNA is unconceivably high, once we can get data written on DNA, you can stick it in a cave and forget about it until you want to read it." The Microsoft corporation has started its research and experiments on DNA data storage too, in fact it had the most successful start in this field of technology.

Demand for data storage has been growing exponentially, but the capacity of existing storage media is not keeping up. Most of the world's data today is stored in colossal data servers. Despite improvements in HDDs and SDDs, storing a zettabyte of data would still take many millions of units, and use significant physical space. If we are to fulfil the growing demand of data storage, we need to seek significant advances in storage density and durability. Using DNA to store world data is an attractive possibility because it is extremely dense (up to about 1 Exabyte per cubic millimeter) and durable [5][6].

II. HOW DNA STORAGE WORKS

DNA data storage is the process of encoding binary data into synthetic, man-made strands or sequences of DNA. DNA storage is ultra-compact due to which data can be stored safely for thousands of years in a cool, dry place. To store a binary digital data in DNA, the bits (binary digits) are converted from 1s and 0s into the letters A's, C's, G's, and T's. DNA molecules are long strands or sequences, which are made-up of nucleotides called Adenine, Cytosine, Thymine, and Guanine which represents letter A, C, G, T.

The physical storage medium is a synthesized chain of DNA containing the A's, C's, G's, and T's in a sequence corresponding to the order of the bits in the digital file. Oligonucleotide synthesis machines are made to store and upload information digital data in DNA. To recover the data, the chain of DNA is sequenced and the order of A's, C's, G's, and T's are retrieved back using DNA Sequencing Machines (Figure 2).

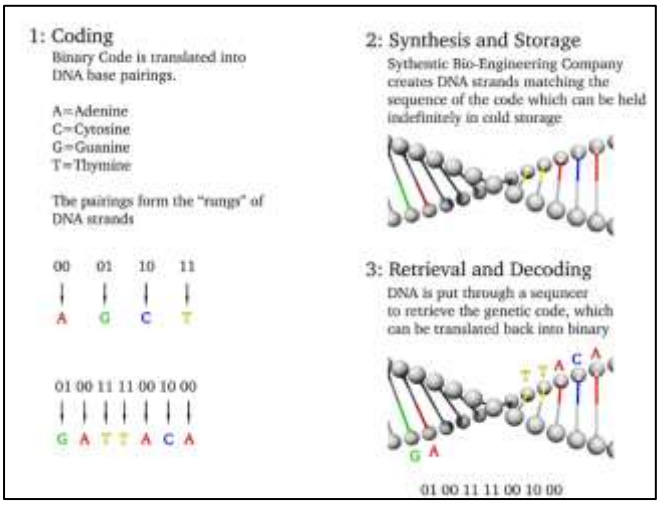

Fig. 2: Process of encoding and decoding digital data from and into DNA [7]

The entire process of encoding and decoding of DNA storage can be divided into following representation or steps:

1) Encoding of digital data
2) Synthesis of DNA
3) Storage of digital data into DNA
4) Retrieval of digital data from DNA
5) Sequencing of digital data
6) Decoding of digital data

Details of these steps are not discussed here since it's not the purpose of this paper. The process of encoding and decoding of data in DNA is being enhanced on every steps. But, currently this process is relativity tedious and expensive as compare to HDD's and SDD's which we will discuss on next section.

III. Drawbacks and Challenges

Every technology which has the potential to revolutionize human society has some challenges and drawbacks, we will now divert our attention to review some of the major drawbacks and challenges faced with DNA data storage around the globe.

Many significant changes are yet to be made before the large scale commercial use of this technology. However, it will not be easy to achieve this potential in order to develop user friendly DNA data storage systems. The entire process of encoding and decoding of information in DNA is far more time consuming than conventional storage technology. Although the durability of DNA is staggering but DNA is also prone to mutation under extreme conditions, the chances of data alteration cannot be ignored. Compare to the conventional data storage system like HDD's and SDD's, the process of retrieving, rewriting and erasing of data from DNA based system is still on its very outset.

The other significant challenges with DNA data storage are cost and efficiency. It would cost $1 trillion to write one petabyte of data into DNA which compare to conventional system is very expensive [8]. Speaking of efficiency, imagine you want a picture stored in your DNA storage systems you are probably going to have a piles of DNA, which is gazillions of images, movies, files and other stuff, finding one picture is like trying to find a needle in a desert. And we don't have any current technology to mitigate this process. Many significant changes are to be made yet trying to improve and develop the system rapidly. The main target of this technology now is to make it faster and cheaper.

IV. Conclusion

The DNA Data Storage technology is a revolutionary idea which reinvents the way we store and retrieve data from devices. It can replace the huge data centers which are very expensive and consumptions of resources, if we can figure out the way to address the drawbacks and challenges of this technology with efficiency and cost, the entire on available on this planet can be stored in your fingertips. Storing data in DNA is a natural way of storing data which saves us tones of resources and enable us to store our memories for thousands of years, sharing knowledge to all the generations ahead. For a civilization that irresistibly stores and creates huge amounts of data, our temporarily digital storage technologies won't satisfy the exponential need of humanity. DNA data storage systems are promising technology which can bridge the gap between the growing demand of storage and big data of future generations [9].

References

[1] Arne Hoist, "Amount of data created, consumed and stored 2010-2025," *statista.com,* June 2021.
[2] George M. Church, Yuan Gao and Sriram Kosuri, "Next-Generation Digital Information Storage in DNA," *science (Newyork),* August 2012. http://arep.med.harvard.edu/pdf/Church_Science_12.pdf .
[3] Sun L, He J, Luo J, Coy DH (2019) "DNA and the Digital Data Storage." *Health Sci J Vol.13.No.3:659. DOI: 10.36648/1791-809X.1000659,* June 2019.
[4] Andy Extance, "How DNA could store all the world's data." *Nature* 537, 22–24 (2016). https://doi.org/10.1038/537022a, August 2016
[5] Antonio Regalado, "Microsoft Has a Plan to Add DNA Data Storage to Its Cloud." *MIT Technology Review,* May 2017.
[6] Dormehl, L, "When we run out of room for data, scientists want to store it in DNA." *Digital Trends*, Retrievable at: https://www.digitaltrends.com/cool-tech/dna-data-catalog-startup/, July 2018.
[7] Cornish, C, "How DNA could store all the world's data in a semitrailer." *Financial Times*. Retrievable at: https://www.ft.com/content/45ea22b0cec2-11e7-947e-f1ea5435bcc7, February 2018.
[8] Anne Trafton, "Could all your digital photos be stored as DNA?" *MIT News Office*, June 2021.
[9] Twist Bioscience, "Could DNA Supercharge the Digital Revolution." *Writing the future*, December 2017

