# DNA- THE NEXT GENERATION DATA STORAGE

Shivangi Batra[1], Anshika Sangwan[2], Indu Khatri[3]
[1,2]Students of Information Technology, [3]Guide
Bhagwan Mahaveer College of Engineering and Management, Sonepat;

***Abstract: Every day we deal with data in many forms and try to organize it in efficient manner. There is an obvious increase in data with advancement of technology. This increase, forces the researchers, to find sustainable solution to data storage. Utilizing the gift of nature (DNA) to store data is the next big investment; DNA fountain, a brief of which will be provided in this research piece.***
***Keywords: DNA fountain "Science and everyday life cannot and should not be separated. - Rosalind Franklin"***

## I. INTRODUCTION

Engineering is the medium through which we can implement and program our imagination into reality. Nature has always been an inspiration towards progress. Currently, we are capable enough to talk about and work, in an extremely sensitive arena; the intersection of the two very diverse fields of engineering and biology which comprises of evolutionary ideas of solving the mysteries of the two disciplines. Biology is an umbrella field with computational and synthetic biology being one of the ribs and stretcher of it.
With computational biology viewing the field of biology with an analytical, mathematical and statistical angle and thereby applying similar aspects to solve different biological problems in concert with computer science concepts. And Synthetic biology deals with engineering basis (standardization, abstraction and modularization and reusability) applied to genetic engineering.
DEOXYRIBONUCLEIC ACID(DNA) acts as our information carrier and transmitter. For engineers dna can be thought of as a programming language where we have nucleotide base pairs (building blocks of DNA) the (Adenine)As, (Thymine)Ts, (Guanine)Gs, and (Cytosine)Cs instead of 0s and 1s(with 0s and 1s being machine understandable and A,T,G,C being understandable by human body) [It is important to note that, Adenine bonds or pairs only with Thymine while Guanine bonds only with Cytosine.]

## II. DATA STORAGE

From the times of Watson and Crick to Bill Inmon, data, either as traditional data or big data has been an area of concern. We have data as vast as its meaning and it is increasing exponentially. From data collection, to recordings and maintaining it thereafter has always been a complex task.
[1]Data storage uses finite resources. Any device that uses flash memory stores data as electrical potential within minute transistors that are etched into silicon wafers. While the number of transistors per unit area of silicon has approximately doubled every two years since their invention, transistors have limits to their minimum size and maximum packing density. If projections hold true, our data boom couldsurpass the data space available in our supply of microchip-grade silicon.[1] Ever since we have been handling data, or particularly storing data we have used floppy disks (with a maximum capacity of 1.44 megabytes), and continue to use flash drives and hard drives with capacity extending to some Terabytes). All these devices today along with online platforms available for storing data undoubtedly act as a strong backbone for data storage, data retrieval and data maintenance.

## III. THE INTERSECTION- DATA and DNA

With advancement of technology we have information and there is a need of robust architecture for storing all the data, along with tackling the maintenance, retrieval, security and related issues. And now we entered the time where humans are capable of storing some petabytes of data in few grams of DNA, something which nature has been doing since years.
DNA is nature's way of storing data The whole idea behind DNA Fountain revolves around getting digital data a new home to stay in with DNA fountain enabling a powerful storage architecture for the ever-increasing data. It involves storing data in synthetic DNA.

Synthetic dna being a laboratory activity, where a group of synthetic biology methods construct and assemble genes (without using the template- non coding dna strand)
The idea behind storing Digital Data in Synthetic DNA is a big evolutionary step, both in the field of biology and engineering.

[1] Data Density and Data Capacity are two important measures in DNA Data Storage. Where, Data Density deals with the theoretical limits of data storage per gram of DNA (with current methods) and Data Capacity deals with amount of data being successfully encoded or decoded in DNA at one time.[1]

## IV. SUCCESSFUL TRIALS

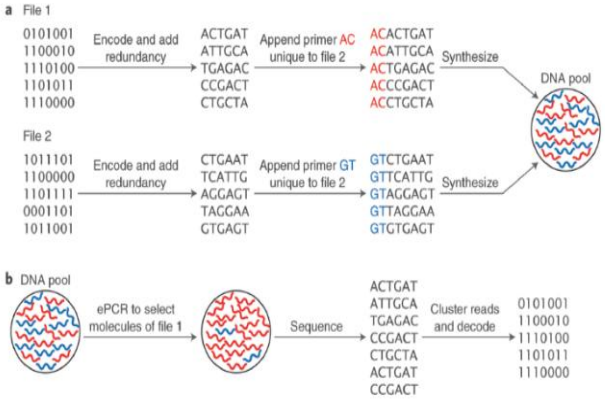
In 1999, the record for DNA storage capacity encoded just a few words[1]

The team from Molecular Information Systems Lab, to date have been able to store 1GB of data in DNA, breaking their previous world record of 200MB [2]

[3]Yaniv Erlich and Dina Zielinski successfully encoded a full operating system, a French film- 'Arrival of Train at La Ciotat', an Amazon gift card, and a lot more in few grams of DNA[5]; which took almost two weeks for synthesis(writing it) and two days to read it(sequencing it)[3]

## V. THE TECHNIQUE




Fig[1]https://www.twistbioscience.com/blog/perspectives/dna-data-storage-setting-data-density-record-dna-fountain

Anything that can be stored in 0s and 1s can be stored in DNA. We need to convert our data, all the 0s and 1s to As, Ts, Gs and Cs and then send to synthesis company. Therefore, we can write it, store it and recover it (sequence it). It involves, random chopping of data and packaging it into droplets. Further translating it into As(00), Ts(11), Gs(10), and Cs(01) (of the same length) which is then sent to (dna) synthesis company. [3]All files are compressed into a single master archive. An algorithm was used to take binary code from that file and split it into short strings of digits. When translating binary into base pair sequences the algorithm is able to drop error sequences. Each bundle being referred as a droplet. To read the data, DNA sequencing technology, special software to reverse encoding process is used.[3] Every time data is retrieved (through sequencing it) we eventually lose the DNA. So, we need to get copies of it for further use. A very powerful technique- polymerase chain reaction is used for DNA amplification.

## VI. DETECTING ERRORS

Coding Theory is the study of properties of codes and their respective fitness for required, specific applications[4]. Encoding a pool of to be designed short DNA strands or oligonucleotides with significant overlap sequence is a possible sollution for error correction[1].

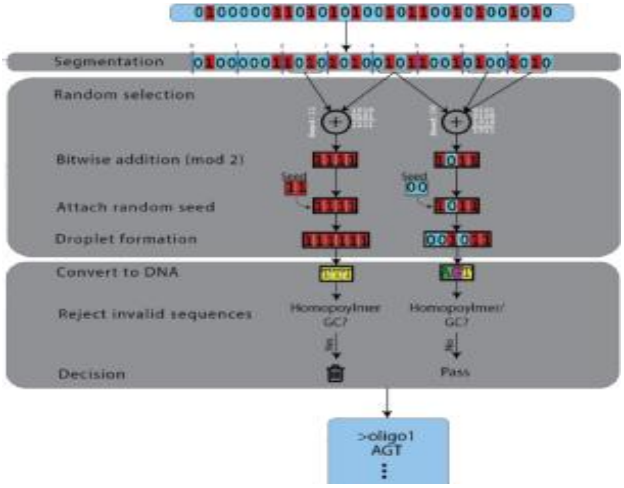



fig.-[5] http://dnafountain.teamerlich.org/

## VII. CURRENT ISSUES

Man has always looked for convenience as a primary factor over any other influencing factors. With advancing technology, a major problem in DNA Data Storage is neither its duplication (creating multiple copies of synthetic dna) nor retrieval of the data but writing it for the first time is both time consuming, currently too expensive and requires a lot of effort. Also, the access time, we get with using current data storing devices or internet storage, one offered with DNA fountain is incomparably large. The aim of upcoming technologies is to make data available anytime, on any device. This is something which is advancing at a very fast rate for dna fountain but has a far future. At present, cost, time efforts and of course humanity are the deciding parameters for the future of this intersecting field of DNA and Data Revolution.

## VIII. CONCLUSION

Initially no technical creativity was thought to reach the level of advancement it has achieved. Be the very first hard drive by IBM, which held about 5MB of data and measured 1.5 sq. meters. But with collective efforts of scientists, researchers and engineers technology continues to advance. Human attempted of creating supercomputers by including the things man can't do and now by this their attempt is to make super-humansthrough this technique. DNA fountain is therefore, one such emerging field which has changed our perspective about efficient data storage. Compactness, durability, consistency and data storage in DNA, acting as a backup of backups are potential reasons which assert the world about it transmogrifying the IT industry.

## REFERENCES

[1] https://www.twistbioscience.com/blog/perspectives/dna-data-storage-setting-data-density-record-dna-fountain
[2] https://news.microsoft.com/innovation-stories/hello-data-dna-storage
[3] https://www.extremetech.com/extreme/245304-researchers-increase-storage-capacity-dna-214-petabytes-per-gram.
[4] https://en.wikipedia.org/wiki/Coding_theory
[5] http://dnafountain.teamerlich.org/