DNA AS FUTURE DATA STORAGE

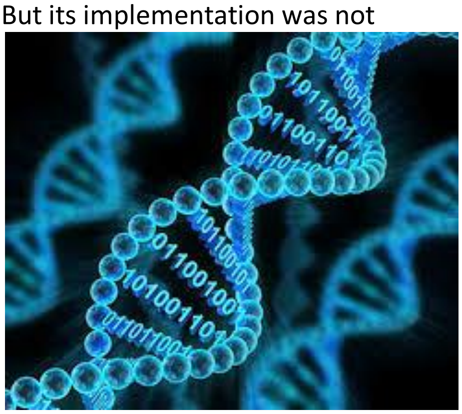
Adish Jain[1,] Pradeep Kumar Gupta[2,]
BCA
College of Computing Sciences & IT,
Teerthanker Mahaveer University, MBD
Asst ProfessorCollege of Computing Sciences & IT,
Teerthanker Mahaveer University, MBD
pradeepg.computers@tmu.ac.in

Abstract-- **Data is Raw information and on analysis it bring us to a conclusion. In present scenario we are creating huge data each and every moment. Present time is also refer as era of Computing and computers are extensively used in all aspects of life. This involvement of Computers in our life we are creating Data in Digital format unlike documental in previous time. Storage of this Digital data is becoming a crucial issue nowadays. We have witnessed a big leap in technology for data storage, there is a big transformation in data storage technology from Magnetic tape to Cloud data storage. In this transformation we have seen Floppy disc, Compact Disc, Digital Versatile Disc and Hard Disc, but more compact data storage discovered more data is created at very fast rate. This need lead us toward exploring more complex and huge data storage technology and biological Data storage became the field of interst in current years.**

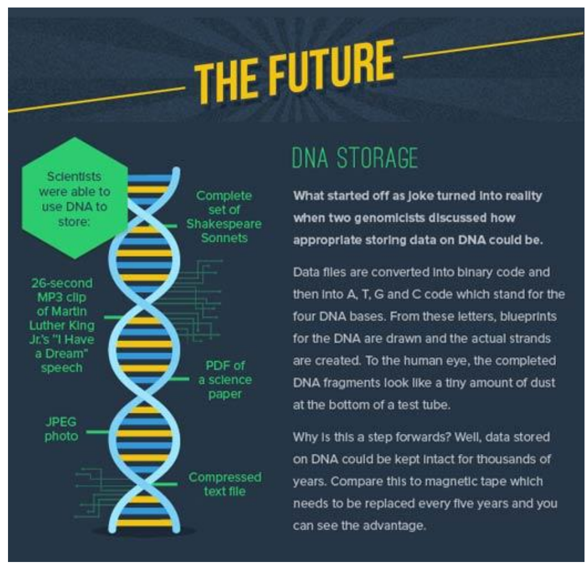
*Keywords--**Information, Digital Data storage, Magnetic tape, Biological Data storage***

1. Introduction

Biological data storage is not a new term it was identified as future technology for data storage. But its implementation was not



executed for technical and practical reasons. With steep rise in rate of data creation this field gain interest for research and development. Researchers search for data storage possibilities and various biological molecules which change their state with external stimuli e.g, Chlorophyll and haemoglobin. But DNA proved to be a stable data storage molecules due to some of its properties such as sequence of



nucleotides and translating the information stored.

Literature Review

The concept od Data Storage in DNA is relatively a new concept and only few universities and professional organisation have worked in this domen. MIT MIT USA and Google are among the pioneer researchers in this field. Following are the sources of literature which we go through for this Review paper:

1-DNA could store all of the world's data in one room
By- Robert F. Service Mar. 2, 2017.

2- New Trends of Data Storage in DNA

By- Pavani Yashodha D Silva Uni of Morathua Sri Lanla

3- DNA Data Storage

By- Luis and Reinhard

4- Is DNA is Future Data Storage

By- Layla Abdel Ilah

5- DNA Fountain Improve data storage

@ Evolving Science

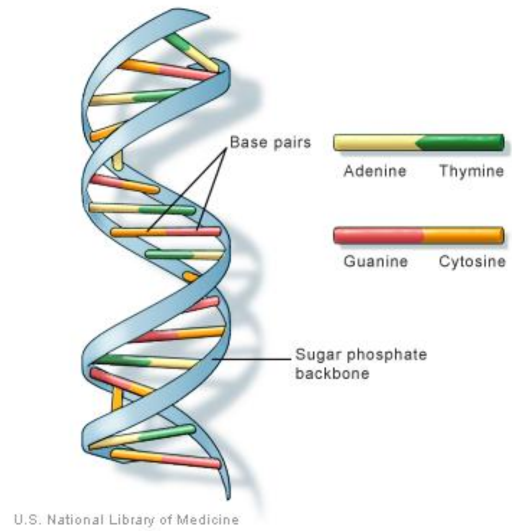
https://www.evolving-science.com/information-communication-computer-science-technology/dna-fountain-improves-data-storage-00207

6- A Living Hard Drive: This GIF Was Stored in the DNA of Bacteria
By- Edd Gent July 18, 2017
https://singularityhub.com/2017/07/16/a-living-hard-drive-this-gif-was-stored-in-the-dna-of-bacteria/

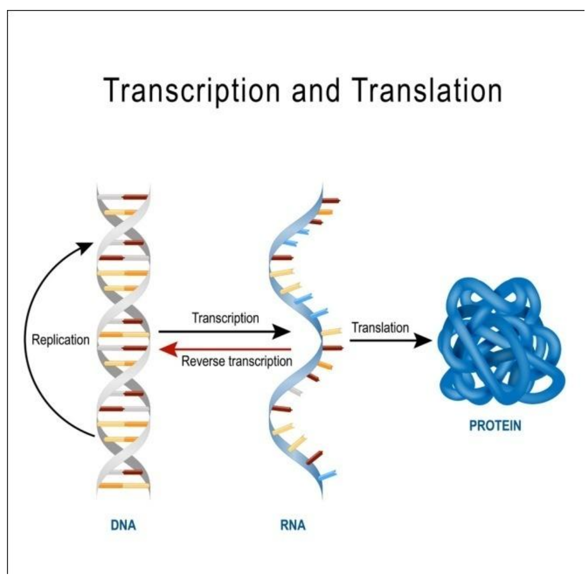
7- DNA Data Storage
From- Wikipedia
https://en.wikipedia.org/wiki/DNA_digital_data_storage

2. FUNCTIONING OF DNA

DNA is short form of Deoxy Ribo Nucleic Acid which is main constituent of Genetic Material of all living beings. DNA is a Double Helical Structure(Watson and Crick). Each helical is a polymer nucleotides and they are complementary and reverse in nature. Both helical are coiled around a common hypothetical axis. There are 5 types of nucleotides names Adenine , Thiamine, Guanine Cytosine and Uracil also referred as A, T, G, C, U. In the formation of DNA Adenine coupled with Thiamine and Guanine with Cytosine or Uracil. The sequence of these monomers decide the



development and functioning of whole life cycle of an organism. All cells of an organism



have same DNA and every cell have potential to be developed into a complete organism.This process is termed as Translation whole phenomenon is defined as Central Dogma.

3.

4. WHY DNA AS DATA STORAGE

In the age of big data, we are quickly producing far more digital information than we can possibly store. Last year, $20 billion was spent on new data centres in the US alone, doubling the capital expenditure on data centre infrastructure from 2016. And even with skyrocketing investment in data storage, corporations and the public sector are falling behind.

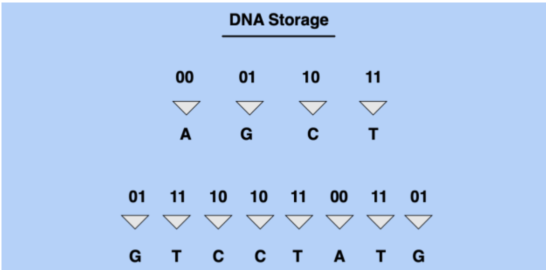
But there's hope, with a nascent technology leveraging DNA for data storage, this may soon become a problem of the past. By encoding bits of data into tiny molecules of DNA, researchers and companies like Microsoft hope to fit entire data centres in a few flasks of DNA by the end of the decade.

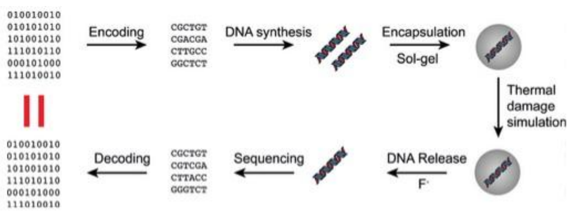
But let's back up. After the 20th century, we graduated from magnetic tape, floppy disks, and CDs to sophisticated semiconductor memory chips capable of holding data in countless tiny transistors. In keeping with Moore's Law, we've seen an exponential increase in the storage capacity of silicon chips. At the same time, however, the rate at which humanity produces new digital information is exploding. The size of the global data sphere is increasing exponentially, predicted to reach 160 zettabytes (160 trillion gigabytes) by 2025. As of 2018, digital users UPLODE over 44 billion gigabytes of data per day. By 2025, the International Data Corporation (IDC) estimates this figure will surpass 460 billion. And with private sector efforts to improve global connectivity—such as One-Web and Google's Project Loon, we're about to see an influx of data from five billion new minds. By 2020, three billion new minds are predicted to join the web. With private sector efforts, this number could reach five billion. While companies and services are profiting enormously from this influx, it's extremely costly to build data centres at the rate needed.

At present, about $50 million worth of new data centre construction is required just to keep up, not to mention millions in furnishings, equipment, power, and cooling. Moreover, memory-grade silicon is rarely found pure in nature, and researchers predict it will run out by 2040.

Take DNA, on the other hand. At its theoretical limit, we could fit 215 million gigabytes of data in a single gram of DNA.

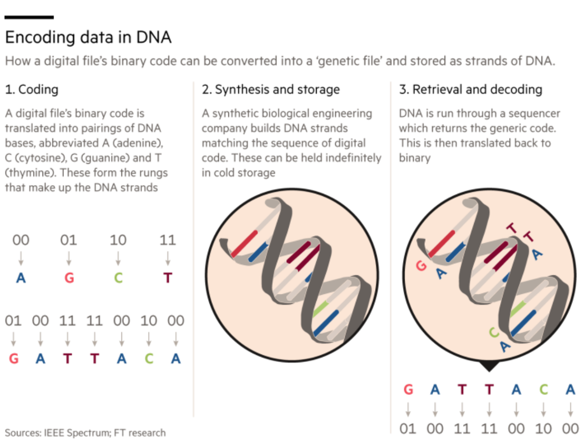
5. BUT HOW?DNA is built from a double helix chain of four nucleotide bases—adenine (A), thymine (T), cytosine (C), and guanine (G). Once formed, these chains fold tightly to form extremely dense, space-saving data stores. To encode data files into these bases, we can use various algorithms that convert binary to base nucleotides—0s and 1s into A, T, C, and G. "00" might be encoded as A, "01" as G, "10" as C, and "11" as T, for instance. Once encoded,



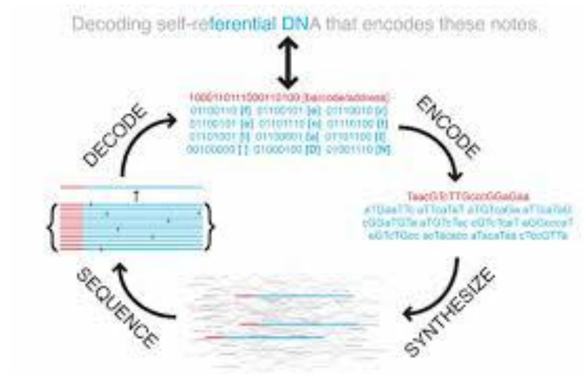


Scheme 1. Digital information is encoded to DNA and encapsulated within silica spheres. Upon release of the DNA from the spheres by fluoride chemistry, the DNA is read by Illumina sequencing and decoded to recover the original information, even if errors were introduced during the procedures.

information is then stored by synthesizing DNA with specific base patterns, and the final encoded sequences are stored in vials with an extraordinary shelf-life. To retrieve data, encoded DNA can then be read using any number of sequencing technologies, such as Oxford Nan pore's portable Minion. Still in its deceptive growth phase, DNA data storage—or NAM (nucleic acid memory)—is only beginning .



to approach the knee of its exponential growth curve. But while the process remains costly and slow, several players are beginning to crack its greatest challenge: retrieval. Just as you might click on a specific file and filter a search term on your desktop, random-access across large data stores has become a top priority for scientists at Microsoft Research and the University of Washington. Storing over 400 DNA-encoded

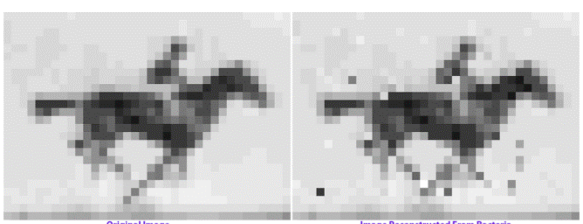


megabytes of data, U Washington's DNA storage system now offers random access across all its data with no bit errors.

6. Applications area of DNA Data Storage Technology

Even before we guarantee random access for data retrieval, DNA data storage has immediate market applications. According to IDC's Age 2025 study a huge proportion of enterprise data goes straight to an archive. Over time, the majority of stored data becomes only potentially critical, making it less of a target for immediate retrieval. Particularly for storing past legal documents, medical records, and other archive data, why waste precious computing power, infrastructure, and overhead? Data-encoded DNA can last 10,000 years—guaranteed—in cold, dark, and dry conditions at a fraction of the storage cost. Now that we can easily use natural enzymes to replicate DNA, companies have tons to gain (literally) by using DNA as a backup system—duplicating files for later retrieval and risk mitigation. And as retrieval algorithms and biochemical technologies improve, random access across data-encoded DNA may become as easy as clicking a file on your desktop. As you scroll,

researchers are already investigating the potential of molecular computing, completely devoid of silicon and electronics. Harvard professor George Church and his lab, for instance, envision capturing data directly in DNA. As Church has stated, "I'm interested in making biological cameras that don't have any electronic or mechanical components,"



(The scientists from Harvard University used the CRISPR genome-editing tool to store a picture of a hand and an animation of a running horse adapted from Eadweard Muybridge's 1878 photographic study *Human and Animal Locomotion* in the genome of *E. coli* bacteria.)

whereby information "goes straight into DNA." According to Church, DNA recorders would capture audiovisual data automatically. "You could paint it up on walls, and if anything interesting happens, just scrape a little bit off and read it—it's not that far off." One day, we may even be able to record biological events in the body. In pursuit of this end, Church's lab is working to develop an in vivo DNA recorder of neural activity, skipping electrodes entirely. Perhaps the most ultra-compact, long-lasting, and universe of DNA storage mechanism at our fingertips, DNA offers us unprecedented applications in data storage—perhaps even computing.

7. Potential of DNA Technology

As DNA data storage plummets in tech costs and rises in speed, commercial user interfaces will become both critical and wildly profitable. Once corporations, start-ups, and people alike can easily save files, images or even neural activity to DNA, opportunities for disruption abound. Imagine uploading files to the cloud, which travel to encrypted DNA vials, as opposed to massive and inefficient silicon-enabled data centres. Corporations could have their own warehouses and local data networks could allow for heightened cyber security particularly for archives. And since DNA lasts millennia without maintenance, forget the need to copy databases and power digital archives. As long as we're human, regardless of technological advances and changes, DNA will always be relevant and readable for generations to come. But perhaps the most exciting potential of DNA is its portability. If we were to send a single Exabyte of data (one billion gigabytes) to Mars using silicon binary media, it would take five Falcon Heavy rockets and cost $486 million in freight alone. With DNA, we would need five cubic centimetres. At scale, DNA has the true potential to dematerialize entire space colonies worth of data. Throughout evolution, DNA has unlocked extraordinary possibilities—from humans to bacteria.

Soon hosting limitless data in almost zero space, it may one day unlock many more.

8. DRAWBACK OF DNA STORAGE TECHNOLOGY:

Each and every technology also have some demerits along with its merits. DNA data storage is not an exception to this.

Here are just a few of the disadvantages of using DNA for data storage.

- Portability - hard to share the data since only a very limited number of places can write / read digital data in DNA.
- Expensive - about 10 cents for encoding a million base pairs, which is equivalent to about a dollar per megabyte (INR 70000 per GB, and INR 70 M per TB). We can buy a 1TB external drive a lot less - Best 1TB External Hard Drive for under INR 4000
- Reliability - DNA is reactive to the environment around it and can degrade over time
- Standards - None exist as yet for encoding data.
- Coding and decoding process is very slow and till now no reliable technology is developed

9. DEVICE FOR DNA DATA STORAGE

Founded in 2013, Austrian firm Kilobaser has raised an undisclosed amount of funding to develop the "Nespresso machine of DNA synthesis". They're already taking reservations for the $9,351 machine which creates DNA primers. (According to Wikipedia, a primer is a short single strand of RNA or DNA (generally about 18-22 bases) that serves as a starting point for DNA synthesis.) Here's the machine along with some basic specs:

Ready to use 25 bases primer:	**< 1 hour**
Scale:	**500 pmol**
Max. lenght:	**100 bases**
Only two consumables:	**Reagent cartridge & microfluidic chip**
Columns:	**Single (Dual)**

The machine is expected to ship in August 2019.

10. COMMERCIAL DATA STORAGE ON DNA

It was in January of 2017 when an Irish startup that was selling a DNA storage device on Amazon. there are now a whole bunch of DNA data storage companies that have popped up.

CATALOG
INFINITE DATA ARCHIVES

evonetix

11. References

DNA Fountain Improve data storage

https://www.evolving-science.com/information-communication-computer-science-technology/dna-fountain-improves-data-storage-00207

Wikipedia
https://en.wikipedia.org/wiki/DNA_digital_data_storage

A Living Hard Drive: This GIF Was Stored in the DNA of Bacteria
By- Edd Gent July 18, 2017
https://singularityhub.com/2017/07/16/a-living-hard-drive-this-gif-was-stored-in-the-dna-of-bacteria/

Disadvantages of storing digital data in DNA
https://www.quora.com/What-are-the-disadvantages-of-storing-digital-data-in-DNA

https://www.nanalyze.com/2018/09/dna-data-storage-companies/

https://www.wired.com/story/the-rise-of-dna-data-storage/