ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 6, May 2018

2nd National Conference on "RECENT TRENDS IN COMPUTER SCIENCE & INFORMATION TECHNOLOGY (RTCSIT'18)"

13th-14th March 2018

Organized by

Departments of Computer Science & Information Science Engineering, Sri Krishna Institute of Technology, Bangalore, India

Next Generation Data Storage Innovation Using DNA as a Medium and Self Referential DNA

Roop Raj Thapa[1], Rahul Kumar Pandit[2] , Rajesh V[3]

U.G. Student, Department of Computer Science and Engineering, Sri Krishna Institute of Technology, Bengaluru, India[1]

U.G. Student, Department of Computer Science and Engineering, Sri Krishna Institute of Technology, Bengaluru, India[2]

Asst Professor, Department of Computer Engineering, Sri Krishna Institute of Technology, Bengaluru, India[3]

ABSTRACT: People have dependably been attached to accessing increasingly information in minimum conceivable time and space. Subsequently New Generation Computers and High Speed Internet have gained ubiquity in the current years. We have been observer to striking accomplishments like the change from the massive hard-drives to the glimmer drives which has made individual data storage effectively reasonable. Yet, with regards to handling enormous data, the data of a partnership or of the world in general, the present data storage innovation comes no place close to have the capacity to oversee it effectively. An earnest requirement for an appropriate medium for information authentic and recovery purposes emerges. Deoxyribonucleic acid (DNA) is viewed as a potential medium for such purposes, basically on the grounds that it is like the consecutive code of 0's and 1's in a PC. This field (DNA Computing) has developed to end up a theme of interest for specialists since the previous decade, with significant achievements in its course. Seeming to come straight out of sci-fi, "a penny-sized gadget could store the whole information as the entire Internet". The broke down data from the investigates uncovers that only four grams of DNA has the capability of storing all the information that the world creates in a year. Here, this subject of 'Data Storage in DNA' is portrayed starting from the primary research to the latest one, their strategies, their favorable circumstances and their defects, the necessity for DNA storage, and how it will at last turn into a paradigm shift in computing.

KEYWORDS: Binarization, PCR, Encoding, Synthesis, Sequencing, Decoding, DNA.

I. INTRODUCTION

DNA data storage alludes to any procedure to store data in the base arrangement of DNA using economically accessible oligonucleotide amalgamation machines for the purpose of storage and DNA sequencing machines forthr retrieval of stored data.

Data storage and retrieval is inevitable, and its protection issue is looming over our information organize. The interest for storing an ever increasing number of data is increasing step by step. In 2012, the aggregate computerized information on the planet was around 2.7 zettabytes. With each passing year it is outgrown by its forerunner by half. The advantage while working with DNA is a million times more proficient than a cutting edge PC. DNA is an extremely powerful material and has a long time span of usability with no constriction in data. Data in DNA is put away in a volumetric manner (using Adenine, Thymine, Guanine, Cytosine bases) which offers access to more storage choices not at all like present mediums which store data in a linear request. Hypothetically DNA can encode 2 bits for each nucleotide or 455 exabytes for each gram of DNA. Since the whole arrangement never gets harmed during denaturation, the remaining succession can be opened up to obtain the original one indeed. DNA likewise has the ability for life span, as long as the DNA is held in cool, dry and dull conditions. It can be appropriately ensured in a spore, for instance, and safeguarded for many years. It can be effortlessly increased by Polymerase Chain Reaction

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology
An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 6, May 2018
2nd National Conference on "RECENT TRENDS IN COMPUTER SCIENCE & INFORMATION TECHNOLOGY (RTCSIT'18)"
13th-14th March 2018
Organized by
Departments of Computer Science & Information Science Engineering, Sri Krishna Institute of Technology, Bangalore, India

systems to get the coveted number of its duplicates. A standout amongst the most noteworthy preferences of using DNA as a storage medium is that the storage thickness is high

What is DNA:deoxyribonucleic acid, a self-replicating material present in nearly all living organisms as the main constituent of chromosomes. It is the carrier of genetic information. DNA is well-suited for biological information storage.

Importance of DNA as storage medium: High capacity for storage, High data storage density , Withstand extreme environmental conditions range from -800 to +800, High memory space for the data storage , Secure as invisible to human naked eye , Can store data for long periods by protecting from water and oxygen

Storage capacity of DNA: Goldman and co have possessed the capacity to store 2.2 petabytes in a single gram of DNA. DNA would permit no less than 100 million hours of top quality video to be put away in a teacup. Around 1kg of DNA can store the whole world's data.

Feature of DNA: Successful power utilization Long-lived, stable and effortlessly combined. Needs no dynamic maintenance. Stores computerized records without power for a great many years. Keep going for a huge number of years. We can store 2.2 peta bytes in a single gram of DNA. Profoundly solid

II. RELATED WORK

In year1994, according to Adelman it was indicated to store information and completed a couple of figurings correspondingly, the DNA may be utilized. And furthermore it is expressed by Ad leman in 1994 that DNA utilized four bases with each input parameter

The few looks into so far have been basically investigated here. The possibility of Digital Storage in DNA was first indirectly actualized in 1999 by Clell and, Risca, Bancroft . They prevailing with regards to storing encoded words in short DNA strands. They prevailing with regards to storing encoded words in short DNA strands (Microdots).The innovation was utilized as a part of World War II to convey mystery data.

The DNA containing imperative information was mainly in light of polymerase chain response arrangements which were encompassed by 109 times its size worth of concealing human DNA atoms. It gave an extremely complex foundation to shroud the message DNA consequently ensuring data security and protection.

The few explores so far have been fundamentally audited here. The possibility of Digital Storage in DNA was first indirectly executed in 1999 by Clell and, Risca, Bancroft [2] and [3]. They prevailing with regards to storing encoded words in short DNA strands (Microdots).The innovation was utilized as a part of World War II to convey mystery data.

The specialists executed it on a DNA scale, i.e. attempted to conceal profitable information in a DNA strand (and consequently were the first to store information onto DNA). They utilized the 4 bases, PCR Primers and an Encryption key (base triplets to speak to English letters in order and Arabic Numbers).

The coveted DNA-encoded message was first covered up within the basically complex denatured human genomic DNA and after that further scaling down this example to a microdot. The DNA containing vital information was mainly in light of polymerase chain response successions which were encompassed by 109 times its size worth of concealing human DNA particles. It gave an exceptionally complex foundation to shroud the message DNA accordingly ensuring data security and protection. Their imperative finding was that if the coveted PCR groundwork arrangements and the particular encryption key are known, at that point the required DNA could be promptly opened up and examined by gel electrophoresis regardless of the cover gave by the denatured concealing DNA.

This finding made it clear that the coveted DNA can be encoded regardless of how much 'commotion' or undesirable arrangements it is encompassed with (Data is painstakingly kept secure within it, it won't be lost). Moreover, this examination on data stirage in DNA had presented the possibility of the DNA Storage is substantially more private and secure than Digital Storage on Silicon gadgets that too without an express encryption system.

III. METHODOLOGY

A System Model

The System model of data storage in DNA contains the following elements: DNA basics, PCR, Binarization, Encoding, Synthesis, Sequencing, Decoding

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 6, May 2018

2nd National Conference on "RECENT TRENDS IN COMPUTER SCIENCE & INFORMATION TECHNOLOGY (RTCSIT'18)"

13th-14th March 2018

Organized by

Departments of Computer Science & Information Science Engineering, Sri Krishna Institute of Technology, Bangalore, India

DNA basics:- Naturally occurring DNA consists of four types of nucleotides: adenine (A), cytosine (C), guanine (G), and thymine (T). DNA strand, or oligonucleotide, is a direct arrangement of these nucleotides.

PCR:- is a strategy for exponentially intensifying the centralization of chose arrangements of DNA inside a pool. A PCR response requires four principle segments: the layout, sequencing preliminaries, a thermostable polymerase and individual nucleotides that get joined into the DNA strand being increased. The layout is a solitary or twofold stranded atom containing the (sub)sequence that will be intensified

DNA synthesis:-Subjective single-strand DNA successions can be combined artificially, nucleotide by nucleotide. The coupling proficiency of a combination procedure is the likelihood that a nucleotide ties to a current fractional strand at each progression of the procedure. In spite of the fact that the coupling proficiency for each progression can be higher than 99%, this little mistake still outcomes 3 out of an exponential diminishing of item yield with expanding length and limits the extent of oligonucleotides that can be productively combined to around 200 nucleotides. DNA sequencing.

DNA sequencing. There are several high-throughput sequencing techniques, but the most popular methods (such as that used by Illumina) use DNA polymerase enzymes and are commonly referred to as "sequencing by synthesis". The strand of interest serves as a template for the polymerase, which creates a complement of the strand. Importantly, fluorescent nucleotides are used during this synthesis process.

Payload. The series of nucleotides speaking to the information to be put away is broken into information obstructs, whose length relies upon the coveted strand length and the extra overheads of the organization. To help translating, two sense nucleotides show whether the strand has been invert supplemented Address. Every datum piece is increased with tending to data to recognize its area in the information string. The address space is in two sections.

Address. Each data block is augmented with addressing information to identify its location in the input data string. The address space is in two parts. The high part of the address identifies the key a block is associated with. The low part of the address indexes the block within the value associated with that key.

Primers. To each end of the strand, we attach primer sequences. These sequences serve as a "foothold" for the PCR process, and allow the PCR to selectively amplify only those strands with a chosen primer sequence.

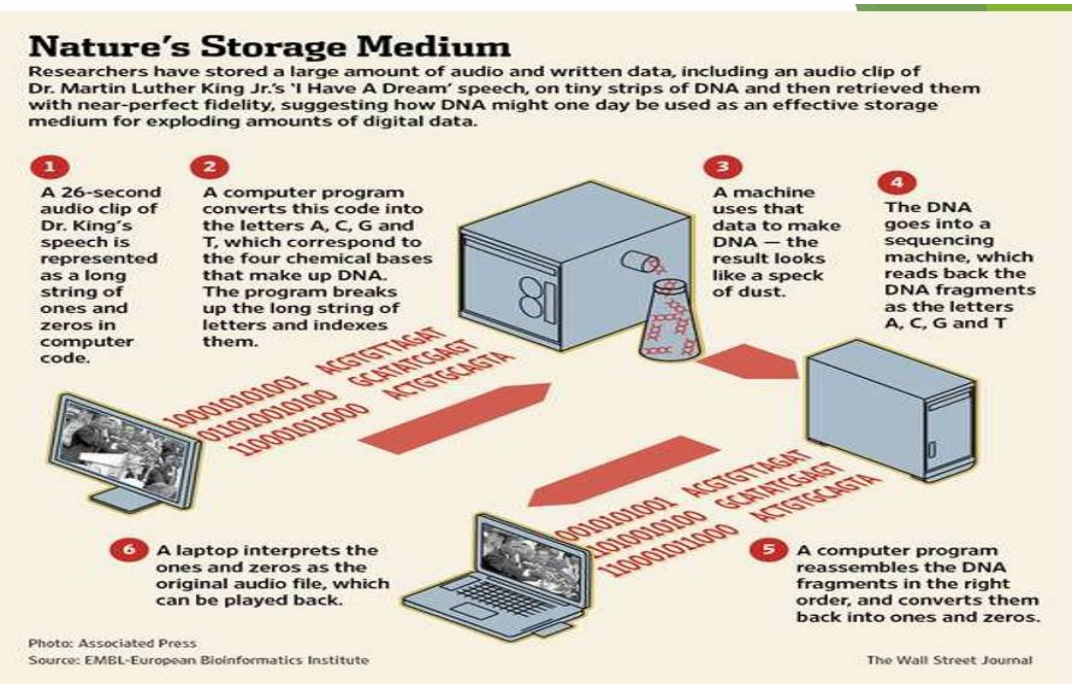

Fig 1 System Model for overall process

We imagine DNA storage as the last level of a profound storage chain of importance, providing exceptionally thick and solid recorded storage with get to times of numerous hours to days. DNA combination and sequencing can be made

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 6, May 2018

2nd National Conference on "RECENT TRENDS IN COMPUTER SCIENCE & INFORMATION TECHNOLOGY (RTCSIT'18)"

13th-14th March 2018

Organized by

Departments of Computer Science & Information Science Engineering, Sri Krishna Institute of Technology, Bangalore, India

subjectively parallel, making the important read and compose data transfer capacities attainable. We now portray our proposition of a framework for DNA-based storage with irregular access bolster.

A DNA storage framework comprises of a DNA synthesizer that encodes the data to be put away in DNA, a storage container with compartments that store pools of DNA that guide to a volume, and a DNA sequencer that peruses DNA arrangements and proselytes them once more into computerized data.

Interface and Addressing A storage framework needs an approach to relegate recognizable proof labels to data questions so they can be recovered later. We pick a basic key-esteem design, where a put(key, esteem) activitypartners an incentive with key, and a get(key) task recovers the esteem relegated to key. To execute a key-esteeminterface in a DNA storage framework, we need:a work that maps a key to the DNA pool

DataDebasement: Due to the regular information access and refresh tasks, the likelihood of information blunders happen increments and the life time of storage medium turns out to be short. In this manner, the information misfortune mishaps may happen out of the blue and the cloud storage suppliers attempt to shroud these mischances to keep up a decent notoriety. The clouds may likewise erase a few information that are once in a while or never went to lease the spaces to different clients for money related benefits. Also, Cloud A may send a bit of the expected information to Cloud B and claim that every one of the information are exchanged,or send manufactured information to cheat the client.

B. Data Format: Another useful issue with representing data in DNA is that ebb and flow amalgamation innovation does not scale past arrangements of low many nucleotides. Data past the many bits subsequently can't be incorporated as a single strand of DNA. Also, DNA pools don't offer spatial confinement, and so a pool contains data for some, unique keys which are immaterial to a single read activity. Isolating just the particles of interest is non-paltry, and so existing DNA storage systems for the most part grouping the whole arrangement, which incurs huge cost and time overheads. To conquer these two difficulties, we arrange data in DNA in a comparable manner to Goldman et al. Segmenting the nucleotide portrayal into squares, which we integrate as particular strands, permits storage of huge esteems. Tagging those strands with identifying groundworks permits the read procedure to segregate particles of interest and so perform random access.

• **Payload:** The string of nucleotides representing the data to be put away is broken into data hinders, whose length relies upon the coveted strand length and the extra overheads of the organization. To help decoding, two sense nucleotides indicate whether the strand has been turn around supplemented (this is done to maintain a strategic distance from certain neurotic cases).

• **Address:** The string of nucleotides representing the data to be put away is broken into data hinders, whose length relies upon the coveted strand length and the extra overheads of the organization. To help decoding, two sense indicate whether the strand has been turn around supplemented (this is done to maintain a strategic distance from certain neurotic cases).

• **Primerss:** To each finish of the strand, we connect preliminary arrangements. These groupings fill in as a "decent footing" for the PCR procedure, and permit the PCR to specifically open up just those strands with a picked groundwork succession.

• **Random Acess:** We abuse preliminary groupings to give random access: by assigning distinctive groundworks to various strands, we can perform sequencing on just a chose gathering of strands. Existing work on DNA storage utilizes a single groundwork succession for all strands. While this outline gets the job done for data recuperation, it is inefficient: the whole pool (i.e., the strands for each key) must be sequenced to

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 6, May 2018

2nd National Conference on "RECENT TRENDS IN COMPUTER SCIENCE & INFORMATION TECHNOLOGY (RTCSIT'18)"

13th-14th March 2018

Organized by

Departments of Computer Science & Information Science Engineering, Sri Krishna Institute of Technology, Bangalore, India

C. PROPOSED DATA STORAGE SCHEME

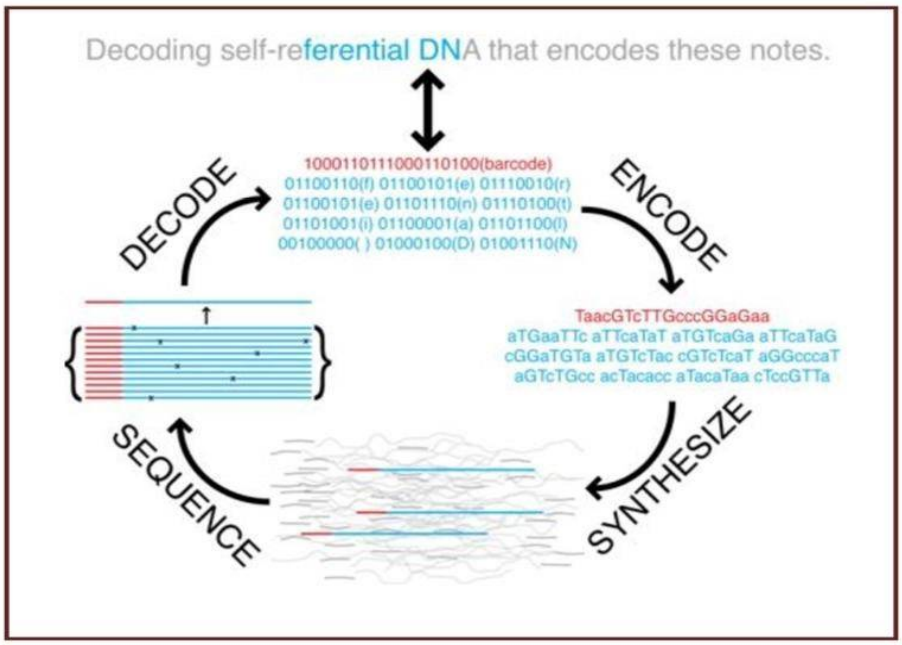

Fig 2: Encoding, Decoding, Synthesize, Sequence process

Example for coding of information

First we should use numbers to represent the letters in ASCII code
From ASCII table
S=83
K=75
I=73
T=84
Change to quaternary numbers
83= 1103
75= 1023
73 = 1021
84 = 1110
Use "A, T, C & G" to represent the numbers
0 = A
1 = T
2 = C
3 = G

This A, G, C, T sequence avoids any reading errors, particularly when encountering repetitive base sequences. Also, rather than synthesize one long string of DNA to code for an entire item of information, they broke the file down to smaller chunks, so that no errors occur during synthesis or read-out. These chunks are then read in an appropriate manner or protocol, providing for 100 per cent accuracy.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 6, May 2018

2nd National Conference on "RECENT TRENDS IN COMPUTER SCIENCE & INFORMATION TECHNOLOGY (RTCSIT'18)"

13th-14th March 2018

Organized by

Departments of Computer Science & Information Science Engineering, Sri Krishna Institute of Technology, Bangalore, India

SKIT 1103102310211110
TTAGTACGTACTTTTA

IV. EXPERIMENTAL RESULTS

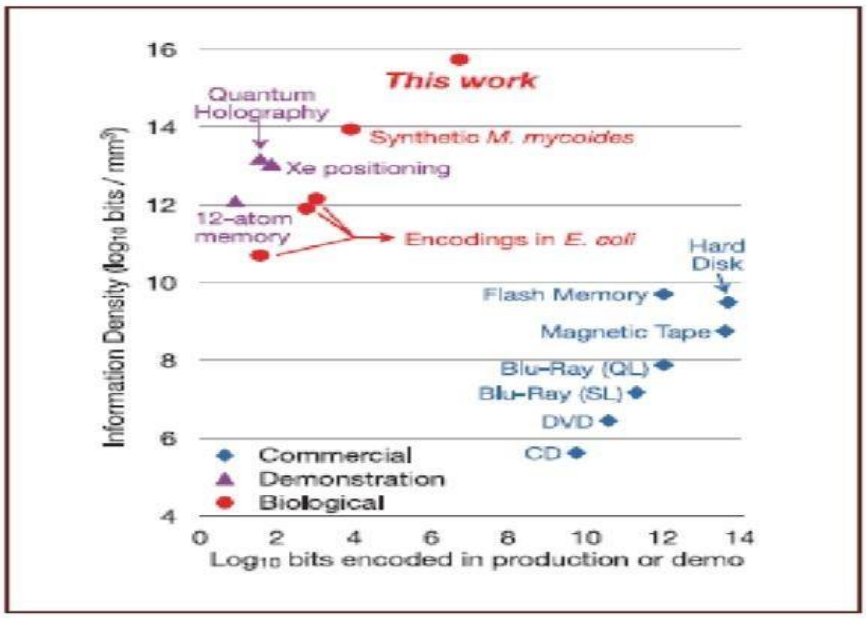

Fig 3 Observations plotted on graph for the experiments by church et al

We imagine DNA stockpiling as the last level of a profound stockpiling chain of command, giving extremely thick and tough chronicled stockpiling with get to times of numerous hours to days (Figure 2). DNA combination and sequencing can be made subjectively parallel, making the vital read and compose transmission capacities achievable. We now depict our proposition of a framework for DNA-based capacity with irregular access bolster.
A DNA stockpiling framework comprises of a DNA synthesizer that encodes the information to be put away in DNA, a capacity holder with compartments that store pools of DNA that guide to a volume, and a DNA sequencer that peruses DNA arrangements and proselytes them once again into advanced information.
A capacity framework needs an approach to allocate distinguishing proof labels to information protests so they can be recovered later. We pick a basic key-esteem design, where a put(key, esteem) task partners an incentive with key, and a get(key) activity recovers the esteem doled out to key. To execute a key-esteem interface in a DNA stockpiling framework, we need:a work that maps a key to the DNA pool (in the library) where the strands that contain information live; and a system to specifically recover just wanted parts of a pool (i.e, arbitrary access), since the DNA compartment will probably store essentially a larger number of information than the coveted question.

V. CONCLUSION

Information stockpiling in DNA is not any more restricted to science fiction however is being acknowledged and ad libbed at extremely encouraging rates by explore groups everywhere throughout the world. This thought has gotten positive feedback from the overall population as can be surmised from their reactions on the distinctive science sites.
Like all transformations in innovation, DNA-based information stockpiling innovation needs to confront significant difficulties to understand its maximum capacity. It is nonetheless, unavoidable that DNA would be perpetually utilized for chronicled purposes for its sheer thickness, heartiness, security and vitality productivity. In principle, grams of

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 7, Special Issue 6, May 2018

2nd National Conference on "RECENT TRENDS IN COMPUTER SCIENCE & INFORMATION TECHNOLOGY (RTCSIT'18)"

13th-14th March 2018

Organized by

Departments of Computer Science & Information Science Engineering, Sri Krishna Institute of Technology, Bangalore, India

DNA can store all the data at any point delivered by humankind. A few leaps forward will be required before it turns out to be financially standard for information recovery.

This field has had a million-overlay change in the current years. Computerized Data Storage in DNA innovation demonstrates huge advance, since perusing and composing it is propelling ten times each year dissimilar to the Electronic Technology which is enhancing around 1.5 times each year (Moore's Law).

Cost productivity likewise demonstrates guarantee as the exponential drop in DNA combination and sequencing cost has been five-overlay and twelve-overlap individually while Electronic media indicate just a 1.6-crease drop every year. As handheld sequencers are ending up economically accessible, the exploration would develop past substantial activities, with more individuals gaining admittance to try different things with this thought. "As DNA is the premise of life on Earth, strategies for working with it, putting away it and recovering it will remain the subject of persistent mechanical advancement", says Nick Goldman. With a specific end goal to deal with huge information a down to earth largescale DNA chronicle would require stable DNA administration and inventive ordering arrangements. This would make the coveted change in outlook in registering as the part of Data Storage would be a fundamental part to understand the possibility of DNA Computing. Proportionally, it would catalyze innovative work in blend and sequencing advances.

Future work could incorporate pressure plans; managing repetition at all levels, checking for equality, rectifying blunders to upgrade thickness and wellbeing. DNA could likewise be substituted with polymers or be altered to suit the requirements of computerized stockpiling. Besides, it will fuel research to search for elective materials for data stockpiling and to help in understanding the requirement for a widespread medium for information. This innovation is staying put and could change the way we have ever taken a gander at DNA and figuring as entirely unexpected elements. Dissecting the ecological conditions under which the investigations are completed. Protection of DNA for long haul stockpiling.

REFERENCES

[1] N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E. M. Le Proust, B. Sipos, and E. Birney, "Towards practical, high-capacity, lowmaintenance information storage in synthesized DNA," Nature, vol. 494, no. 7435, pp. 77-80, 2013.
[2] C. T. Clelland, et al., "Hiding messages in DNA microdots," Nature, vol. 399, no. 1033, pp. 533–534, 1999.
[3] C. Bancroft, T. Bowler, B. Bloom, and C. T. Clelland, "Longterm storage of information in DNA," Science, vol. 293, no. 5536, pp. 1763-1765, 2001.
[4] P. C. Wong, K. K. Wong, and H. Foote, "Organic data memory, using the DNA approach," ACM, vol. 46, no. 1, pp. 95-98, 2003.
[5] D. A. Huffman, "A method for the construction of minimumredundancy codes," in Proc. IRE., vol. 40, 1953, pp. 1098-1101.
[6] M. Ailenberg and O. D. Rotstein, "An improved Huffman coding method for archiving text, images, and music characters in DNA," Biotechniques, vol. 47, no. 3, pp. 747–754, 2009.
[7] G. M. Church, Y. Gao, and S. Kosuri, "Next-generation digital information storage in DNA," Science, vol. 337, no. 6102, pp. 1628, 2012.
[8] J. P. L. Cox, "Long-term data storage in DNA," Trends Biotechnol., vol. 19, no. 7, pp. 247–250, 2001.
[9] Reading and Writing a Book with DNA. [Online]. Available: http://spectrum.ieee.org/biomedical/imaging/reading-and-writinga-book-with-dna
[10] R. A. LEO. Writing the Book in DNA. [Online]. Available: http://hms.harvard.edu/news/writing-book-dna-8-16-12
[11] E. Regis and G. M. Church, Regenesis, How Synthetic Biology Will Reinvent Nature and Ourselves, Basic Books, ch. 2, 2012, pp. 101-103.
[12] Digital Archiving. History Flushed. [Online]. Available: http://www.economist.com/node/21553410
[13] J. Gantz and D. Reinsel, "Extracting value from chaos," IDC IVIEW, June 2011.
[14] S. Murray, V. Bahyl, G. Cancio, et. al "Tape write-efficiency improvements in CASTOR," presented in J. Phys.: Conf. Ser. 396 042042 in 2012.
[15] D. G. Gibson et al., "Creation of a bacterial cell controlled by a chemically synthesized genome," Science, vol. 329, no. 5987, pp. 52-56, 2010.