Data Storage in DNA

Siddhant Shrivastava and Rohan Badlani
Birla Institute of Technology and Science, Pilani, India
Email: sidhu94@gmail.com, rohan.badlani@gmail.com

Abstract—**Human-beings have always been fond of accessing more and more information in minimum possible time and space. Consequently New Generation Computers and High Speed Internet have gained popularity in the recent years. We have been witness to remarkable achievements like the transition from the bulky hard-drives to the flash drives which has made personal data storage efficiently manageable. But when it comes to handling big data, the data of a corporation or of the world as a whole, the present data storage technology comes nowhere near to be able to manage it efficiently. An urgent need for a proper medium for information archival and retrieval purposes arises. Deoxyribonucleic acid (DNA) is seen as a potential medium for such purposes, essentially because it is similar to the sequential code of 0's and 1's in a computer. This field (DNA Computing) has emerged to become a topic of interest for researchers since the past decade, with major breakthroughs in its course. Seeming to come straight out of science fiction, "a penny-sized device could store the entire information as the whole Internet". The analyzed data from the researches reveals that just four grams of DNA can store all the information that the world produces in a year. Here, this topic of 'Data Storage in DNA' is described starting from the very first research to the most recent one, their techniques, their advantages and their flaws, the need for DNA storage, and how it will ultimately become a paradigm shift in computing.**

Index Terms—**encoding, DNA, nucleotides, encryption scheme, oligonucleotides, storage mechanisms, coding theory, digital data**

I. INTRODUCTION

Data storage and retrieval is inevitable and its preservation problem is looming over our information network. The demand for storing more and more data is increasing day by day. In 2012, the total digital information in the world was about 2.7 zettabytes. With every passing year it is outgrown by its predecessor by 50%. The journey of data storage began from Rocks, Bones, Paper, Punched Cards, Magnetic Tapes, Drums, Films, Gramophone records Floppies etc. Data storage has in the present scenario extended to optical discs including CDs, DVDs, Blu-ray Discs to Portable hard drives and USB flash drives. But all of these techniques are prone to obsolescence and decay. Moreover, Silicon and the other non-biodegradable materials pollute the environment, are limited in resources and would exhaust one day. The Maximum Storage Density on these devices

is 1 Terabyte per square inch while the projected data demand would be 8000 Exabytes till 2015. Libraries, Corporations and File Sharing Systems are in favor of shifting to newer technologies for archival purposes. The current Storage Technologies definitely are not competent enough to handle it efficiently and archive it for the distant future. For instance, the European Organization for Nuclear Research-CERN's CASTOR (CERN Advanced STORage manager) system stores 0.08 exabytes of Large Hadron Collider data and grows at 15 PB every year. To store all this information disks are used only for 10% data, and magnetic tapes need to be used which have access rates reducing every two years. Thus potentially important information is lost for a lack of better archival systems.

So to find new solutions to the issues of Digital Data Storage, new technologies and principles are in a state of innovative experimentation throughout the world.

Scientists and Researchers from different parts of the world, over the past decade, have been testing to develop a robust way of storing non-biological data on a medium that is dense, universal, non-obsolete, ever-lasting and enduring. They are sticking to the basics by considering Mother Nature's Storage medium, DNA (Deoxyribonucleic Acid). There are several reasons to use DNA as the storage medium. Its storage density and small size (Occupying just 1cm i.e. 1 gram of dry DNA has a storage potential of 455 exabytes of information), something which would take conventional media roughly 2 million times that volume for the same amount of information. Thus data on DNA can be conveniently stored. It can sustain a wider range of temperatures (-800 to 800 C). A gram of DNA contains 1021 DNA bases which can correspond to 108 terabytes of binary data.

The Power Usage while working with DNA is a million times more efficient than a modern personal computer. DNA is a very robust material and has a very long shelf life with no attenuation in data. Data in DNA is stored in a volumetric fashion (using Adenine, Thymine, Guanine, Cytosine bases) which gives access to more storage options unlike present mediums which store data in a linear order. Theoretically DNA can encode 2 bits per nucleotide or 455 exabytes per gram of DNA. Since the entire sequence never gets damaged during denaturation, the remaining sequence can be amplified to obtain the original one once again. DNA also has the capability for longevity, as long as the DNA is held in cold, dry and dark conditions. It can be suitably protected in a spore, for example, and preserved for millions of years. It can be easily amplified by Polymerase Chain

Reaction techniques to get the desired number of its copies. One of the most significant advantages of using DNA as a storage medium is that the storage density is very high. For example, it was found in a research by Hoch and Losick that the density to contain characters (char/m^2) of a Bacillus subtilis bacterium (genome size 4.2 Mega Base Pairs, with 1 μm diameter) spore is twenty million times that of a 200 Megabyte ZIP disk of diameter 10 cm. DNA sequences can contain more information than their binary counterpart because DNA with 4 bases has 4X representations possible for a X-character long string while binary system can represent only 2-X times that information.

As stated by Goldman *et al* [1]: "Isolated DNA fragments are easily manipulated in vitro and that the routine recovery of intact fragments from samples that are tens of thousands of years old indicates that well-prepared synthetic DNA should have an exceptionally long lifespan in low-maintenance environments. Moreover, DNA can be read as a code (strings stored as bases) in both directions, a feature of DNA that promises more chance of data extraction and improved latency. DNA is nearly invisible to the human eye, a fact which ensures, it can't be harmed easily unlike present silicon devices. These inevitable advantages set the ground for research in the following years. (1999-2013 are discussed here)

II. A REVIEW OF RESEARCHES SO FAR

The several researches so far have been critically reviewed here.

The idea of Digital Storage in DNA was first indirectly implemented in 1999 by Clell and, Risca, Bancroft [2] and [3]. They succeeded in storing encoded words in short DNA strands (Microdots).The technology was used in World War II to communicate secret data. A microdot was a downscaled picture of a typed page encoded in a period (.) in a harmless letter. The researchers implemented it on a DNA scale, i.e. tried to hide valuable information in a DNA strand (and hence were the first to store information onto DNA). They used the 4 bases, PCR Primers and an Encryption key (base triplets to represent English alphabets and Arabic Numbers). The desired DNA-encoded message was first hidden within the structurally complex denatured human genomic DNA and then further scaling down this sample to a microdot. The DNA containing important information was mainly based on polymerase chain reaction sequences which were surrounded by 109 times its size worth of concealing human DNA molecules. It provided a very complex background to hide the message DNA thus ensuring data security and privacy. Their important finding was that if the desired PCR primer sequences and the specific encryption key are known, then the required DNA could be readily amplified and analysed by gel electrophoresis irrespective of the mask provided by the denatured concealing DNA.

This finding made it clear that the desired DNA can be encoded no matter how much 'noise' or unwanted sequences it is surrounded with (Data is carefully kept secure within it, it won't be lost). Additionally, this research brought forward the idea that DNA Storage is much more private and secure than Digital Storage on Silicon devices that too without an explicit encryption mechanism. The following Fig. 1 summarizes their research, coding scheme and the code they stored.
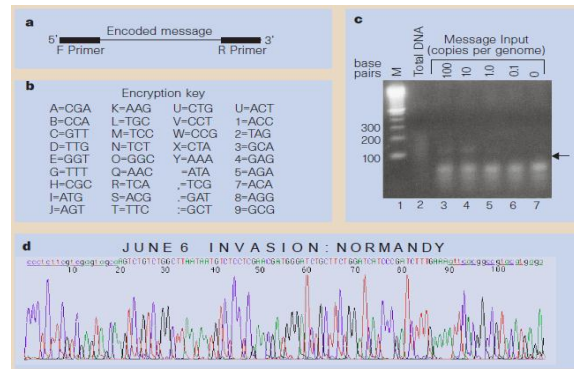

Figure 1.   Summarizes the research work and coding schemes.

The same idea was carried forward by Bancroft, Bowler, Bloom, and Clelland [3] in 2001. They used the similar mechanism for encryption and storage. Using iDNA (information DNA – the encoded data) and the Poly primer Key (The primer base sequence to access the information on the iDNA) they devised an experiment which included amplification by PCR giving the 'universal' forward and reverse primers in sufficient amount to analyse it using the decided encryption scheme. Each linearly analysed sequencing primer would yield ordered fragments of iDNA. Analyzing these orderly arranged sequences inside, the data could be decoded using the designed Encryption Scheme. The data which they stored and retrieved from DNA were the opening lines of Charles Dickens's 'A Tale of Two Cities'. Though this experiment was a direct consequence of the Microdot experiment, it was for the first time a DNA-based data storage mechanism was used. Thus this finding proved that development of such storage devices was possible using the simple principles of biology. However it did not take into account the safety and tolerance of the DNA to extreme external and internal conditions, since it was done under laboratory conditions. Nevertheless this research paved the way for future advances to come in this exciting new field of DNA Computing. The Fig. 2 shows the structure of DNA molecules used for information storage and reading.
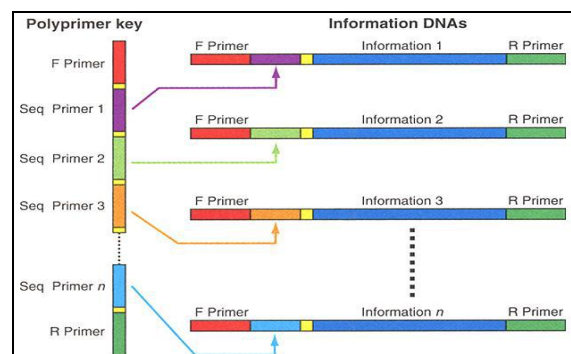

Figure 2.   Structure of DNA molecules used for data storage.

The next research [4] was published in 2003, regarding the retrieval of data stored on DNA which was successfully conducted by Wong et al. They advocated that DNA strands can break at both ends, which is fatal to its structure which can lead to loss of information. To prevent that, a dependable medium to store and protect the encoded strands from harsh conditions and addition of synthesised gene sequences was required. For that purpose, they needed a vector which contains the DNA with the data and is able to grow and multiply to ensure permanence of information. The team used Escherichia coli and Deinococcus radiodurans as vectors due to their rapid regeneration rate and tolerance towards vacuum, radiation (Ultraviolet which is otherwise harmful to humans even in one tenth of that amount). They used the oligonucleotide sequences as a digital arrangement similar to the 1's and 0's that make up the ASCII scheme for text representation in silicon based devices. Next, they identified those 'safe sequences' which are foreign to both bacteria but don't cause unnecessary protein production which can harm the safety of the encoded message (found 25 out of ten billion sequences). Then they created the complimentary strands (twenty bases long), used restriction enzymes to insert the desired encoded fragments and finally cloned the double strands into a recombinant plasmid. The stop codons act as 'sentinels' to protect the message. PCR is used to extract the data whenever needed. This way they were able to procure 7 chemically synthesized DNA fragments with 57–99 base pairs (bp) of foreign encoded information in the bacteria. This research was fundamentally important in the view that it showcased the methods of protection of the desired data from extremities in environment, radiation, vacuum, nucleases otherwise harmful for the fragile DNA fragments. This experiment founded the idea that DNA could be ultimately used for archival science given the density it had to offer by storing the encoded data in a suitable tolerant host. Their encoding scheme and idea (of protecting the sequence) are shown in Fig. 3.
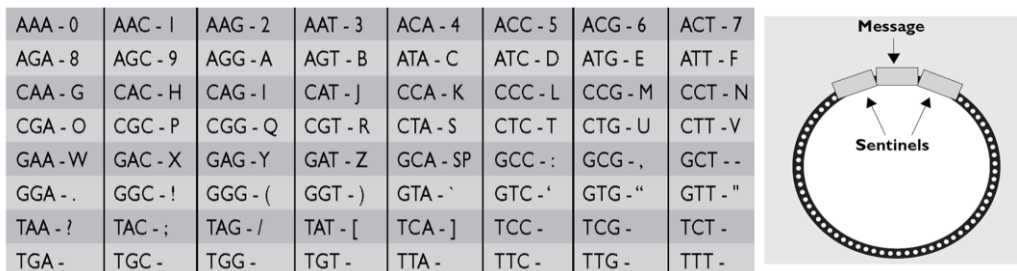

Figure 3. Encoding scheme.

III. ENCODING SCHEMES

The following six years (2003-2009) majorly focused on the methods of encoding DNA. Till 2009 digital data was being encoded in the form of base triplets (since encoding was not the main focus then, proving the feasibility of the technique was the primary aim). So the interim period widely ranged on the different encoding schemes which could be used. Applying Combinatorics, Linguistics, Logic, Mathematics there emerged a need to form a universal scheme of communication (between the DNA bases and other languages) which is robust, free from semantic barrier, is representative of all possible data we need to store and also has further scope to expand to fit new formats of data. Simultaneously it should not compromise on the conventional reading technologies while simultaneously maintain an optimum processing time and occupy less memory. In brief a proper and organized coding scheme was needed. Due to the breakthroughs obtained in the field of DNA based cryptography, various encoding schemes have been tested and successfully implemented in several data storage applications. These include the Huffman Code, Comma code, The Alternating Code, The Three Base per Letter encryption key.

One of the prerequisites for a good DNA coding method is the judicious use of nucleotide bases per character. It has been mathematically proven that the base to character ratio of around three is most optimum and economical for a coding system. This is the reason why many researchers preferred (and still continue to do) the Huffman Coding Scheme [5]. It is a base 3 representation which is widely applied for lossless data compression in digital communication and data storage. As the coding scheme to encode messages into the base 4 DNA form, it had to be modified. Ailenberg and Rotstein [6] proposed an improvement over the Huffman Coding to facilitate storage of different types of digital media- text, pictures and audio characters in DNA. Using the principles of Huffman Code, they defined DNA codons for all the characters on the keyboard to prevent ambiguity while coding. They synthesized and arranged the plasmid library for storing and retrieving information by designing primers which are incorporated in the message segment with resembles the structure of an exon or intron which facilitates speedy, efficient, reliable retrieval of information. They used 7 bases for every 2 characters (ratio was nearly 3.5). This is precisely where Huffman Codes are better than their earlier counterparts. Other methods had higher ratios (higher bases required) which were uneconomical but they worked smoothly because they were representing only the English Alphabet .On the other hand, this team set their rules for encoding text, music and images.

A 2010 project from the Craig Venter Institute was able to encode a 7920-bit watermark in the genome

sequence of the bacterium Mycoplasma mycoides. Synthesizing artificial DNA is a prerequisite for researchers to be able to produce synthetic cells so that they can be clearly differentiated (watermarked) from natural cells and DNA. Although this project did not involve digital data storage, it was the largest project to date to encode a larger amount of information into DNA. Moreover, it was also the first time a completely synthetic cell was produced, so it was a significant achievement.

These were the several different data storage milestones till 2010.The scale of the work till 2010 was relatively smaller as compared to the two major research projects completed in the last two years which can be clearly called as successful breakthroughs in DNA Computing. Earlier, writing and decrypting long and perfect DNA sequences was tough, since the technology wasn't developed to that extent.

These projects moved from bits to megabytes of information, both a quantitative and a qualitative paradigm shift in data storage. Their methods are described in detail.

A team of three researchers Church, Gao and Kosuri published a milestone paper [7] on 16th August 2011. They first converted 11 JPG images, an html coded draft of an approximately 53,000 words book ("Regenesis"- by Church, and a JavaScript program into a 5.27 megabit bit stream (set of binary 0's and 1's). They used variety in their file selection to demonstrate the potential of DNA to store all forms of data, which were properly encoded using their innovative encoding scheme. They assigned 0 as A or C (Adenine and Cytosine) and 1 as T or G (Thymine and Guanine). This was due to the fact that a sequence like AAACTGG read in one direction (say 5'-3') would give the same result as that of CCAGTTT read in reverse (3'-5' direction) . The converted stream of 0's and 1's were encoded onto oligonucleotides, each containing a 96 bit data block. These bits were encoded onto 54,898 159-nt oligonucleotides (oligos). Since the length is huge, the sequence was chunked into 96 nucleotide pieces (corresponding to a 96-bit data block), where 19 nucleotides (bits) represented the location of the text in the book. The 22 nucleotides which were potentially common in their sequences were amplified and sequenced. Ink-jet printed DNA glass microchips were used to synthesize the oligonucleotide library. The book was read by amplifying the library by limited-cycle PCR and then sequenced on a single lane of commercially available sequencing technology Illumina HiSeq©. The recovery was 100% with an error of only 10 bit per 5.27 million which were located at the ends of the oligonucleotide. The error was due to the single sequence coverage. Such long stretches of DNA are particularly expensive to work with, so Church and his colleagues split the DNA sequence into short chunks that were each 96 bases long. Each chunk included a 19-bit bar code, or address, to show where that chunk belonged in the whole of the book. The DNA was synthesized, inkjet-printed on a glass DNA microchip, and then cleaved off and dried to form a 50-nanogram clump smaller than a speck of pollen. These methods had distinct advantages over the storage approaches described earlier. Instead of encoding two bit per base, the team used one bit per base (A, C =0 and T, G=1). An 'f' for example, which meant "01100110" in binary was encoded in DNA as "ATGAATTC". This enabled efficient reading of even difficult sequences (with high CG content or secondary structures and repeated sequences) since the complementary strand reading would contain the same information when read backward (due to this encoding scheme). Thus multiple way reading made the process faster than the previous experiments here, the entire process from encoding, amplifying, reading and retrieving took just 14 days, which is quite unexpected given the relatively large data involved. Secondly, the idea of splitting the set of 0's and 1's in the form of addressed data blocks made it quite analogous with the storage on silicon and optical hard drives, because it did away from the resource intensive task of assembling a single long DNA sequence. Several copies of individual oligonucleotides were synthesised and stored which did away with the extra task of cloning and recertifying the sequence. This method was feasible because a correct copy would automatically correct errors in other copies as the errors in synthesis and sequencing are usually independent of each other. Furthermore, the entire procedure was done artificially in the in vitro and didn't involve any living organism (like the in vivo approach used in data storage in vector described previously). Thus no cloning or instability issues were involved. Ultimately, they used the State of the art Next-generation technologies in synthesis and sequencing which facilitated the entire process to consume 105 times less cost compared to first-generation encodings. The unique aspect of this work is that the density and scale of information stored using this method is comparably better than other experimental as well as available commercial data storage techniques.

The advantage of the technique resulting from this work is that a general approach of addressing data blocks, synthesizing and sequencing them would be compatible with future technologies too. The Fig. 4 and Fig. 5 show this simple principle as well as the comparison of this work with available as well experimented technologies.
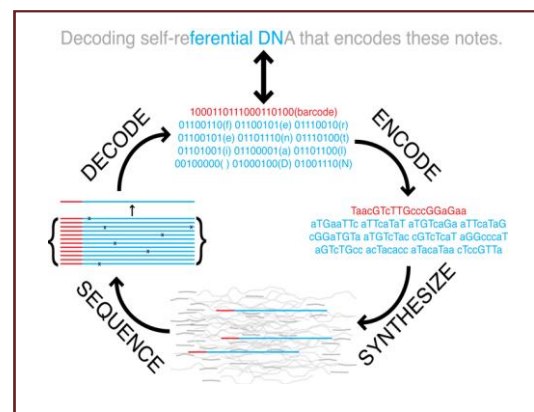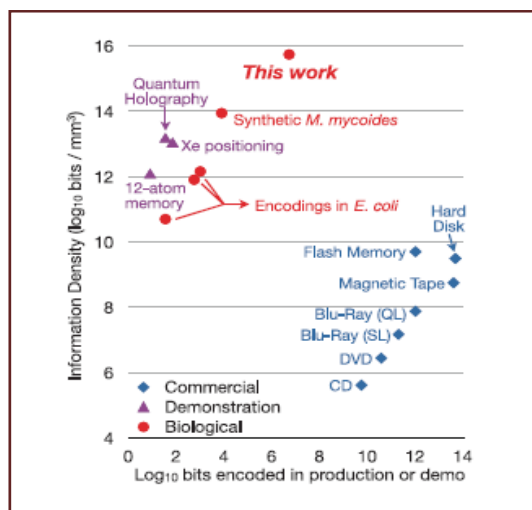

Figure 4.   Overall process.

Figure 5.   Observations plotted on graph for the experiments by church et al

DNA-Data Storage witnessed yet another milestone on 23 January 2013, when the exemplary work of Nick Goldman and his team at the European Bioinformatics Institute was published in Nature. Their aim was to create an information system which is feasible, has a lot of capacity and which requires lesser maintenance than present storage media. Their breakthrough was to create a scalable method which could ultimately become cost-effective within a decade and be an appropriate medium for archival science. They encoded around 740 kilobytes of hard-disk stored digital data to a DNA code, synthesised it, sequenced it and retrieved it completely with proper genuine reconstruction techniques. Their data comprised of Shakespeare's Sonnets in ASCII format, Watson and Crick's 1954 Classic Paper, a medium resolution colour photograph, a snippet of the audio from Martin Luther King's famous speech in MP3 format and a Huffman code (again in ASCII text scheme) for an aggregate data size of 757,051 bytes. Their method is briefly described here. The efficiency of their Encoding Scheme was 88%. It was found out that as the data size increases, the decrease in encoding efficiency is very slow owing to unavailability of technology which can synthesise longer fragments to maintain the index as well as data (70% for data storage on petabyte (PB, 1015 bytes) scales and 65% on exabyte (EB, 1018 bytes) scale). Larger data means that each base is read a lesser number of times thus reducing the sequencing coverage while the growth rate of errors being slower than the data size. In spite of the several errors predicted through this experiment, it is firmly ensured that DNA-based storage schemes could be scaled beyond present world data volumes to become practical mediums for large-scale, long-term and sequentially accessed digital archiving.

The coding scheme used in this project, projected efficiency for real-time solutions and the comparison of this method with that of previous years with regard to the number of bits encoded and decoded was very complex.

IV. CHALLENGES

Considering all these major findings, it is inevitable that DNA would become a universal archival medium one day. But it presents several challenges, some due to its own physical composition, while some due to our technological ineptness to unleash its full potential at present.

The overall process of encoding, amplifying, sequencing, restructuring and decoding takes significantly more time than their conventional counterparts. According to Cox [8], "Assume reading the sequence at enzymatic rates (say 150 nucleotides per second), the retrieval process would still be six orders of magnitude slower than that of a personal computer" (which can read data from the hard drive at nearly 100 Megabits per second). Consequently, DNA is unlikely to compete with optical, magnetic or quantum formats in the foreseeable future [9]-[11].

Many types of errors are associated with the current machines dealing with DNA. For instance Presence of Homo polymers, sequencing errors, error due to lower access rate are some examples. Though DNA in living cells have auto correction enzymes, no such artificial enzymes exist for artificial DNA [12]. DNA Strings need to be discarded if the decoding scheme is inefficient, thus leading to a loss of data and consumption of more DNA to ensure the same theoretical completeness. Due to its structure, it is prone to mutations in extreme conditions, thus the data might get altered in a mutation. It is a base 4 storage device, so it is fundamentally inefficient since the best storage and lossless compression occurs for base 3 (Huffman Encoding). Another major challenge for practical DNA-based information storage is the difficulty of synthesizing [13] long sequences of DNA de novo (simulating on the computer) to a specified design.

Even with insignificant computational costs and adequate use of the technologies current costs are estimated to be \$12,500 per MB for information storage in DNA and \$220 per MB for decoding information while that of conventional hard-disk is 8.21 cents (as of 2010).

V. CONCLUSIONS

It is clear that data storage in DNA is no more confined to science fiction but is being realized and improvised at very promising rates by research teams all over the world. This idea has received positive criticism from the general public as can be inferred from their responses on the different science websites.

Similar to all revolutions in technology, DNA-based data storage technology has to face major challenges to realize its full potential. It is however, inevitable that DNA would be invariably used for archival purposes [14] for its sheer density, robustness, stability and energy efficiency. In theory, grams of DNA can store all the information ever produced by mankind. Several breakthroughs will be required before it becomes commercially mainstream for data retrieval.

This field has had a million-fold improvement in the recent years. Digital Data Storage in DNA technology

shows immense progress, since reading and writing it is advancing ten times every year unlike the Electronic Technology which is improving roughly 1.5 times a year (Moore's Law).

Cost efficiency also shows promise as the exponential drop in DNA synthesis and sequencing cost has been five-fold and twelve-fold respectively while Electronic media show only a 1.6-fold drop per year. As handheld sequencers are becoming commercially available, the research would grow beyond large projects, with more people getting access to experiment with this idea. "As DNA is the basis of life on Earth, methods for working with it, storing it and retrieving it will remain the subject of continual technological innovation", says Nick Goldman. In order to handle big data a practical large-scale DNA archive would need stable DNA management and innovative indexing solutions. This would create the desired paradigm shift in computing as the aspect of Data Storage would be an integral part to realize the idea of DNA Computing. Reciprocally, it would catalyze research and development in synthesis and sequencing technologies.

Future work could include compression schemes; dealing with redundancy at all levels, checking for parity, correcting errors to enhance density and safety. DNA could also be substituted with polymers [15] or be modified to suit the needs of digital storage. Furthermore, it will fuel research to look for alternative materials for information storage and to aid in realizing the need for a universal medium for data. Overall, this technology is here to stay and could transform the way we have ever looked at DNA and computing as totally different entities.

REFERENCES

[1] N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E. M. Le Proust, B. Sipos, and E. Birney, "Towards practical, high-capacity, low-maintenance 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, "Long-term 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 minimum-redundancy 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-writing-a-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.

Siddhant Shrivastava, is a sophomore at BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, Pilani, India pursuing B.E.(Hons) in Computer Science and Engineering. He attended Junior School and High School at ST. JOSEPH'S ACADEMY, Dehradun. Being born in Mumbai, and brought up in different parts of India, he has inculcated cosmopolitan views about his environment. His hobbies are broad and varied which include Photography, watching Science Documentaries, Graphics Designing, Reading works and biographies of Scientists and listening to Classical Music. One day, he aims to contribute to the field of "Nanotechnology" using Artificial Intelligence and Robotics. Currently he is endeavoring to develop and propose alternative sources of Digital Data Storage and Studying Agent Communication Protocols. He proactively believes that Scientific Research should be as sound and complete as possible using interdisciplinary expertise. He is an active member of the Technical Team, BITSEmbryo, has a sound knowledge of Video Conferencing Infrastructure and has handled several popular talks before a packed lecture theatre.

Rohan Badlani, is a sophomore at BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, Pilani, India, pursuing B.E.(Hons) in Computer Science and Engineering. He did his schooling from DELHI PUBLIC SCHOOL, MATHURA ROAD, Delhi, and R.N. PODAR SCHOOL, Mumbai. Born in Baroda, and raised all over India – he has wide variety of interests. His hobbies include playing Lawn Tennis, reading books, swimming and social work. He is a blue belt in Karate. Currently, he is working towards developing efficient data storage devices and an optimized architecture for network security devices. He believes in meaningful scientific research for the betterment of mankind. His leadership qualities include organizing events for IEEE BITS Pilani Student chapter and ACM BITS Pilani. He is also a member at Web Team, Embryo Club, BITS Pilani.