International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 9 Issue 2 ǁ 2021 ǁ PP. 68-72

DNA computing storage implementation

Nikita Sabne[1], Dr. Prakash Kene[2]
1. Student, Department of MCA,
2.Assistant Professor, Department of MCA, Modern College of Engineering, Pune, MS 411005, India

A B S T R AC T
Rapidly day-by-day growing technologies will need huge memory in future, biomolecular computing gives a modern scientific turn for memory storage, computational operations, parallel processing, extra-ordinary storage capacity. Using single nucleotide we are able to use 4 bases which composes the DNA to encode data into byte. Data can be encoded using two-bit. DNA computing has been introduced instead of Silicon-based technology. For Example, 10 terabytes of computer data can be stored in 10 trillion DNA molecule which can fit in one cubic centimetre.

KEYWORDS*: DNA Computing, Bio-Molecular Computing, Molecular Computing.*

Date of Submission: 10-02-2021 Date of acceptance: 24-02-2021

I. INTRODUCTION

Mikhail Naiman, 1964 published first idea on using synthetic DNA to record, store and retrieve digital information. Joe Devis, 1988 put first Artificial data storage design and synthesized an 18-bp message and transformed into E-coli. DNA is Double Helix Structure having two strands and four bases that is Adenine (A), Guanine (G), Thymine (T), Cytosine (C) and Phosphate backbone. In which there is fix pairing that is Adenine always paired with Thymine (A-T) and Guanine always paired with Cytosine (G-C). The chemical structure of the bases allows an efficient formation of hydrogen bonds only between A and T or G and C; this determines the complementarily principle, also known as Watson-Crick base pairing of the DNA double helix. Encoding of data into nucleotides called genetic coding. DNA present in cell that cell contains nucleus, nucleus contains chromosome and that chromosome contains DNA molecules. So basically DNA is a Nano particle we cannot see by our naked eyes. Deoxyribonucleic acid (DNA) is a molecular structure which provides genetic instructions for functionality of cells and development of living organisms. DNA undergoes replication in which two strands get separated and the hydrogen bond between bases breaks. Free nucleotides (A, T, G and C) attract to their complementary bases. Separated strands receives new nucleotides and they joined together with the help of enzyme. That is how two identical DNA strands forms and this process repeats continuously. Polymerase Chain Reaction allows us to produce many copies of a specific sequence of DNA. PCR is an iterative process that cycles through a series of copying events using an enzyme called polymerase. Polymerase will copy a section of single stranded DNA starting at the position of a primer.

DNA storage need static environment very low or high thermal energy and UV energy can damage it and error can occur. In case of error at one strand, that error can resolve with the help of enzyme because in double stranded DNA every complementary strand is mirror of another strand. DNA solution/molecules in test tube looks like drop of water but if we see that one drop contains trillions of molecules and it can store extremely huge data. We can perform computer operations in molecule.

A	**G**	**C**	**C**	**T**	**G**	**A**
T	**C**	**G**	**G**	**A**	**C**	**T**

From the above example, we have **AGCCTGA** so it's complementary s' will be **TCGGACT** i.e. Adenine- Thymine and Guanine-Cytosine. Hence we can substitute 0's and 1's for A, T, G, C that is two-bit encoding.

DNA has computational potential to solve mathematical problems like the directed Hamilton Path problem also Known as the "travelling salesman problem". Logic Gates are a vital part of how our computers carries out functions that we command it to do. These gates convert binary code moving through the computer into a series of signals that the computer uses to perform operations. Currently logic gates interpret input signals from silicon transistors and convert those signals into an output signal that allows the computer to perform complex functions. DNA logic gates are the first step towards creating a computer that has a structure similar to that of an electronic PC. Instead of using electrical signals to perform logical operations, these DNA logic gates

rely on DNA code. They detect fragments of genetic material as input, splice together these fragments and form a single output. These logic gates might be combined with DNA microchips to create a breakthrough in DNA computing. Researches in DNA Computing composed of enzymes and DNA molecules instead of silicon microchips

DNA, with its unique data structure and ability to perform many parallel operations, allows us to look at a computational problem from a different point of view. Silicon-based computers typically handle operations in a sequential manner. Although multi-processor computers and modern CPUs incorporate some parallel processing. Computer instructions are handled extremely fast but basically sequentially. DNA computers has a different way from ordinary computers. Typically, increasing performance of silicon computing means faster clock cycles where the emphasis is on the speed of the CPU and not on the size of the memory. For DNA computing, though, the power comes from the memory capacity and parallel processing. For example, if we consider the read and write rates of DNA, DNA can be replicated at a rate of about 500 base pairs a second. Biologically this is quite fast and assuming low error rate, an impressive achievement. However, this is corresponds to only 1000 bits/sec, which is quite slow when compared to the data throughput of an average hard drive.

DNA Parallelism

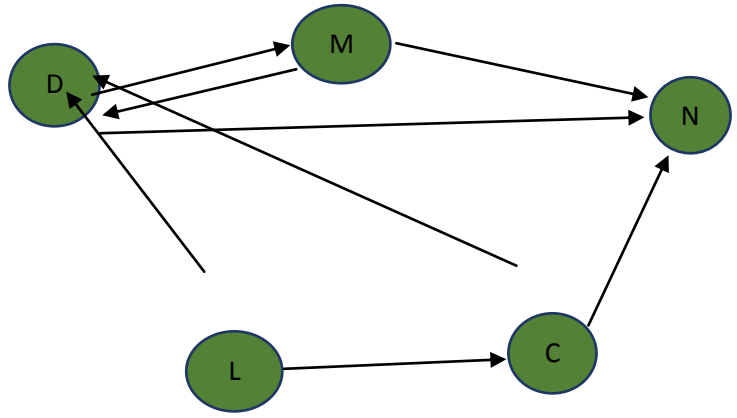
DNA can work in a massively parallel fashion. The operations like addition, logical operators, bit-shifting, also allows complex-calculations. It has capacity that many copies of enzyme can simultaneously work on many DNA molecules. There are variety of operational proteins which manipulates DNA and variety of enzymes which are tiny protein machines that read and processes DNA according to design. Test tube is collection of DNA molecules which can perform operations like cut, paste, copy, repair and also perform operations between two different test tubes t1 and t2 like:

1. Append-tail (T, Q): it append Q onto the end of every strand in tube T.
2. Discard (T): it discards test tube T.
3. Copy (T1, T2): it contains Test tube T2 same contents as T1.
4. Detect (T): in given test true output ‘yes’ if T contains at least one strand.
5. Merge (T1, T2): stores union of two test tubes T1 U T2 in T1 and leave T2 empty.

Adleman’s Experiment

Adleman used to solve directed Hamiltonian Path problem using the DNA technology. He first generated all the possible paths and then selected the correct path. Specifically, the method has the following steps

1. Generate all possible routes.
2. Select paths that start with the proper city and end with the final city.
3. Select paths with the correct number of cities.
4. Select paths that contain each city only once.



Encode city names in short DNA sequences, and then encode the paths by connecting the city sequences for which routes exist. DNA can simply be treated as a string of data. L-GCTACG C-CTAGTA D-TCGTAC M-CTACGG N-ATGCCG the entire path can be encoded by simply stringing together these DNA sequences represents specific cities. This route is L - C – D - M - N. DNA that starts with L and ends with N by using the Polymerase Chain Reaction. To accomplish this we can use technique called Polymerase Chain Reaction (PCR). A primer is a short piece of DNA complimentary to one end of a section of the DNA. By selecting primers at the start and end of the section of DNA we want to amplify the DNA between these primers, doubling the amount of DNA containing this sequence. Sort the DNA by length and select the DNA whose

length corresponds to 5 cities. To accomplish this Adleman used a technique called Gel Electrophoresis (GE), which is a procedure used to resolve the size of DNA. Gel Electrophoresis is to force DNA through a gel matrix by using an electric field. DNA is a negatively charged molecule under most conditions, so if placed in an electric field it will be attracted to the positive potential. Since the charge density of DNA is constant long pieces of DNA move as fast as short pieces when suspended in a fluid. The gel is made up of a polymer that forms a meshwork of linked strands. In the end the test tube will contain all paths that each starts at L and ends at N, and which have a total of exactly 5 cities encoded. DNA containing a specific sequence can be purified from a sample of mixed DNA by a technique called affinity purification. This is accomplished by attaching the compliment of the sequence in question to a substance like a magnetic bead, for the first run we use L'-beads to filter out DNA sequences which contain the encoding for L, the next run we use D'-beads, and then C'-beads, M'-beads, and finally N'-beads. The order isn't important. If a path is missing a city, then it will not be filtered out during one of the runs and will be removed from the pool.

Adleman's experiment has not really solved a problem, but it has shown that DNA and the procedures mentioned above can be used to solve a mathematical problem.

Experiment on Algorithm to find out the Polypurines

Two copies (3' and central) of an important region of the genome, termed the Polypurines tract (PPT), are re-factory to RNase H Cleavage. They are left intact so they can be used as primers in the synthesis of the second DNA strand. DNA is composed of four nucleotides, also called bases: adenosine (A), cytidine(C), guanosine (G), and thymidine (T), each of which consists of a phosphate group, a sugar (deoxyribose), and a nucleobase (pyrimidine – thymine and cytosine, or purine – adenine and guanine). The nucleotides are covalently linked through the sugar (deoxyribose) and phosphate residue and form the backbone of one DNA strand. These two different elements (sugar and the phosphate group). DNA the end with 3' with hydroxyl group of the deoxyribose, and the other end is 5' with the phosphate group. Two single DNA strands assemble into a double-stranded DNA molecule, which is stabilized by hydrogen bonds between the nucleotides. The two single DNA strands are complementarily aligned in a reverse direction: the one, called also a leading strand, has a 5' to 3' orientation, whereas the complementary strand, called lagging strand, is in the reverse 3' to 5' orientation.

- First create a file consist of randomly sequenced DNA molecules
- Through the program verify that the file is exist or not, that file will acts as the database for the Polypurines.
- Then enter the sequence of "A" and "G". Then we will iterate to find out the sequence of "A" and "G", If found then it will print the sequences and count the number of Polypurines.

Experiment on Data Storage in DNA

First we took a picture, we had it as digital file it was JPEG. We turn it into a string of binary. We broke it into different blocks. Information gets translated into pool of words, which are pool of DNA sequence that represent it. Then we combine the right piece, the right building block together, and the next thing we add is an enzyme. And this is a biological that then takes the pieces of DNA and stick them together so form a molecule. This molecule would encode a piece of information, a piece of the picture. On the machine we transfer each of those assembled identifier and put them in one big pool and now we have all our data stored in one tiny little test tube. Once we have all the DNA written, it just look like translucent droplet. But it actually contains millions if not billions if not trillions of copies of your information. If down the road you want to look at your data again, you can take your DNA out of storage and put it through a sequencing machine, which just read back the DNA molecules. It is kind of dictionary of sort for DNA, where every word has an inherent order to it. Then computer can read it back into a JPEG from the binary

Using single nucleotide. In this way, we are able to use 4 bases that compose the DNA strand to encode each byte of data. We can simply encode any data by using two-bit to Nucleotide conversion.

Two bits	nucleotides
00	**A (Adenine)**
10	**G (Guanine)**
01	**C (Cytosine)**
11	**T (Thymine)**

APPLICATION

1. By developing algorithms in programming languages we can find out the Polypurines which are continuous sequence of purines in a protein.
2. The regions of DNA lying between genes may be powerful triggers for diseases and may hold the key for potential cures.
3. Fusion of DNA Computing and Artificial Intelligence could results into an expert system, lead to development of Face Recognition Systems.
4. DNA can help in secure transmission of huge amount of data.

ADVANTAGES

1. The DNA computers come with a great parallel processing capability. This feature has the potential to speed up those polynomial time problems that demand relatively less operations.
2. The performance of DNA strands is allowed to increase exponentially by performing millions of operations simultaneously.
3. DNA computers are light weight.
4. Power electricity required by the DNA computer are very less, power is needed only to prevent DNA from denaturation
5. DNA computers would be able to solve the hardest problems faster than the traditional ones.
6. The DNA computers come with a greater ability to hold a tremendous amount of information by using very small space. The DNA molecules only need just a single one cubic Nano meter for storing a single bit of information. This proves that the DNA computers have clearly greater advantages over the traditional computers.
7. For example, 10 terabytes of computer data can be stored in 10 trillion DNA molecule which can fit in one cubic centimetre.

LIMITATIONS

1. Generating solution sets, even for some relatively simple problems, may require impractically large amounts of memory (lots and lots of DNA strands are required).
2. DNA storage needs static environment, in case of increasing of thermal energy or UV sun-light we could loss the data.
3. DNA computer take much time to solve simple problem compared to traditional silicon computers.
4. Sometimes there may be error in the pairing of nucleotides present in the DNA strands.

II. RESULT AND CONCLUSION

In the travelling salesman problem, or "TSP" for short, a hypothetical salesman tries to find a route through a set of cities so that he visits each city only once. As the number of cities increases, the problem becomes more difficult until its solution is beyond analytical analysis altogether, at which point it requires brute force search methods. TSPs with a large number of cities quickly become computationally expensive, making them impractical to solve on even the latest super-computer.

I designed this algorithm in "C" language. On applying this algorithm the filtered sequences of "A" and "G" can be found along with the total Polypurines found. It helps in finding the large sequences of molecules in a database and gets the result. In place of text files we can take some other databases for storing large sequences and issues related to pattern matching can be resolved. DNA has one million times the data density of SSDs with a shelf-life of thousands of years. Massively-parallel DNA-based computers that make trillions of truly simultaneous calculations each moment. It just look like translucent droplet. But it actually contains millions if not billions if not trillions of copies of our information.

Considering all the attention that DNA has garnered, it is not hard to imagine that one day we might have the tools and talent to produce a small integrated desktop machine that uses DNA as a computing substrate along with set of designer enzymes. This field has got many obstacles and drawbacks. It will not be used to do things that traditional computers are good at, but it might be used in the study of fields such as encryption, genetic algorithms, language systems, and more. This area of study, called forensic biotechnology uses a method called DNA fingerprinting. DNA Computing devices could revolutionize the pharmaceutical and biomedical fields. Fusion of DNA Computing and Artificial Intelligence could results into an expert system

REFERENCES

[1]. Adleman, L. (1994) "Molecular computation of solutions to combinatorial problems". Science 266,
[2]. Lee, J.Y., Shin, S.-Y., Park, T.H., Zhang, B.-T. (2004). "Solving traveling salesman problems with DNA molecules encoding numerical values". BioSystems 78.
[3]. Amos M, Paun G, Rozenberg G, Salomaa A(2002) Topics in the Theory of DNA Computing, J. Theoretical Computer Science, 287

[4]. Beielstein T, Ewald T-P,Markon S (2003) Optimal elevator group control by evolution strategies, In Cantu-Paz E, Foster JA (ed) Genetic and Evolution Computation-Gecco 2003, Proc. Of Genetic and Evolutionary Computation, Conf. July 12-16, 2003, Chicago, Springer-Verlag, Berlin.
[5]. Ito Y, Fukusaki E (2004) DNA as a 'Nanomaterial', J Molecular Catalysis B: Enzymatic 28.
[6]. G. Gamow, A. Rich, and M. Y˘cas. The problem of information transfer from the nucleic acids to proteins. Adv. Biol. Med. Physics, 4.
[7]. M. Ogihara and A. Ray. DNA-based parallel computation by counting. In H. Rubin and D. H. Wood, editors, DNA Based Computers II
[8]. J. Sambrook, E. F. Fritsch, and T. Maniatis. Molecular Cloning: a Laboratory Manual. Cold Spring Harbor Press, NY, 2nd edition, 1989.
[9]. J. Reif. Parallel molecular computation. In Proc. 7th Symp. on Parallel Algorithms and Architecture.
[10]. Bornholt J, Lopez R, Carmean DM, Ceze L, Seelig G, Strauss K. A dna-based archival storage system.
[11]. Choi Y, Ryu T, Lee AC, Choi H, Lee H, Park J, Song S-H, Kim S, Kim H, Park W, et al. High information capacity dna-based data storage with augmented encoding characters using degenerate bases.