International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 4 22 - 26

DNA Computing and Implementations

Vijay Prakash Tiwari
Department of Computer Engineering
Bharati Vidyapeeth's COE,
Lavale, Pune, MH, INDIA
vijayptiwari28@gmail.com

Vikas Tiwari
Department of Computer Engineering
Bharati Vidyapeeth's COE,
Lavale, Pune, Maharashtra, India
vikastiwari.raghav@gmail.com

Prof. U. C. Patkar, HOD
Department of Computer Engineering
Bharati Vidyapeeth's COE,
Lavale, Pune, Maharashtra, India
patkarudyac@gmail.com

***Abstract*—** DNA Computing aims to harness the molecules at the Nano level for computational purpose. DNA Computing features high data density and massive storage capability therefore, its approach can be used to solve various combinatorial Problems like solving Non Deterministic Problems (i.e. NP- Complete and NP-Hard). This Molecular Level Computational involve input and output both in the molecule form. Since DNA has already been explored as an exquisite material and is a fundamental block for manufacturing large scale Nano mechanical devices. DNA Computing is an approach towards the Biomolecular Computation where the aim is not only to process the information but also to transfer it to other molecular structures for utilization. DNA Computing is slower when an individual DNA Computes in compare to silica based chips. Its Efficiently and throughput increases as the number of DNA increase. DNA provides the possibility of massive parallelism. Starting with the Introduction about the DNA Structure, followed up by DNA Computers, this paper will discuss some recent advancements and challenges of DNA Computing. We will also discuss the possible future scope and implementation as well how the Artificial Intelligence approach can be used with DNA Based Computers to achieve a better and efficient Machine Learning.

Keywords- *DNA, DNA Computers, Biomolecular Computing, Genetic Algorithm, Swarm Intelligence*

*****

I. INTRODUCTION

Nature has always been headmaster of all learning that we humans have. DNA (Deoxyribonucleic Acid) Computation is one of the inspiration from the nature. It is inspired by the way DNA works. It works just like the theoretical Turing Machine [5]. Turing machine can be formally defined as a 7 – tuple, M = (Q, Γ, b, Σ, δ, q0, F) where

1. Q is a Non-Empty, Finite set of states.
2. Γ is a Non-Empty, Finite set of Tape Alphabet Symbols.
3. b Є Γ is the Blank Symbol
4. Σ Є Γ \ {b} is the set of Input Symbols.
5. δ: (Q \ F) x Γ → Q x Γ x {L, R} is the Transition Function.
6. q0 is the Initial State.
7. F is the Final State.

Neurosciences has enabled the development in the field of biomolecule information processing like genetic regulation, Cellular Signal, Morphogenesis, Protein Trafficking and Evolution. The first demonstration involving the explanation of importance of information processing in evolution was addressed in Aldeman's Experiment in 1994 [1][2]. The experiment demonstrated the laboratory molecule biology can be enough to have a programmed computation on a DNA. The DNA Computing involves two types of approaches. First, Purely Chemical, known as, *vitro*. Second, Inside Cellular Life Forms, known as *vivo*.

DNA has ability to save massive information and have very low energy dissipation which eventually leads to immense interest in massive DNA based Computation.

Starting with the introduction of DNA computation, followed up by the how DNA computing have been applied to solve huge and complex mathematical problems like complex combinatorial problems. Currently, Molecular Computing is a highly anticipated topic of research not only within the context of solving Non Deterministic Problems i.e. NP-complete an NP-Hard Problem, but also implementing it on existing silica era of Digital Computers.

II. DNA STRUCTURE

DNA molecule is the Biological Storage Device which by default stores the Genetic Code of the living creature. This information can be manipulated by using various types of enzymes and nucleic acid interaction. Today, the information is saved in long strings or sequences of 0s and 1s similarly in DNA all the information is stored in the long sequences of bases called Nucleotides. There are four types of Nucleotides, Adenines (A), Guanine (G), Cytosine (C) and Thymine (T)

According to chemical convention, each strands has a 5' and a 3' end. Therefore, every DNA strand has its natural orientation. Figure 1 Elaborates the Chemical Convention and its Orientation of DNA Molecule.

Each Nucleotide is located 0.35 nm apart in DNA Molecule. This compactness gives the data density of 1 bit per cubic nm and a potential of 10^{18} bit or an Exabyte in just 1-gram DNA. These four bases or Nucleotides follow connecting rules of 'Watson-Crick Complementarity' where A can only connect with T while G can only connect with C. These connecting properties leads to Double-Strand Structure or Helix Structure of DNA. Therefore, each DNA Sequence has its natural compliment.

Figure 1: Example of DNA Molecular Orientation

III. DNA COMPUTERS

DNA Computers are the Biological Process Driven Computers that has been the most efficient computing platform. In this, the data is represented in the form DNA strands. In Spite of being DNA Computation as a slow

computation technique than silicon based computation, DNA Computing provide the massive parallelism. This advantage of DNA Computing can be used to solve Non Deterministic Mathematical Problems (i.e. NP-Complete or NP-Hard Problems). [2][3]

It is very common to have $6x10^{16}$ DNA Molecules per ml for a Biological Experiment. If we assume that each DNA molecule expresses one character than we can realize the $6x10^{16}$ bits or 60,000 TB storage capacity or memory. DNA Computation involves DNA Molecule based procedure (i.e. DNA Molecule Reactions) therefore execution time of a DNA Molecule is high than traditional Electronic Computation but due to massive parallelism the total execution speed exceptionally high.

The Important thing to note is that, computation is done using DNA not on DNA. In DNA Computing the chemical and physical properties of DNA bases is used to achieve the output. Just like every computer, DNA has also two important components. First, Enzymes as Processing Unit. Enzymes perform Denature (Cutting), Replication (Copying) and Anneal (Pasting). Second, DNA itself works as Storage Unit.

Since ages Researchers are proposing various Miniature Computer that are aimed to harness the storing and parallel processing capability of DNA Molecule. Models used for such Molecular Computing is known as Automatons. Earlier Automatons utilizes the ATP (Adenosine Triphosphate), which was the common source of energy for Cellular Level Reaction. Later on after 2003, Benenson designed a new model which provides two things, firstly the initial value or initial data and secondly, sufficient energy input to work on and compute the result [6]. For Automatons, two state input are given in the form of string. To answer a that particular question, Automatons uses predefined rules. Then for Cutting an input state molecule at particular position and releasing the energy stored in the bonds, enzymes are used as a computer hardware. The released heat energy is then used for power generation.

When Position Operation Control is embedded properly with Molecular Tools like Enzymes, can result in efficient way of building a general purpose and robust programmable device. Von Neumann depicted the model for Self-Replicating System as shown in Figure 2. The only thing with the Self-Replicating System is that these systems must have reasonable and acceptable complexity. The purpose of building a Self-Replicating System in context of manufacturing is just to have low-cost replication of a flexible and programmable manufacturing system.

DNA tiles, macromolecular building block used for bottom-up nanoscale construction of patterned structure which provide a programmable approach towards nanostructures. DNA tiles have sticky ends. DNA tiling lattices, is a process of attaching this sticky end of DNA tiles with one another to have a long structure assembly [7].

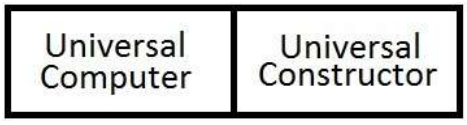

Figure 2: Von Neumann Architecture of Self-Replicating System

IV. DNA COMPUTNG

DNA Computing is a fascinating platform which enables development of computing technologies by interfacing Computer Science and Molecular Biology. This technology has evolved not only as an information processing technology, but also as a catalyst that can be used for transferring processed information between nanotech and biology. DNA Computing is a type of Biomolecular Computing where input and output all are in molecular form.

Biomolecular Computers are molecular sized, programmable and autonomous Computers in which the input, output, software and hardware, all are made up of biological molecules [8]. Biomolecular Computer uses biological process to compute a problem. Biomolecule Computing utilizes naturally occurring chemical and physical properties of the biomolecule.

The main steps involved in DNA Computing are as follows:

a) Melting: Breaking the weak hydrogen bonds in a double helix to form two DNA strands which are complement to each other.
b) Annealing: Reconnecting the hydrogen bonds between complementary DNA strands
c) Merging: Mixing two test tubes with many DNA molecules.
d) Amplification: DNA replication to make many copies of the original DNA molecules.
e) Selection: Elimination of errors (e.g. mutations) and selection of correct DNA molecules.

Since DNA behave like Turing Machine therefore it can work as Data Storage Device. Usually 8-20 base pairs are used to represent bits. Various methods have been introduced to manipulate these bits. Following are some of the most used techniques which can be used as computational operators for copying, sorting, splitting and concatenating the information embedded in a DNA Molecule [1][2]

a) Ligation
b) Hybridization
c) PCR (Polymerase Chain Reaction)
d) Gel Electrophoresis
e) Enzyme Reaction

A. Encoding

Encoding is the main challenge in the DNA Computational Procedure. Designing the correct DNA Sequence is essential in order to have an optimal solution.

- Ligation and Hybridization: When DNA Sequences are dropped into the test tube, DNA Sequences taken, started to recombine with one another by enzyme reaction. This process is known as 'Ligation'. DNA Sequences and their corresponding complements are mixed together in one test tube. Then for hybridization process, the DNA Mixture taken I the test tube is heated up to 95^0 C and cooled to 20^0 C with 1^0 C per minute. Then the reaction is subjected to Ligation. After the completion of the process, a particular DNA Sequence will ligate together with another DNA Sequence just to produce a new sequence.
- PCR (Polymerase Chain Reaction): Polymerases can be used to repair and duplicate the DNA. It is a

method to amplify the DNA in *vitro.* This process quickly amplifies the amount of a specific DNA molecule in a given solution by using primer extension. Each cycle doubles the quantity of this molecule. Following are the process are involved

1. Initialization: A mix solution of template, primer, dNTP and enzyme is heated to 94^0 C – 98^0 C for 1 – 9 Minutes to ensure that most of the DNA templates and primers are denatured.
2. Denaturation: Heat the solution to 94^0 C – 98^0 C for 20 – 30 seconds for separation of DNA Duplexes.
3. Annealing: Lower the temperature in between the 50^0 C – 64^0 C for 20 – 30 seconds for primers to anneal specifically to the ssDNA template.
4. Elongation: Raise the temperature to optimal elongation temperature of about 70^0 C – 74^0 C, for the polymerase add dNTP from the direction of 5' to 3' that are complementary to the template.

- Gel Electrophoresis*:* It is a technique of sorting the DNA Strands on basis of their size [4]. DNA Molecules have basically negative charge. Therefore, when placed in Magnetic Field, they tend to migrate towards the positive pole. Since DNA Molecule has the uniform distribution of charge per unit length there for the force applied is same for all the DNA Strands but due the length variation therefore DNA Strand of a particular length tends to move only to limited distance from its pace to positive pole. Hence we separate the DNA Strands on the basis of their size.

B. Solving Problems using DNA Computing

- Finite State Problems: To compete with the already existing silicon based Computing Technology, Biomolecular Computation Technology has to be fast, robust, and reliable. Therefore, DNA Computing has the advantage of using its most important property that it supports massive parallelism. Guarnien and Bancroft in 1999 developed a DNA based addition algorithm that uses successive primer extension reactions to implement the carries and the Boolean logics was required in Binary Addition [9]. The same method was used to show the subtraction. But unfortunately there comes some limitations to it. Firstly, they were unable to utilize the massive parallelism capabilities of biomolecular computer. Therefore, at a time only 2 numbers can be added together. Secondly, the outputs were encoded distinctly from input, therefore repetitive operations were not allowed.
- Combinatorial Problems: DNA Computing operations and its procedure was employed in solving various Mathematical Complex Combinatorial Problems like Non-Deterministic Problems. The first successful demonstration was done in 1994's experiment known as Adleman's Experiment. Adleman solved the HPP (Hamiltonian Path Problem) successfully [1][2]. Then DNA Computing was used to solve the Chess Problems. The advantage of these approaches of massive parallelism that DNA Computing provides.

The Combinatorial Problem that was solved by Adleman was a simple instant of most popular problem HPP (Hamiltonian Path Problem) also known as TSP (Travelling Salesman Problem). In this, Adleman first assigned a genetic code to every city. For Example, 'DELHI' might be coded as 'GCCACGT' or 'PUNE' might be coded as 'CCATAGG' [1][2]. If two cities are to be connected, then the genetic sequence will have first 3 letters of one city and last 3 latter of genetic sequence of another city. For Example, connecting 'DELHI' to PUNE' will result in the connection between of 'GCC' from 'DELHI' with 'AGG' from 'PUNE'. The TSP may seem a simple puzzle but the most Advanced

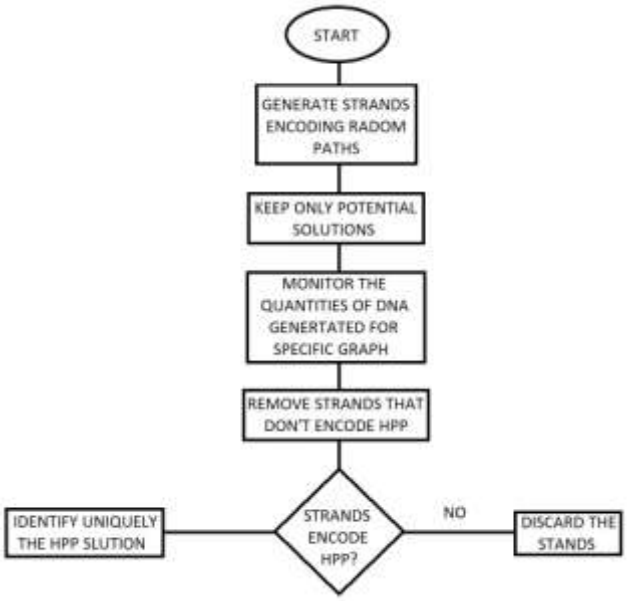

Figure 3: Adleman's Approach to HPP

Supercomputer would take years to calculate the optimal solution of a TSP with 50 – 100 cities. In Adleman Experiment, Adleman solved this problem with 7 cities within second. He assigned every city with their unique Single-Stranded DNA Molecule having unique Genetic Code. Each Stand was 20 nucleotides long and only 10 nucleotides are used connect when there exists a path as a solution between two cities. When the DNA Strands is mixed with ATP (Adenosine Triphosphate) and DNA ligase, this results in the generation of all possible path through cities [1][2]. Since all paths are not solution to the problem. Therefore, Adleman filtered all such undesired paths. Remaining Molecules are the solution of the HPP.

The DNA Sequence Generation happens exponentially every second. They are set to replicate and create trillions of new sequences in just matter of seconds (DNA Hybridization). Based on the Adleman's Experiment, of there is about 300 cities than by repetitive generation of various DNA Sequences, TSP would take probably the amount of

DNA that weighs equally to the mother Earth. Hence the error rate at each iteration is a big hurdle that need to be overcome.

C. *Classification of DNA Computing*

DNA Computing is currently classified in three categories that are Intermolecular, Intramolecular and Supramolecular.

- Intermolecular DNA Computing: Focusing on Hybridization of different DNA Molecules as a basic step of Computation.
- Intramolecular DNA Computing: Construction of programmable state machines in single DNA Molecules which operate by means of intramolecular transitions.
- Supramolecular DNA Computing: Focusing on the Self-Assembly of rigid DNA Molecules with Different sequences to perform Computation.

V. INTELLIGENT SYSTEMS

A. *Smart DNA Chip*

Smart DNA Chips are applied to the prediction and diagnosis of Cancer so that Health Analysis in direction of early detection/ diagnosis of cancer can be done. To accurately classify the Cancer, one have select genes related to cancer because genes, extracted from DNA Chips are noisy [10]. Alone DNA Computing approach was not enough therefore researchers included various Machine Learning approach. Hence Combining this both approaches, researchers were able to explore various features and classifiers several benchmark datasets to systematically evaluate the machine learning classifiers such as m-nearest neighbour and features of selection methods and support vector machine.

Intelligent DNA memory Chips are designed to utilize the massive data storage capacity of a DNA Molecule [11]. DNA has been proved as a most reliable and robust medium to save data even up to millions years. We Still can read genetic Code of millions years old living organism. The DNA Based Storage device need Cold Storage area for being preserved. The DNA Memory incorporates intelligent processing and reasoning capabilities. The high data storage capability and massive parallelism support capability of DNA Molecule is used for drawing inferences on the entire in *vitro* knowledge bases. The DNA Chip is called intelligent because it is not only able to detect the gene patterns and expressions but also the logical formulae for them.

B. *AI (Arificial intelligence) in DNA Computing*

Whenever the pool of DNA Molecules is used to solve the NP-Complete problem like the HPP, the pool of molecular data sequences increases at every iteration exponentially. The more the number of variable, bigger the pool of molecular data we have. Most of the molecules create in this brute-force solution finding process are undesired sequences or are not the solution to the problems. As the problem size increases, this brute force solution find approach becomes infeasible. Therefore, AI (Artificial Intelligence) can be used to break the barrier of this brute force-attack and find the optimal solution in just a small initial data pool. By this we can avoid all the enumerating candidate solution. We can achieve this by two possible ways. Firstly, by using GA (Genetic Algorithm) or Secondly, by introducing Swarm Intelligence.

- GA (Genetic Algorithm): The Genetic Algorithm works on principle of evolution and natural selection. GA uses famous Darwin's Principle of natural selection. Evolution is the phenomenon that obtains the adaption through the interplay of selection and diversity. Low-Fit Molecules are ruled out at every iteration of the brute-force solution finding approach. In the end we are left with the solution with very small data pool. The Genetic Algorithm manipulates the population of but strings by using crossover and mutation operators. The success comes from the advantage of having massive parallelism and very densely data storage capability of DNA when directed with the search capability of Genetic Algorithms. 1 gram of Single-Stranded DNA Molecules is equivalent to 10^{22} bytes.
- Swarm Intelligence: This is an another approach that can be used to break the brute-force barrier of DNA Computing. Swarm Intelligence is inspired by the collective intelligence in social animals. These animals require no leader. Their interaction among individuals decides their collaborative behaviour. Such process is known as Self-Organization. Here, each individual is not intelligent but when they are observed together we find a collaborative intelligence. There exist, 3 main types of Swarm Intelligence Techniques that are Models of Bird Flocking, Ant Colony Optimization (ACO) and The Particle Swarm Optimization (PSO) Algorithm. Earlier PSO was to designed to understand more about the human social behaviour by simulating bird flocking. Perceptive Particle Swarm Optimization (PPSO) Algorithm is an advancement of PSO. This advancement not only considers the social interaction but also the environmental interaction of the creature. It was designed to handle the real-world problem of programming and controlling the physical world agents of nanotechnology and DNA Computers.

VI. CONCLUSION AND FUTURE SCOPE

DNA Computers are the new edge technology that has evolved from ages and we are still years away from understanding it completely. Electronic Computers theoretically and practically can implement the parallelism only to a limited extent but Biomolecular Computing is the next generation computing technology that can implement massive parallelism therefore standing above Electronic Computers in area of parallelism. The main advantage of DNA Computing is not only that it supports Massive Parallelism but also that it provides high density data storage capability.

DNA's data storing reliability can be understood by the scenario where we can still read the genetic code of the living organism that have extinct million years ago and can be replicated. The main benefit of DNA Computers is that it can solve the combinatorial problems much faster and efficient than any Silica Computer. DNA Computer provides energy efficient approach towards the computation. In spite of all these advantages and fascinating capabilities, DNA's

capabilities are still unleashed. There are tremendous scientific, engineering and technological challenges that need to be overcome to actually come up with one fully functional one packaged machine just like the current computers that a layman with little computer knowledge can use.

Industries could get transformed by manufacturing the suitable chemically balanced nanoscale materials for DNA Computing based Devices. The implementation of intelligent computing strategies like Genetic Algorithms in DNA Computing can provide and efficient alternative to the current computing approaches of supercomputers. The engineered biochemical circuits could transform the industries that utilizes the chemical and nanostructured materials.

There is a new multi-disciplinary branch, name Bioelectronics, that also uses Biomolecules in electronics and photonic devices. This branch aims to integrate the biomolecules with electronics to develop the broad range of bioelectrical devices that can revolutionize the medical, engineering and military working. Since Biomolecular Computation is powerful enough to replace the current Silica Based Electronic Computers but at the same time the costing of such advancement would be high enough. In nearby future DNA Computing will open up the doors towards the new research problems in mathematics like combinatorial, theoretically complex problems and also in intelligent system designing. The efficiency that a DNA Computing shows in finding the solution for the HPP, like in Adleman's Experiment, it is expected that in nearby future we would have much efficient Switching-Routing application of DNA Computing.

Important events that have been taken place in the field of DNA Computing has initiated the possibility of exploiting the massive parallelism, dense storage capability and nanostructures inherent in natural phenomena to solve computational problem. The core thing to understand is that, DNA is used as Storage Device but the chemical and biological properties of the nucleotides is used for computational purposes.

ACKNOWLEDGMENT

This survey is done by the group of 2 people which includes, Vijay Prakash Tiwari and Vikas Tiwari under the guidance of Prof. U. C. Patkar. We would like to thank each member of the group for their contribution in this survey. Special thanks to our Professor for his assistance and guidance in this survey.

REFERENCES

[1] Adleman LM (1994) Molecular computation of solutions to combinatorial Problems, *Science*, 266: 1021-1024.

[2] Adleman LM (1998) Computing with DNA, *Scientific American*, 279(2): 54-61.

[3] Lipton RJ (1995) DNA solution of hard computational problems, *Science*, 268 : 542-545.

[4] Amos M, Paun G, Rozenberg G, Salomaa A(2002) Topics in the Theory of DNA Computing, *J. Theoretical Computer Science*, 287(1): 3-38.

[5] McCaskill J 2000 Biomolecular computing (WWW document)http://www.ercim.org/publication/ErcimNews/enw43/mccaskill1.html (accessed 20th May 2005).

[6] Benenson Y, Adar R, Paz-Elizur T, Livneh Z and Shapiro E 2003 DNA molecule provides a computing machine with both data and fuel *Proc. Natl Acad. Sci. USA* **100** 2191–6.

[7] Reif J H, LaBean T H, Sahu S, Yan H and Yin P 2005 Design, simulation, and experimental demonstration of self-assembled DNA nanostructures and motors *Lecture Notes in Computer Science* vol 3566, pp 166–80.

[8] Benenson Y and Shapiro E 2004 Molecular computing machines *Encyclopedia of Nanoscience and Nanotechnology* ed J A Schwarz, C I Contescu and K Putyera Dekker (New York: Dekker) pp 2043–56.

[9] Guarnieri F and Bancroft C 1999 Use of a horizontal chain reaction for DNA-based addition *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* vol 4, pp 105–11.

[10] Cho S B andWon H H 2003 Machine learning in DNA microarray analysis for cancer classification *First Asia-Pacific Bioinformatics Conf. (Adelaide)* http://crpit.com/confpapers/CRPITV19Cho.pdf (accessed 20th october 2005).

[11] Chen J, Wang Y and Deaton R 2004 Large scale genomic monitoring or profiling using a DNA-based memory and microarrays *24th Army Science Conf. (Orlando, FL, 2004)* (WWW document) http://www.asc2004.com/Manuscript/SessionA/AP-06.pdf (accessed 27th May 2005).