Hasan et al LGURJCSIT 2019
LGURJCSIT
ISSN: 2519-7991
LGU Research Jounral for
Computer Sciences & IT
Vol. 3 Issue 1, January - March 2019

Future Computing: DNA Hard Drives

Muhammad Zulkifl Hasan[1], Muhammad Zunnurain Hussain[2], Zaka Ullah[3] ,Taimoor Hassan[4]
[1]Department of Computer Science
[1]Lahore Garrison University, Lahore, Pakistan, [2]Bahria University Lahore Campus
zulkifl.hasan@lgu.edu.pk, Zunnurain.bulc@bahria.edu.pk,
zakaullah@lgu.edu.pk, taimoorhassan@lgu.edu.pk

Abstract:

DNA is a hard drive, the memory in every cell of every living organism that has the instructions on how to make that cell. With the exponential growth of our media, the growth of data storage is also increasing. DNA is incredible compact due to its molecular structure and can be used to achieve data attractively through genome sequencing. A raw limit of 1 Exabyte/mm3 (109 GB/mm3) having half-life of over 500 years. With this, all the data of the world can be stored in just a small place of a room.

Key Words: Molecular computing: DNA, data storage.

1. INTRODUCTION

The idea of storing data into the DNA started out as a jock on 16 February 2011 by Nick Goldman, a group leader at the European Bioinformatics Institute (EBI). In 2013, scientist proved the they can write computer data into synthetic DNA with zero errors. DNA is a digital storage medium, it is a sequence of discrete alphabets of four letters and data can be stored by manipulating it. DNA is A, T, C and G but the machine language is 1's and 0's. To convert binary code to DNA code, algorithms were used. Although DNA pairs A with T and G with C, it can also be reversed means T with A and C with G. all these relations are different from each other making dense data in less space. Whereas, hard drives, cds and flash drives have a small shelf life. Tape backups usually used by special data centers need special climate control facilities with constant maintenance. Meanwhile DNA can survive with minimal effort. Paper and science showed a single gram of DNA can store 215 petabytes of data. All the internet traffic in the whole world in 2016 is equal to 1.1 zettabytes and it can be stored in 1 gram of DNA. In 2013 scientists encoded few 100 kilobytes of data in a DNA.

DNA based data storage is emerging future technology for storing data having features such as high density, durability, efficient in replication. The storage uses high throughput sequencing mechanism for maintaining user data. The encryption for DNA storage will change the digital data into nucleotide sequences into DNA. The process through which we change the encrypted data into its original shape is through decryption .

2. Information storage

In order to store information into the DNA, a code was invented that could store any kind of information in the digital codes just like in the computer or mobiles. Long pieces of DNA are not possible to make as humans so the scientists invented a code that could make a message out of the small fragments of DNA. Scientists also had to device a code that could store any kind of information and the problems and error that might Accor in writing and reading of that information, just like digital transmission in TV's or mobile phones that have error corrections in them .

3. CRISPR TECHNOLOGY

CRISPR is a technology that can edit or

add a new DNA anywhere in a genome with precision and store information in it. This technology also lets scientists to store data on the DNA of a living organism. By using a technology 'CRISPR' a renowned geneticist George Church hijacked the genomes of E. coli bacteria to store the information. CRISPR is actually a viral defense system in bacteria and it requires usual sequences from varices. In order to put constrains on the sequences, scientist had them contain information and the information they used were images like:

- Static images
- Moving images

Which were delivered to the living bacteria over time. Then scientist was able to sequence those bacteria and reconstruct the images. This was done by splitting the pixels of the image and then encoding those pixels in the DNA through CRISPR mechanism .

4. WRITING PROCESS ON DNA

The strands of the DNA are usually stored in pool's having stochastic spatial organization (non-structured addressing). The data entered is in the form of a key (.txt) and a value (in 1's and 0's) to obtain PCR primer sequence and finding the pool in a DNA where information is to be stored. The data is translated into DNA language and is then stored onto the DNA .

5. READING PROCESS FROM DNA

To read the data from the pool in the DNA on which the information is stored, the DNA language is first converted into digital or binary numbers using DNA sequences. The identification of the DNA molecules is associated with the key .

6. DNA DATA COPY

The data on the genome can also be copied by using laboratory equipment and grow exponentially many copies

7. DIGITAL MONEY STORAGE

Money is the digital information now a days. A bit coin is a form of money that only exist on computers and cryptography. This money can be stored in the DNA .

8. Methodology

DNA Computing is already proven powerful for technical problems. Parallel nature of biomolecular operations gives this power to DNA computing. This would be much more powerful and general purpose when combined with other existing well-known algorithmic solutions that exist for conventional computing architectures using a suitable ALU. Thus, a specifically designed DNA ALU that can address operations suitable for both domains can mitigate the gap between these two. The novelties of the proposed method as well as the logic and arithmetic operations for this DNA ALU proposed in this work may be summarized as follows.

Number of depictions: Executions are based on sticker-based DNA model, and hence can explore the power of the binary number representation .

Affluence of implementations: Operations applicable on sticker-based DNA model are simple to implement, except the clear operation. Proposed operations avoid the clear operation of sticker model, and are using the simple operations, and hence physical implementation is easier .

IO of the operations are of alike structure: The input operands are being taken as separate strands in separate test tubes, and output is also provided in separate test tube. Hence, the output of one operation can be used ready-made for any next operation, without any modification. Also, the uniform representation can be used, i.e., the way an user needs to represent the input, the output can be decoded in the user-readable format in the similar manner without any modification .

Opportunity of automation of DNA computer: If the basic operations of sticker model, i.e. combine, separate and set can be automated, the proposed implementations are highly suitable for the design of a fully automated complete ALU .

Plan of a complete DNA ALU: All logic operations, integer, and floating-point arithmetic operations have been implemented. Also, all representations follow conventional ideology, such as floating-point operations follow IEEE 754 floating point format and so on. So any algorithm or program logic that users can perform can be implemented directly on the

DNA computer without any modification .

9. CONCLUSION

Right now, no one is currently long term archiving digital information. Yet most information is now being created digitally but we cannot archive that information because some information are only in digital form like movies (3D, 2D etc.). when editors want to make an archive copy then they make an angular, celluloid copy but it is not long lasting. Whereas, DNA works for a long time over 500 years minimum. We can easily read data and make multiple copies of it. computer scientists are designing the best way to secure humanity's eve-increasing digital data collection and turning the molecular biology.

10. REFERENCES

[1] E. D. Green, E. M. Rubin, and M. V Olson, “The future of DNA sequencing,” Nat. News, vol. 550, no. 7675, p. 179, 2017.

[2] N. Mills, F. A. Feltus, and W. B. Ligon III, “Maximizing the performance of scientific data transfer by optimizing the interface between parallel file systems and advanced research networks,” Futur. Gener. Comput. Syst., vol. 79, pp. 190–198, 2018.

[3] S. Zhao, K. Watrous, C. Zhang, and B. Zhang, “Cloud computing for next-generation sequencing data analysis,” in Cloud Computing-Architecture and Applications, IntechOpen, 2017.

[4] J. Bornholt, R. Lopez, D. M. Carmean, L. Ceze, G. Seelig, and K. Strauss, “A DNA-based archival storage system,” ACM SIGARCH Comput. Archit. News, vol. 44, no. 2, pp. 637–649, 2016.

[5] Y. Erlich and D. Zielinski, “DNA Fountain enables a robust and efficient storage architecture,” Science (80-.)., vol. 355, no. 6328, pp. 950–954, 2017.

[6] S. Selvi, M. Gobi, M. Kanchana, and S. F. Mary, “Hyper elliptic curve cryptography in multi cloud-security using DNA (genetic) techniques,” in 2017 International Conference on Computing Methodologies and Communication (ICCMC), 2017, pp. 934–939.

[7] S. Greengard, “Cracking the code on DNA storage.,” Commun. ACM, vol. 60, no. 7, pp. 16–18, 2017.

[8] M. Sarkar, P. Ghosal, and S. P. Mohanty, “Exploring the Feasibility of a DNA Computer: Design of an ALU Using Sticker-Based DNA Model,” IEEE Trans. Nanobioscience, vol. 16, no. 6, pp. 383–399, 2017.