**The Entire World in A Shoebox – DNA Storage system**

*By G. S. Harsha Anirudh, BE 3/4 CSE-2*

Everyday millions of pictures, videos, and other multimedia are stored on hard drives and secondary storages. It is estimated that around 600 TeraBytes of data are stored only on Facebook servers every day. Imagine how much data is stored on the servers of all companies, and how much data is stored in a year! The amount of data stored exceeds our imagination. The place required to store this humongous data requires warehouses of millions of square feet and with huge maintenance costs.

Scientists have come up with a solution: storing the data in human DNA. This idea of recording, storing and retrieving data from DNA was first given by Mikhail Neiman in 1964 in the *Radiotekhnika* journal. On August 16 2012, George Church and colleagues at Harvard University published a journal which described how DNA was encoded with digital information, like an HTML draft of a 53,400-word book, 11 JPEG files, and a JavaScript program. They showed that 5.5 Petabits can be stored in a cubic millimetre of DNA. They used a very simple code for retrieval which is usually quite error-prone. This was the very first instance which proved that DNA can be used in place of traditional storage devices.

DNA data storage uses a base sequence of nucleic acids. It uses oligonucleotide synthesis machines for storage and DNA sequencing machines for retrieval. DNA is much more compact and is perennial compared to hard drives, and also has resistance to obsolescence. Tapes use 0’s and 1’s to store data, where the nucleotides adenine (A), cytosine (C),guanine (G) and thymine (T) constitute a strand whose combination is used for storing the data. It is much better than tapes in terms of longevity because of the three-dimensional packing in DNA, which isn’t possible in transistors because of heat issues.

The major drawback here is that the read latency is much higher than the write latency, as the entire pool must be sequenced to read a small block of data. Selective DNA amplification with polymerase chain reaction (PCR) technique can be used to access a block of data, but the cost of using this cutting edge technology is too high. Recent research shows that data can last centuries when kept in suitable environment, and all of the data present in the world today can be packed in a single kilogram of DNA.

