How can DNA be used to store data?

DNA DATA STORAGE

Why

We have more data than we know what to do with. Over 90% of all data that has been created throughout the history of humanity has been created over the last 2 years. The world's population is creating 2.5 quintillion (10^{18}) bytes of data every day, and every person will soon be generating 1.7 MB of data each second of their lives. And that's just humans. The earth is generating orders of magnitude more data than that every minute.

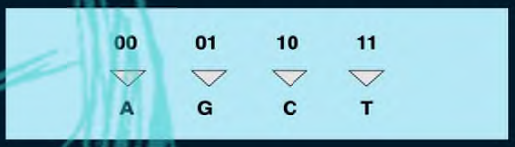
The speed at which we make data is outpacing our ability to store, transport, and access it. This is problematic because managing this massive increase requires energy (the IT industry burns between 7–12% of global energy every year) and physical space. It's slightly counterintuitive, but digital storage takes up space and energy, too. Saving to the cloud requires enormous warehouses of servers that consume massive amounts of energy to maintain at the right temperature and humidity.

DNA solves these problems by providing a more condensed form of data storage.

What

Digital data is stored in binary (strings of 1s and 0s). The basic unit of digital data is the bit (b), and eight bits make up a byte (B). Alphanumeric characters are coded in bytes, typically one per character (ASCII character set). Alternatively, images are videos are stored using a binary code representing each pixel in the media. Since videos are a collection of image frames, they require a lot more storage space. Files can be quite large (even can be measured in Terabytes).

DNA, however, is made up of four base components: Adenine, Guanine, Cytosine and Thymine (known as AGCT). A base is a part of a monomer called a nucleotide and many can be synthetically concatenated to create DNA strands. Each strand can only hold 20 bytes each. Therefore, digital data needs to be broken down into smaller pieces, each with an indicator (sequence number) of where in the sequence it falls into. This is similar to memory storage in a computer where logical addresses have to be converted to physical addresses as chunks of data are often not stored contiguously. The conversion from binary to DNA bases can be seen below.



DNA vs Current Technologies

Magnetic storage uses rigid rotating platters coated with ferrous (iron) particles- which are polarised to become a north or south state (representing a 1 or 0). The disk is divided into tracks into concentric circles. The disk spins at up to 10000 RPM. The driver head, similar to a record, moves across the disk, allowing data to be read from or written to it. This has large capacity but is easily damaged as due to its moving parts it is susceptible to wear and tear.

Optical Storage works by using a high-power laser to burn pits in to the surface of a CD, making it less reflective. The untouched areas jut out and are called lands. Reflective and non-reflective areas are read as 1s and 0s. The pits and lands are arranged as a tight spiral rather than circles. The smaller the wavelength used to create the pits, the smaller the pits so more data can be stored. This is why Blu-rays and standard CDs have such different sizes.

Solid state makes use of an array of chips on a board. They comprise of millions of NAND cells (NOT AND gate) present in digital circuits. Along with those there is a controller that manages the blocks of memory. Each cell works by delivering a current that activates the flow of electrons from the source to the drain. If the current is high enough it pushes a few electrons across an insulated oxide gate and they are trapped there. The state of a NAND cell is determined by measuring the charge in the gate. The presence absence of charge decides if it is a 0 or 1.

ADVANTAGES

+ **Making back-ups and copies is much easier -** Chemical reactions can make identical strands simultaneously rather than one by one. Reading process can destroy DNA but backup copies are easily available.

+**Very long lasting and durable** - DNA lasts for 1000s of years (seen by the fact that DNA is preserved from animals hundred of thousands of years ago) while traditional hard drives can get corrupted or wear out in less than 30 years

+**Large storage capacity**- DNA has a 1 million times greater information density compared to solid states so storing and transporting would require significantly fewer resources, reducing costs and is more environmentally friendly as there is a reduction in e-waste.

CHALLENGES

- **Slow and inefficient**- Building and reading DNA must require human input, and needs to be automated to become commonly used. Reading and building is relatively slow. DNA synthesis lets us write hundreds of bytes per second while a hard drive can write hundreds of millions of bytes per second.

- **Prone to error**- Reading and writing to DNA is prone to corruption. However, digital data doesn't need perfect accuracy like biology does so it is not a massive downside.

We should focus on long term storage of historical material so that it can be preserved for future generations rather than short term use. Instead of asking what we should save and preserve, we could preserve everything in DNA.