**Unlimited Data Storage thru DNA\***

**The Growing Storage Gap**

A global data sphere of 175 ZB (175 x 1021 bytes) projected by 2025 – five fold increase from 2018.

Kryder’s Law: storage densities doubled every 13 months since 1956.

Over the past decade though storage density growth slowed to paltry 16%, because of laws of physics. Some techniques may buy time but limit is thought to be 100 TB per square inch. Limits to solid state storage too.

World can only build so many storage facilities – energy to power them another limit. Mainstream storage also degrades over decades or even years.

**The Potential of DNA Storage**

DNA data storage uses the four bases of adenine (A), cytosine (C), guanine (G) and thymine (T) to form a quaternary (as opposed to binary system to represent data),

**Encoding**: pairs of binary digits are translated to corresponding bases, which are molecularly synthesized and appended to a DNA strand and then stored until needed.

**Decoding:** Upon retrieval, the DNA strand is sequenced and the detected bases decoded into binary digits. Note: The binary pair-base encoding method is the most basic encoding approach but more complex coding algorithms are used in practice.

**DNA Advantages:**

**Density:** Each base pair comprised of 29 or 30 atoms. Very compact and efficient structure. Work at Columbia University in New York densities of 25 PB/gram of DNA (PB is 1015 bytes). Work at University of Washington demonstrated file recovery equating to a density of 17,000 PB/g of DNA.

So 2025’s 175 ZB could be stored in just over 10 kg of DNA. This makes DNA storage more than ten million times denser than magnetic storage or solid state storage today.

**Durability:** In 2013 scientists sequenced the genome of a 700,000 year old horse from the Yukon. DNA was frozen in tundra in less than ideal conditions but could be fully sequenced. If DNA is kept dry and cool it has a shelf life measured n millennia. Today optical storage has a shelf life of 50 years, the longest available.

**Low Power:** Needs minimal energy compared to today’s technology.

**Maturation:** Death knells for Moore’s and Kryder’s laws are premature for the next decade at least. However However DNA synthesis and sequencing has surpassed Moore’s law, more than triple the growth in Moore’s law over the past decade. U of Washington holds record of storing/retrieving 1GB of data, in 2016 was 200 MB, a five fold increase in 3 years.

**Retrievability:** DNA synthesis/sequencing technology likely to last longer than storage technology of any given year (punch cards, 9 track, CD-ROM…).

**DNA Data Storage Challenges**

**Throughput:** Leading synthesis products do over a million synthesis operations in parallel yielding base pairs representing a MB per hour. Much slower than 9600 baud modems of 1980’s. Chemistry is slow – use parallelism.

**Noise:** Substantial synthesis and sequencing error rates. May need 20% error correction bits. Exacerbates throughput problem. But error correction reduce noise to acceptable levels.

**Cost:** Too expensive now, $3500 for 1 MB in 2017 study. Automation of steps people do and people don’t do is the key to lower costs.

**Random Access:** DNA storage is sequential, not random, like in early days of computing. Some early attempts at random access.

-------------------------------------------

\*M. Campbell, “DNA Data Storage: Automated DNA Synthesis and Sequencing are Key to Unlocking Virtually Unlimited Data Storage”, *Computer,* April 2020, 63-67.