FUTURE DATA STORAGE MEDIUMS FOR HIGH-CAPACITY SOLUTIONS

Data storage, storage density, future technology

Abstract

The amount of data produced by humanity each day is set on an exponential trend. As data production increases, the demand for data storage also grows. Current storage technologies cannot keep up with the extreme rate of growth, so new approaches to data storage must be considered. Quartz glass optical storage, holographic storage and DNA storage are three promising technologies that may become widely used in the future. The crucial factors necessary for these storage technologies to succeed are their storage density, transfer speed and commerciality. As of now, quartz glass optical storage leads the way as the most promising solution for large volume, low access data storage. However, research into DNA sequencing from other unrelated fields could make DNA archival storage more appealing than quartz glass in the longer run.

Theodore Glavas

April 18, 2021
Word Count: 2478

Introduction

As the product of the digital revolution, humanity collectively generates enormous amounts of data. The data created by companies, consumer devices, datacenters and more is known as the global datasphere, and its size was estimated to be 33 zettabytes (1 ZB = 10^{21} bytes) in 2018 [1]. By 2025, the International Data Corporation (IDC) predicts that the global datasphere will grow to 175 ZB [1]. The exponential rise in data production poses a serious question as to how the data will be stored. Innovation is required in the data storage industry to keep up with rising demand.

Although current storage technologies benefit from incremental improvements each year, they are not commercially viable in the long run [2]. Hard drives, which are still used for over half of all storage, are approaching the end of their product lifetime [2]. While recent advances in the storage industry have popularized high-speed mediums like NVMe SSDs, a large portion of demand is still for cold storage [3]. Cold storage refers to low cost, high-density storage used for keeping data that is not frequently accessed. This paper explores which emerging technology has the most potential to disrupt the cold storage market.

The criteria chosen to evaluate each technology are storage density, throughput and commerciality. Among the multitude of storage technologies in development today, quartz glass optical storage, holographic storage and DNA storage stand out as the most promising solutions to keep up with exponential demand for data storage.

Storage Technologies

The growing need for data storage has not gone unnoticed by large data corporations. Large scale data storage in private and public datacenters represents the majority of all data storage, creating a very large and profitable market for companies to compete in [1]. For this reason, many companies are investing into research for new data storage technologies. Among these initiatives is Microsoft's Project Silica, a research group investigating the use of quartz glass as an optical storage medium [4]. Quartz glass optical storage is a way to encode data on pieces of silica glass using specialized lasers and cameras to read and write information. The process and characteristics are well known. Femtosecond laser pulses directed at the glass to form tiny structures named voxels. These voxels can be made to vary in size and polarization, effectively creating a 5-dimensional storage medium: the 5 dimensions are obtained by encoding data in each voxel's X, Y and Z coordinates, as well as its size and polarization. Since the size of the voxels are only a few micrometers, a large amount of data can be encoded in a very small volume. Data can then be extracted from the glass using specialized optical cameras, which can measure each voxel's position, size and polarization. One downside of this technology is that once data is written to the medium, it cannot be erased or replaced. This is known as a write-once-read-many (WORM) medium [5].

One of the storage technologies competing with quartz glass optical storage is holographic storage. Research into holographic storage commenced in the 1960s and is still ongoing today [6]. The fundamental principle behind holographic storage is using light interference to encode data. In brief, data is first encoded inside a data beam. Then, a reference laser beam is made to interfere with the data beam inside a very thin sheet of photosensitive material called a hologram. The interfering beams leave a permanent 2-dimensional data pattern, called a page [7]. By using different reference beams and varying the orientation of the material, multiple pages of data can be superimposed on top of each other. This process is known as multiplexing, and it allows for large amounts of data to be encoded in a small volume [6]. To read the data, the reference beam is focused back onto the hologram. The beam is diffracted by the hologram in such a way that the original data beam is recreated and read by an optical sensor [6].

The final technology which holds promise in the future is DNA storage. The advances in DNA sequencing and manipulation from other fields have allowed researchers to encode data into DNA. This is done by using the four nucleobases of adenine, cytosine, guanine and thymine to represent data in a base-4 system [8]. Data is written using DNA synthesis, which is the process of creating artificial DNA with a specific arrangement of the four base pairs. The DNA can then be stored and read at a later date using DNA sequencing, which consists of reading the base pairs and extracting the encoded information [8, 9]. This process of writing and reading data from DNA is quite complex, but it enables large quantities of data to be stored in a very small volume.

Criteria

The first criterion investigated in this paper is storage density. Storage density refers to the amount of data (in bytes) that can be stored in a storage device per unit volume or per unit mass. Data density is critical in assessing the merit of a cold storage solution since datacenters need to fit as much data as possible in the limited space available to them.

The second criterion to be evaluated is throughput. Throughput refers to the speed at which data can be encoded into a storage device or decoded back, measured in bits per second. Throughput is not as crucial as data density for cold storage because the data being stored does not need to be accessed often. However, throughput is still an important factor to consider since it determines how long users must wait for data to be securely stored or extracted when it is needed.

The final criterion in this paper is commerciality. In this context, commerciality is an indication of how practical it would be to commercialize this technology and implement it on a world scale, assuming it has developed past the prototype stage. There are no empirical measurements for this criterion, but some important factors to consider are: lifespan of the medium, cost, energy consumption and equipment needed to operate.

Analysis

Storage density. Of the three technologies proposed, DNA storage can achieve the largest storage density. A paper from Erlich and Zielinski demonstrated an experimental storage density of 214 PB/g of DNA (equivalent to roughly 130PB/cm^3) by encoding and successfully recovering a series of digital files [10]. The theoretical maximum for DNA storage density is over 400,000 PB/g, leaving a lot of room for growth [11]. One of the main limiting factors to DNA storage density is the relatively high error rate which occurs during both DNA synthesis and sequencing. Error-correcting techniques are used to ensure that the data can still be read, but this comes at the cost of larger required file sizes and thus reduced storage density [8].

Quartz glass optical storage has yielded experimental storage densities of 132 TB/$inch^3$ (8.25 TB/cm^3), while some proposed methods are estimated to reach over 25TB/cm^3 [12, 13]. Despite these figures being a large improvement over today's technology (see Figure 1), it is still 4 orders of magnitude lower than what has been shown by DNA storage. Holographic storage has produced experimental storage densities of 50GB/cm^2, and methods for 750GB/cm^2 have been proposed [6, 7]. These densities were measured in terms of surface area rather than volume, because the holographic medium is extremely thin.

Figure 1 shows the relative storage density of these three technologies compared to the highest storage density drives available today. DNA storage is by far the highest in terms of storage density. Quartz glass optical storage still offers 2 order of magnitude improvement, while holographic storage is only one order of magnitude ahead of today's best commercial magnetic tape drives.

Figure 1

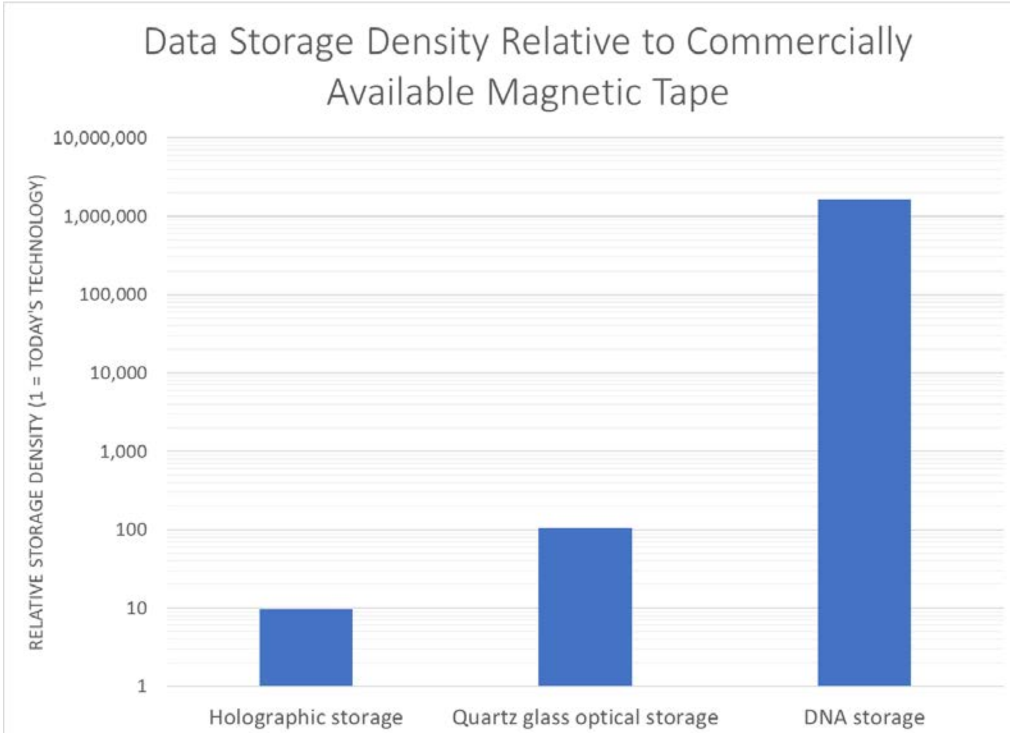

Figure 1 Data storage density of competing technologies (experimental) relative to leading magnetic tape archival drive (Quantum Ultrium 9) [14]

Throughput. When it comes to throughput, holographic storage has a significant edge over competing solutions due to its multiplexing properties. Data from an entire page, which can be composed of several million bytes, can be written or accessed simultaneously from the hologram [7], allowing Hoshizawa and his team to develop a holographic storage system capable of over 1Gbps throughput [7]. Quartz glass optical storage is also capable of simultaneous voxel reading using optical imagery, but it is not able to write data simultaneously. Each voxel must be written individually, resulting in slower write speeds compared to read speeds. However, the use of clever optics and light polarization techniques can mitigate this effect considerably [13]. Slower write speed is not of crucial importance for archival storage, since data only needs to be written once, and time is of higher importance when the data needs to be accessed. In contrast, DNA storage currently suffers from completely impractical read and write speeds. DNA synthesis is an incredibly complex process and still requires human intervention in most scenarios. Write speeds are in the order of 1MB/hour, which is severely limiting even for archival storage [8]. A lot of progress in DNA synthesis and sequencing needs to occur before the process can accelerate to a reasonable level.

In summary, holographic storage offers the most compelling throughput in both read and write speeds. Quartz glass optical storage offers competitive read speeds, but it falls slightly behind in write speeds compared to holographic storage. Finally, the throughput of DNA is currently too low to offer any practical use.

Commerciality. Holographic storage has many attributes that make relatively easy to commercialize. The materials needed to create the hologram are inexpensive, and stored data can last for more than 50 years without needing replacement [15]. Additionally, the technology requires no power when idle, and the optical equipment needed to read and write data is compact. In fact, a commercial holographic storage drive was launched by InPhase Technologies in 2008, demonstrating the ability to commercialize the technology. Unfortunately, the product did not catch on due to the rapid improvements in other storage technologies.

Quartz glass optical storage also offers promise to be adopted on a world scale. The quartz glass itself is extremely cheap to manufacture and has a seemingly unlimited lifespan if kept in controlled conditions [13]. This storage medium requires no power when idle but does consume significant power when writing data. The equipment needed to operate is relatively expensive and takes the space equivalent to a 2U server [5]. Although quartz glass optical storage might not be practical for small operations, it could still offer a competitive solution to larger datacenters. By integrating and optimizing the technology in large datacenters for mass storage, the cost of the technology can become very competitive compared to hard drives or magnetic tape, which need to be periodically replaced [5].

DNA storage is the most difficult of the three technologies to commercialize. The current cost to encode a single MB of data in DNA is in the thousands of dollars, and the equipment needed for both reading and writing is not scalable with today's technology [10]. However, once data is stored inside DNA, it is incredibly stable and can be stored with no power [8].

Holographic storage is the only technology that can be commercialized with today's technology. In the near future, quartz glass optical storage can be implemented in large-scale operations, while DNA storage needs several orders of magnitude of cost improvement before becoming competitive.

Discussion

There are some external factors to consider when looking at the three solutions holistically. Current archival storage technologies, which include hard drives and magnetic tape, are still making incremental improvements every few years. Holographic storage density was very compelling in the 1960s when it was first proposed, but the technology seems to have stagnated in recent years, allowing hard disk commercial technology to get very close to holographic storage in storage density. The significant technical challenges that need to be overcome to commercialize holographic storage significantly outweigh the marginal improvements the technology would bring to the storage industry.

Quartz glass optical storage also has significant challenges to overcome, but the potential rewards are much greater than holographic storage. Datacenters use tiered storage solutions, keeping the most frequently used data on high throughput storage, while a large portion of data is kept in cold storage. The use of quartz glass optical storage for the bottom tier of a datacenter can result in several orders of magnitude lower cost, making the investment today worthwhile. The prospect of having long term storage that never needs replacement provides enough incentive to companies like Microsoft to invest this technology.

DNA storage is in a unique position compared to other technologies. Unlike quartz glass and holographic storage, DNA storage does not rely on direct investments for the technology to advance. DNA sequencing and manipulation is at the heart of the fields of medicine, biology and others. Research on DNA from other topics can indirectly contribute to improving DNA data storage. For example, research into the human genome, genetically modified organisms (GMOs) and sequencing of DNA from fossils are all actively studying DNA and are developing better tools to work with it. Bornholt et al. summarized the inherent advantage of DNA storage by saying that "DNA-based storage [has] the benefit of eternal relevance: as long as there is DNA-based life, there will be strong reasons to read and manipulate DNA" [3, p. 637]. Even today, advances in DNA sequencing are surpassing Moore's Law [3]. Therefore, it is not impossible to conceive a future where DNA could be synthesized and sequenced several orders of magnitudes cheaper and faster than it is today, which would make DNA storage a viable candidate for the long term.

Conclusion

The rise in humanity's data production is creating the need for new storage solutions. As of now, none of the technologies discussed in this paper are ready to disrupt the data storage market. In the medium-near future, quartz glass optical storage leads the way as the most promising future solution for large volume, low access data storage. The technology strikes a good balance between high storage density, good throughput and low cost, giving it the highest likelihood of disrupting the market. Looking far into the future, research into DNA manipulation from other unrelated fields could make DNA storage a more appealing solution due to its unrivaled data density. Although holographic storage showed promise during its initial conception, current storage mediums are on pace to soon surpass it, making it the least likely to become mainstream.

This paper only considered some data storage technologies for the specific purpose of cold storage. The results in this paper came from laboratory conditions and theoretical calculations, which cannot be directly applied to commercial use cases. However, given the rapid pace of technological innovation and our unrelenting demand for more data storage, it seems inevitable that a new technology–quartz glass optical or otherwise–will come to replace what is mainstream today and propel humanity into the next age of digital information storage.

References

[1] D. R. J. G. J. Rydning, "The digitization of the world from edge to core," *Framingham: International Data Corporation,* 2018.

[2] W. A. Bhat, "Is a Data-Capacity Gap Inevitable in Big Data Storage?," *Computer,* vol. 51, pp. 54-62, 2018.

[3] J. Bornholt, R. Lopez, D. M. Carmean, L. Ceze, G. Seelig, and K. Strauss, "A DNA-Based Archival Storage System," presented at the Proceedings of the Twenty-First International Conference on Architectural Support for Programming Languages and Operating Systems, Atlanta, Georgia, USA, 2016. [Online]. Available: https://doi.org/10.1145/2872362.2872397 https://dl.acm.org/doi/pdf/10.1145/2872362.2872397.

[4] "Project Silica." Microsoft. https://www.microsoft.com/en-us/research/project/project-silica/ (accessed March 10, 2021).

[5] P. Anderson, "Glass: A new media for a new era?," (in eng), *10th USENIX Workshop on Hot Topics in Storage and File Systems, HotStorage 2018, co-located with USENIX ATC 2018,* 2018.

[6] W. Wilson, A. Hoskins, M. Ayres, A. Hill, and K. Curtis, "Introduction to Holographic Data Recording," in *Holographic Data Storage*, 2010, pp. 17-43.

[7] T. Hoshizawa, K.-i. Shimada, K. Fujita, and Y. Tada, "Practical angular-multiplexing holographic data storage system with 2 terabyte capacity and 1 gigabit transfer rate," *Japanese Journal of Applied Physics,* vol. 55, no. 9S, p. 09SA06, 2016.

[8] M. Campbell, "DNA Data Storage: Automated DNA Synthesis and Sequencing Are Key to Unlocking Virtually Unlimited Data Storage," (in eng), *Computer,* vol. 53, no. 4, pp. 63-7, 2020.

[9] C. K. Lim, S. Nirantar, W. S. Yew, and C. L. Poh, "Novel modalities in DNA data storage," *Trends in Biotechnology,* 2021.

[10] Y. Erlich and D. Zielinski, "DNA Fountain enables a robust and efficient storage architecture," *Science,* vol. 355, no. 6328, pp. 950-954, 2017. [Online]. Available: https://science.sciencemag.org/content/sci/355/6328/950.full.pdf.

[11] G. M. Church, Y. Gao, and S. Kosuri, "Next-Generation Digital Information Storage in DNA," *Science,* vol. 337, no. 6102, pp. 1628-1628, 2012, doi: 10.1126/science.1226355.

[12] Y. Lei, M. Sakakura, L. Wang, Y. Yu, H. Wang, and P. G. Kazansky, "Towards the smallest anisotropic structures by ultrafast laser writing in silica glass," in *Conference on Lasers and Electro-Optics*, Washington, DC, 2020/05/10 2020: Optical Society of America, in OSA Technical Digest, p. STh1H.1, doi: 10.1364/CLEO_SI.2020.STh1H.1. [Online]. Available: http://www.osapublishing.org/abstract.cfm?URI=CLEO_SI-2020-STh1H.1

[13] J. Zhang, A. Cerkauskaite, R. Drevinskas, A. Patel, M. Beresna, and P. G. Kazansky, "Eternal 5D data storage by ultrafast laser writing in glass," in *Laser-Based Micro-and Nanoprocessing X, February 16, 2016 - February 18, 2016*, San Francisco, CA, United states, 2016, vol. 9736: SPIE, in Proceedings of SPIE - The International Society for Optical Engineering, pp. The Society of Photo-Optical Instrumentation Engineers (SPIE), doi: 10.1117/12.2220600. [Online]. Available: http://dx.doi.org/10.1117/12.2220600

[14] "LTO Media." Quantum. https://www.quantum.com/en/products/tape-storage/lto-media/ (accessed March 10, 2021).

[15] K. Curtis, L. Dhar, A. Hill, W. Wilson, and M. Ayres, *Holographic data storage: from theory to practical systems*. John Wiley & Sons, 2011.