EXPLORING THE DEPTHS OF THE ARCHIVAL DATA STORAGE INDUSTRY

by

DIANA JONES

Submitted in partial fulfillment of the requirements for the degree of

Master of Science

Physics Department

CASE WESTERN RESERVE UNIVERSITY

May, 2024

We hereby approve the thesis/dissertation of

Diana Jones

Candidate for the degree of Master of Science

Committee Chair

Edward Caner

Committee Member

Robert W. Brown

Committee Member

Michael Martens

Data of Defense

April 30, 2024

****We also certify that written approval has been obtained for any proprietary material contained therein***

Table of Contents

SECTION PAGE

List of Figures

FIGURE PAGE

EXPLORING THE DEPTHS OF THE ARCHIVAL DATA STORAGE INDUSTRY

Abstract

by

DIANA JONES

Data storage is the practice of preserving data for long-term retention, such that it can be accessed and utilized at a later time. This field is experiencing rapid growth due to a surge in big data analytics, cloud computing, and internet of things. Innovations, such as advanced optical technologies, are being introduced to this space to mitigate the environmental effects and economic costs of preserving this data.

I. Preface.

During my internship tenure with Folio Photonics, a pioneering optical data storage startup founded by Dr. Kenneth Singer, Ambrose Swasey Professor of Physics at Case Western Reserve University, I embarked on a multifaceted journey beginning in June 2022 and concluding in November 2023. Commencing as a Sustainability Intern, my primary focus was to delve into the environmental advantages of the company's cutting-edge technology, culminating into an execution of a comprehensive Life Cycle Assessment. Transitioning into the Materials Development Group in September 2022, I assumed the role of Materials Engineering Intern where I resided for the remainder of my tenure. In this role, I examined the physical and optical properties of diverse materials, evaluating their alignment and optimizing their performance to meet the stringent requirements for our minimal viable product.

This paper primarily aims to comprehend the intricacies of data storage itself and the evolving market dynamics. Additionally, it explores how Folio Photonics' optical technology, the Datafilm Disc (DFD), aligns with incumbent optical technologies, while scrutinizing the DFD's features in comparison to other media technologies on the market.

II. Introduction to Data Storage.

Data.

Data is frequently hailed as "the central currency of science", signifying its role as the primary medium in which humanity exchanges information [1]. It manifests in diverse forms, spanning from handwritten notes in a notepad to expansive digital databases teeming with vast amounts of information to the streaming of movies online [1]. Data is increasing exponentially, which necessitates the need for cost-effective and energy-efficient data storage solutions, ensuring accessibility of data in the future.

Data Storage.

According to IBM, data storage encompasses the digital recording and preservation of information on a technological storage platform, allowing for future retrieval. [2] Data is physically stored on magnetic, optical, or mechanical hardware, and for this reason, it is imperative that data is backed up to safeguard against the potential loss or damage of that device, preventing the corruption of stored data [2].

The process of backing up data refers to the re-recording information on an alternative form of media in the event of an error like hardware failure, incorrect media configurations, and/or encryption issues [1]. To ensure data is backed up correctly, effective storage practices are essential for data retrieval because insufficient storage can lead to what is known as "data entropy," where data gradually becomes lost or corrupted over time [1]. To mitigate data entropy and guarantee future retrieval of data, it should be stored in open formats in its raw, unchanged form, eliminating the need for proprietary hardware or software for accessing it when necessary [1].

Types of Data Storage.

There are two general types of data storage, direct area storage (DAS) and network-based storage [2]. DAS refers to data storage that is in the immediate vicinity of the computing machine generating the data, such as the physical data storage medium housed within a laptop [2]. Network-based storage refers to data storage with off-site capabilities, meaning that the data is stored at an off-site location [2]. This type of data storage can be accessed by more than one computer within the network at the same time, for example cloud storage or “the cloud”, which is a collection of remote servers and computers [2, 3]. The cloud facilitates the storage and retrieval of data via the public internet and private networks [3].

Within network-based storage, there are two different types, network-attached storage (NAS) and storage area network (SAN) [2]. NAS is typically a data storage platform that is made up of redundant storage devices of the same type [2]. An example of this is RAID or redundant array of identical disks, in which discs of the same type are used [2]. SAN, on the other hand, is a platform made up of redundant storage devices of varying types [2].

Types of Data Storage Media.

Types of data storage media for direct-attached storage include optical disks like compact discs (CDs) or digital video discs (DVDs), hard disk drives (HDDs), solid-state drives (SSDs), flash drives, and magnetic diskettes [2]. NAS storage may use a mix of any of the above with the addition of magnetic tape, or may opt for a cloud approach to data storage [2].

For the purposes of this paper, optical discs, HDDs, SSDs, and flash are the primary focus of investigation in the data storage market.

Archival Data and Its Importance.

Data archiving refers to the preservation of digitized information in its rawest form, ensuring its reliability and integrity over time [4]. Raw data refers to data that is unaltered, meaning that it has not been processed or changed in any way from when it was collected [4]. Data archiving allows this kind of data to be accessed and retrieved at a later time [4].

The significance of data archiving stems from its crucial role in facilitating data analyses, thereby allowing conclusions to be drawn from the original data [4]. Drawing conclusions allows for informed-decision making [4]. Without the ability to archive data, it becomes difficult to re-visit, re-analyze, and further scrutinize the data to draw new or existing conclusions [4].

III. Market Dynamics: The Need for Data Storage

Market Size and Key Players

The global data storage market, according to a report by Fortune Business Insights, was valued at $186.75 billion in 2023 and is expected to reach $744.00 billion by 2032, with an expected CAGR of 17.1% [5]. Key players in the market storing new and existing data generation include IBM Corporation, Microsoft Azure, Hewlett Packard Enterprise (HPE), Equinix, Dell Technologies, Amazon Web Services (AWS), Hitachi Vantara, NetApp, Huawei, Veritas, and Infinidat [5].

Strengths: Demands for Data Storage

The demand for extensive and scalable storage solutions is on the rise due to two main factors. The first stems from the proliferation of big data analytics [6]. The global market for big data analytics reached $307.51 billion in 2023 and continues to expand rapidly with a growth rate of 13.0% through 2032. [7]. Businesses frequently rely on these analyses from big data analytics to generate insights into consumer behavior and patterns, correlations within a given market, and general market trends, allowing companies to make informed marketing and advertising decisions [6,7].

While big data analytics continues to rise, the real driver of the data storage market is cloud computing due to its projected growth rates in the coming years [8]. Cloud computing was valued at $587.78 billion in 2023 globally, but has projections to reach $2,291.59 billion by 2032 [8]. Cloud computing offers businesses an efficient and cost-effective method of storing data remotely without the need for on-site servers and storage media [8].

Weaknesses: Privacy Issues with Big Data Analytics

While there is a notable demand for data storage, there is a growing concern for privacy in this sector as companies seek to understand market insights on their consumers [9]. Much of the data used in big data analytics is collected passively through digital footprints of consumers and sensor-enabled objects [9]. A *digital footprint* refers to one's distinctive array of digital engagements, behaviors, and communications, which create a data trail on the internet or digital devices, enabling the identification of the specific user or device used [10]. This information is used to track and understand both individual and group consumer behavior [9]. Likewise, sensor-enabled devices like Internet of Things (IoT) often record and store personal information for the purpose of analyzing [11]. In data analytics, datasets are often amalgamated to gain deeper insights, but the amalgamation of datasets may re-identify specific individuals or groups, thereby jeopardizing their right to the privacy of their own choices [9].

Opportunities: Future Sources of Data Generation

The integration of Machine Learning, Artificial Intelligence, and the widespread adoption of Internet of Things presents significant opportunities for the data storage industry. Machine

Learning and Artificial Intelligence are transforming how computers learn and make decisions, generating a wealth of valuable data, thereby increasing the amount of data that needs to be stored [12]. They are anticipated to experience robust annual growth rates of 17.15% and 15.83% respectively from 2024 to 2030 [13, 14]. Simultaneously, the surge in Internet of Things adoption is driving a substantial need for large data storage systems to effectively manage the vast influx of information from interconnected devices [12]. This sector is seeing substantial growth, boasting an annual growth rate of 13.19% extending up to 2030 [15].

Threats: Potential Problems Utilizing The Cloud for Data Storage

Cloud computing is a distributed computing system of interconnected computers and servers, allowing users to remotely run applications and store data [16]. It offers on-demand, highly available services without the need for local hardware and software management [16]. Cloud computing is multi-tenancy in nature meaning it can serve many clients on one single platform [16]. Moreover, it involves storing data on multiple third-party servers instead of a dedicated server, with the specific storage locations typically unknown to users, known as virtualization [16].

The popularity of cloud computing has surged due to its cost-effectiveness and simplified resource administration, however, this system introduces risks to the data storage industry [16]. Because cloud computing utilizes an Infrastructure-as-a-Service (IaaS) model by offering a virtual, remote platform to store data, users are required to relinquish control over their data and applications to that cloud provider [16]. Because cloud computing is multi-tenancy, any external breach jeopardizes the data of all users utilizing it, leading to the potential loss or corruption of cloud users' data [16]. Private and public cloud computing companies must ensure the security and integrity, while maintaining availability of their clients' data [16].

End-users of Data Storage

End-users, specifically those generating, storing, and then later accessing that data, can be broken down into four categories, namely individual, commercial, industrial, and governmental [17]. Among these four segments, the commercial segment holds the largest market share as of 2022, and is expected to have the highest growth rate in the next decade [17]. The commercial sector primarily includes service businesses like restaurants, hospitals, and lodging facilities [18].

Storage Media in Data Storage Market

The data storage market can further be broken down into five different types of storage media. These include direct area storage, network-attached storage, storage area network, cloud storage, and software-defined storage, with DAS holding the biggest market share of 28% according to Fortune Business Insights [17]. However, given that SAN and NAS make up the category of the network-based storage, network based storage dominates over DAS with a 40% market share [17]. This means that the data storage market prefers using media with off-site capabilities as opposed to direct-attached media [17]. Figure 1 displays the breakdown of global market share of the various types of storage media utilized to store data long-term [17].

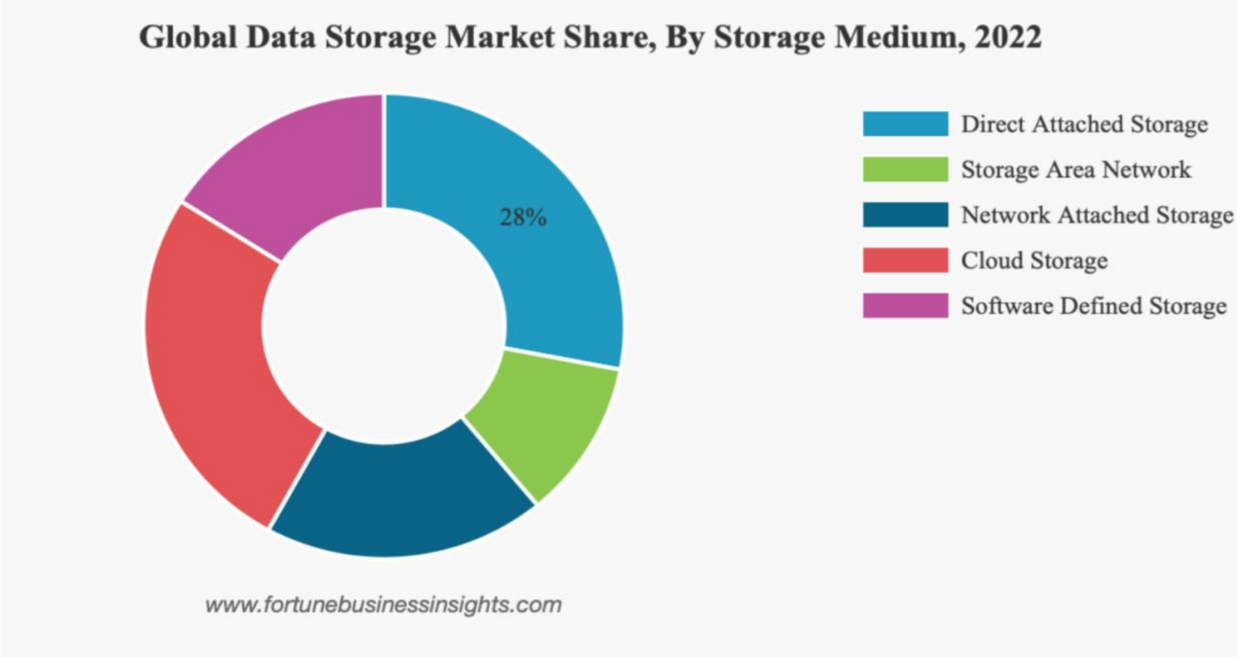

Fig. 1: Breakdown of the global data storage market by storage media as of 2022 [17]

Regional Developments in Data Storage

North America holds the largest market share, with the market size valued at $105.24 billion and expected annual growth rate of 14.3% from 2023 to 2030 [17]. This is because North America benefits from a strong technological infrastructure of data storage and a culture of innovation [17]. Fig. 2 displays the forecasted growth for data storage in North America as predicted by Business Fortune Insights [17]. North America's dominance is further supported by the presence of major industry players including IBM Corporation, Microsoft Azure, HPE, Equinix, Dell Technologies, and AWS and the continuous adoption of advanced storage solutions [17]. In Europe, the market is influenced by a focus on sustainability and data security, leading to an interest in energy-efficient and eco-friendly storage solutions [17]. The Asia-Pacific region is experiencing rapid growth, driven by the ongoing digital transformation and technological advancements in emerging economies [17].

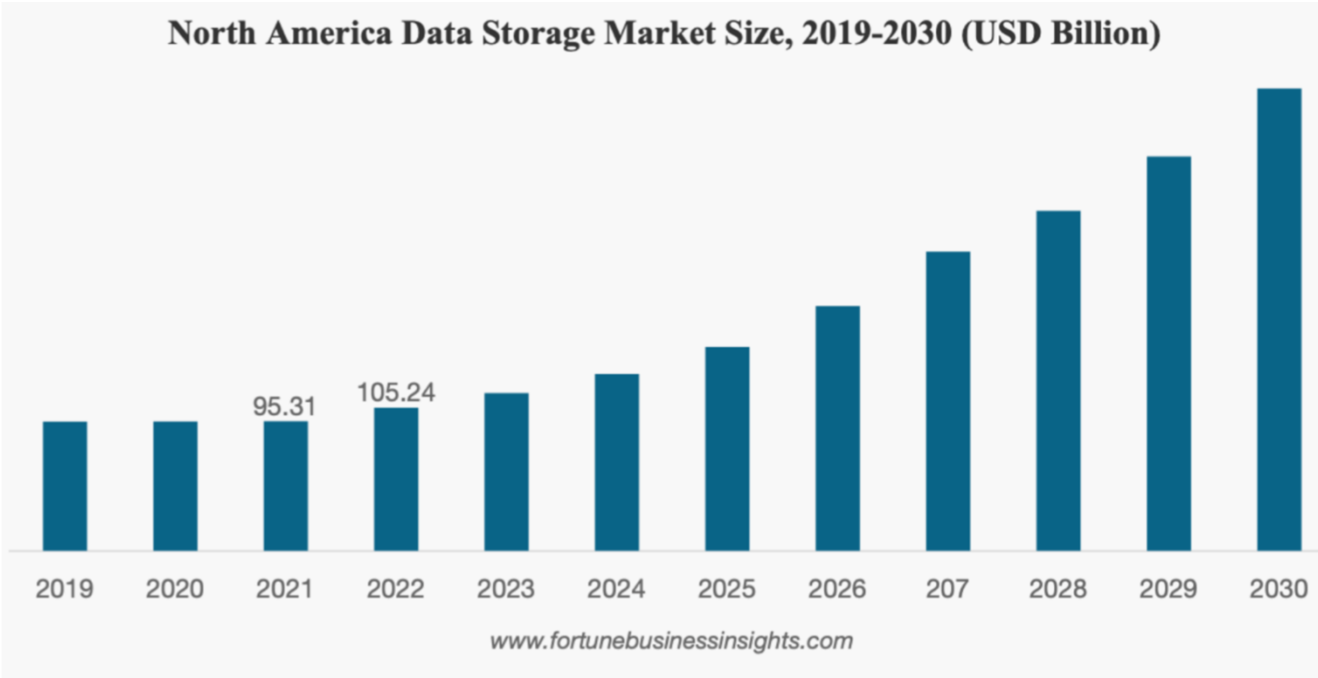

Fig. 2: North American market size and projected growth from 2019-2030 [17]

IV: Market Challenges: Current Problems with Data Storage

The Problem with Data Centers.

Data centers grapple with imperative issues that demand attention: the longevity of storage technologies employed for data retention, the environmental impact of storing vast amounts of data, and the immense costs associated with storing data.

Lifetime of Storage Media in Data Centers.

The anticipated lifetime of various storage media is relatively short in comparison to the extended duration in which data needs to be preserved. This means that data needs to be periodically migrated to preserve the stored information. The lifetime of HDDs ranges from 3 to 5 years, while flash media, on average, lasts between 20 and 30 years [19]. For SSDs, it is common to see drive failure rise after one year, with an average of almost 6% of drives failing after 6 years [20]. This demonstrates an SSD's unreliability in storing data long-term, similar to HDDs and magnetic tape.

The external environment in which storage media is placed plays a crucial role in the longevity of the media [21]. Maintaining an optimal environment with temperature conditions between 21°C and 27°C and humidity levels between 45 to 70% is crucial for data center storage media [21]. If data centers operate outside these ranges, there is a heightened risk for hardware failure and decrease in commuting performance, thus underscoring the importance of reliable data storage needed in data centers [21].

Environmental Impact of Data Centers.

Data centers house all the servers, storage media, and cooling technologies associated with storing and accessing data, thus creating a large demand for energy [22]. The electricity demand of data centers is substantial, representing about 1% of the global electricity demand in 2018 [22]. In the United States alone, data centers consume a significant amount of electricity, accounting for approximately 1.8% of the country's total electricity use [22]. According to the U.S. Department of Energy, data centers are extremely energy-intensive on a per-square foot basis; it is estimated that for every one square foot in a data center, the energy consumed is up to 50 times higher than a typical commercial building [22].

In the last two decades, there has been an unprecedented surge in the energy consumption of data centers[23]. In a five-year span at the turn of the century (2000-2005), the total electricity demand doubled, and in the following five years, grew 56% [23]. From 2010 to the present day, it is estimated that this energy consumption has risen 15 to 20% annually and is expected to rise every year [23].

The carbon footprint of data centers is also a growing concern. In 2018, data centers contributed to almost 0.5% of total greenhouse gas (GHG) emissions in the United States, with a total of 3.15×10^7 tons of CO_2-equivalent emissions [22].

Additionally, water usage is a crucial component to consider when evaluating the environmental implications of data centers [24]. The availability of water is a growing concern worldwide, and it is expected that from 2000 to 2055, water usage will increase by 55% due to an increase in manufacturing, thermoelectric power generation, and everyday domestic use [24].

The world has a heavy reliance on water, but so do data centers. Water is essential for humans to survive; it provides us with the necessary nutrition and hydration needed to sustain life, but data centers continually look to compete with these industries for water supply [24]. In 2014, data centers used 626 billion liters of water [24]. Given that the world drinks 147 billion liters of water per day and the world population is approximately 8 billion, data centers consume the equivalent of what nearly 94 million people would drink as shown in Eqs. 1, 2, and 3.

$$147\ billion\ liters\ water\ per\ day \times 365\ days\ per\ year = 53{,}655\ billion\ liters\ per\ year$$

(Eq. 1)

$$\text{Percentage of DC Water Usage to Global Consumption} = \frac{\text{Annual Data Center Water Usage}}{\text{Annual Global Water Consumption}} \times 100\%$$

$$\text{Percentage of DC Water Usage to Global Consumption} = \frac{626 \text{ billion liters}}{53{,}655 \text{ billion liters}} \times 100\% = 1.17\%$$

(Eq. 2)

$$1.17\% \times 8{,}019{,}876{,}189 \text{ people} \approx 94 \text{ million people}$$

(Eq. 3)

According to the International Monetary Fund, the current worldwide gross domestic product (GDP) has reached $109 trillion USD in 2024 [25]. The current data storage market is valued at $186.75 billion [5]. The size of the data storage market is a mere 0.17% of the total world GDP, yet consumes the equivalent of 1.17% of the world's water, showcasing the imminent problem of data centers' water consumption.

The reliance on water in data centers arises from both direct factors such as cooling and indirect factors like electricity generation [24]. As data center equipment operates, generating heat, it becomes crucial to regulate internal temperatures to prevent overheating [24]. This is typically achieved through two main methods: employing air chillers that use evaporated water to cool the surrounding air, which is then circulated throughout the data center, or by directly spraying water into the air within the data center [24]. Both approaches aim to achieve the same outcome: maintaining cooler temperatures inside the data center to safeguard equipment from overheating [24]. However, both methods involve the consumption of water through evaporation, thereby withdrawing it from the environment without returning it [24].

Similarly, in electricity generation for data centers, water is also consumed from the environment and not replenished [24]. Power plants generate electricity by heating water to produce steam, which subsequently powers turbines [24]. Although this water may be discharged as effluent or recirculated within the power plant system, it is never returned to its original source.

Total Cost of Ownership of Varying Storage Media.

To fully understand the varying costs associated with each storage media, the total cost ownership (TCO) is investigated. Yan et al. illustrates a scenario where a data center exclusively deploys one type of storage media [26]. Over a 50-year data center operational span holding 100 petabytes of data, a data center relying solely on HDDs would incur a cost of $300 million [26]. In contrast, a data center relying solely on SSD would elevate the cost to $600 million, while opting for flash storage would incur a cost of $150 million [26]. The figures encompass all expenses associated with a data center including IT, network, power, and cooling facility costs, building expenditures, energy costs, data migration costs, and human labor costs [26].

It is evident that the TCO of operating a data center is costly, but it is important to mention that the type of media primarily deployed makes a stark difference in the overall cost [26]. SSDs generate the highest cost of ownership by being two times higher than HDDs and four times higher than tape [26]. The choice in media is important in incurrings savings in the long-term for both the operators and the end-users of data centers.

V. Market Solution: Folio Photonics

Overview.

Folio Photonics is pioneering two cutting-edge technologies consisting of a polymeric multi-layer optical data storage media, a Folio disc, and a corresponding optical drive to read and write data [27]. This new technology harmonizes cost-effectiveness, competitive data-density capacities, and a low carbon footprint, while boasting a lifespan up to 100 years, making this a viable choice for archival data storage [28].

Cost-effectiveness.

In the data storage industry, the cost of data storage is often presented in price per terabyte, $/TB, for each storage media [28]. According to Folio Photonics, the media price for HDDs hovers around $20/TB, while LTO flash, one of the most common types of flash deployed in data centers, sits around $8/TB [28]. Folio Photonics is pricing their DFD technology at $3/TB when it enters into the market, but has roadmap plans to eventually hit that $1/TB mark, making it exceptionally cost-effective to deploy in data centers [28]. While SSDs are commonly utilized in data centers, Folio Photonics is specifically focused on HDDs and magnetic tape/flash as direct competitors in their business plans, thus the price per terabyte of SSDs is not included in their model [28].

This “price per TB” model may not always provide a complete picture of the expenses tied to data storage because it does not capture all the expenses incurred in preserving data long-term. Instead, it is essential to delve into the total cost of ownership (TCO) to gain a comprehensive understanding of the economics involved. As mentioned previously, a data center fully saturated with HDDs or flash would incur a cost of $300 million or $150 million respectively over a 50-year lifespan [26]. See IV: Market Challenges: The Problems with Data Storage, *Total Cost of Ownership of Varying Storage Media* for more information. If that same data center fully deployed optical under the same parameters, the TCO would amount to $20 million according to Yan et al., demonstrating the significant price disparity between commonly deployed storage media and optical media [26].

It is also worth noting that Yan et al. synthesized these findings based on mainstream SONY's Blu-ray optical discs with a capacity of only 300 gigabytes (GB) [26]. The study also used IBM LTO Ultrium 8 series magazine holding up to 12 TB for their flash storage calculations, and a single Shingled Magnetic Recording HDD holding 14 TB for their HDD calculations [26]. Had this analysis been conducted using a DFD, the contrast in TCO of optical compared to flash and HDD may have been even greater given the higher data density capacites and lower upfront costs of the Folio Photonics' DFD [26].

Data-density Capacities.

Traditional optical disc storage media offer modest data-density capacities, in comparison to alternative media storage technologies. The limited data capacity, typically around 50 GB, is attributed to the constraints imposed by the diffraction-limited behavior of light used to read and write data bits to the disc [29]. This limitation arises from the necessity to maintain a readable spot size of the bit written onto the disc [29].

Despite SONY's achievement in developing a higher data-density Blu-ray that can hold up 300 GB on a single disc, these capacities fall short in competition with the more robust data capacities of HDDs and flash storage [26]. This is precisely where Folio Photonics comes in.

Revitalizing innovation in the optical data storage sector, Folio Photonics has positioned itself as a trailblazer by introducing an 8-layer disc capable of storing up to a single TB of data [28]. Not only can the DFD hold a remarkable amount of data, making it competitive with HDDs and flash, but its price per TB supercharges its appeal, making it a viable option in the archival data storage industry.

Carbon Footprint.

Deploying optical data storage (ODS) can greatly reduce the carbon footprint of data centers [29]. In data storage, there are two general modes of operation: active mode where data is being recorded and/or read back and idle mode where the technology remains unused [29]. According to Yan et al. the peak powers used in active mode for optical, tape, and HDDs are 0.9, 0.8, and 6.4 kW respectively [26]. While optical and tape are similar in peak power consumption, optical does not require intense cooling measures that tape does [29]. This is due to the resiliency of optical storage media; it can withstand large temperature (10°C to 40°C) and humidity (20% to 80%) fluctuations and still function properly unlike HDD and tape [29]. Moreover, optical does not require any energy in idle mode unlike HDD, further showcasing its sustainability benefits [29].

Longevity of Optical Disc Storage.

According to the International Organization for Standardization, the suggested lifespan of an optical disc is approximately 30 years, but in reality it can range from as little as one year to beyond a century [30]. In this expansive spectrum, Folio Photonics promotes a lifespan leaning towards the upper-limit of this range [28]. Folio Photonics' Datafilm Disc advertises a 100-year lifespan in ideal environmental conditions, and up to 50 years in less-than-ideal conditions where environmental control may be challenging [28].

Not only does optical data storage boast a longer lifetime, but Folio Photonics' DFD is specifically categorized as *WORM* media [28]. WORM is an acronym standing for "write once, read many", meaning that this type of media allows data to be written once, but read back many times [31]. This is important because WORM media's lifetime may be less impervious to the numbers of accesses, further amplifying its long-term data storage benefits [28]. Additionally, optical storage media like the DFD are also resilient to moisture and water immersion and remain unaffected by electromagnetic pulses, making them less susceptible to natural and man-made disasters that may arise [28]. A longer lifetime and greater resiliency means less frequent data migration and data loss, making optical data storage suitable for archival data.

VI. Optical Technology

Blu-ray Disc:

Folio Photonics' DFD closely emulates Blu-ray technology. Consequently, a thorough examination of Blu-ray technology will be investigated to unravel the operational intricacies of a Blu-ray.

Composition of Blu-Ray Disc:

A Blu-ray disc is crafted from a polycarbonate substrate, with a height of 1.2 millimeters (mm) and an approximate diameter of 120 mm [32]. Similar to other optical discs, the substrate undergoes a process where a single track, a physical indentation, is intricately stamped, traversing from the inner to the outer diameter as depicted in Fig. 3 [33]. The space between these spiral tracks, commonly referred to as the track pitch, measures 320 nanometers (nm) [32]. A reflective metal layer is positioned atop the substrate to reflect laser light used to read and write data to the disc [34].

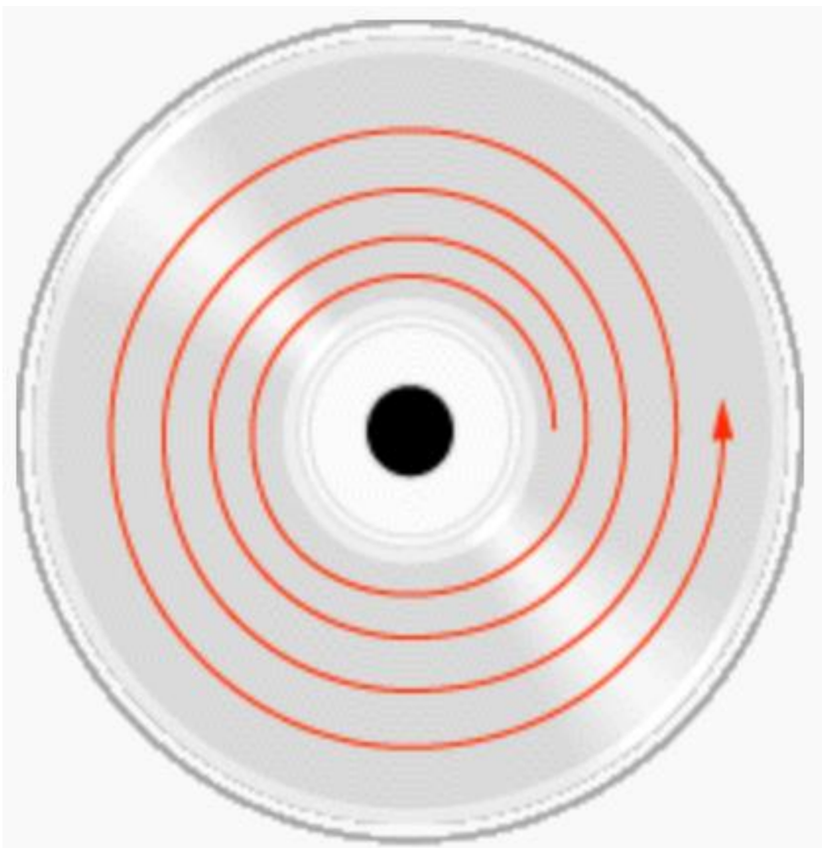

Fig. 3: A continuous spiral is stamped onto the Blu-ray disc [35]

How Data Is Recorded on a Blu-ray.

Information is stored in the form of binary digits called bits [36]. A bit is the basic unit of digital information in computing [36]. Data bits are physically imprinted onto the substrate using a blue laser diode, and they are situated between the indented tracks [37]. The laser emits short pulses, resulting in the formation of microscopic grooves where the laser light touches the surface [33]. These grooves are known as ‘pits’ on the disc’s surface [33]. Due to the intermittent nature of the laser pulses, there are sections of unmarked surface area between the pits, and these are commonly known as ‘lands’ [33]. Pits and lands are the physical features on an optical disc that represent the data bits.

Laser light is reflected off the metallized layer of the disc, creating changes in reflected light due to the pits and lands [34, 38]. The change in reflected light off of the metallized layer is interpreted by the photodetector inside the optical drive, and it converts the laser light into electrical signals using semiconductor materials [34, 38].

How Data Is Read on a Blu-Ray.

A solid state photodetector operates on the principle of the photoelectric effect to convert incoming light into an electrical signal to read data off of a disc [39]. When photons strike a semiconducting material, they can transfer their energy to electrons within the material, promoting them to a higher energy state [39]. This results in electron-hole pairs, i.e. negatively-charged electrons displaced by the photons and positively-charged holes where the electrons once were once located [39]. These displaced electrons transition to the conduction band; they act as free carriers that can move under the influence of either an intrinsic or externally applied electric field [39].

What were once photons reflected off the metallized surface of the disc is now interpreted as a stream of electrons, otherwise known as current, in the optical drive [39]. This current is used to read back the data bits containing information off the disc [39].

Binary Data Encryption.

This distinction between pits and lands forms the basis of encoding data onto an optical disc [33]. The transition between from pit to land or vice versa corresponds to a 0, while the absence of this transition corresponds to a 1 in the world of binary data encoding [33]. These 0s and 1s are interpreted from the change in reflected laser light off of the disc onto the photodetector [33].

Data Density.

The data density on the surface of an optical disc is defined by the number of data bits stored within a given area [40]. As the number of bits in a specific surface area increases, so does the data density [40]. Achieving higher data density involves packing data bits closely together, but this spatial arrangement is contingent on the spot size of the laser diode used for reading and writing data [40]. The laser's spot size must not be too large to ensure individual bits can be accurately read; hence, the spot size dictates how closely bits can be packed [40].

The spot size, denoted as D, is influenced by both the wavelength of the laser source, λ, and the numerical aperture (NA) of the lens within the laser diode [40]. The relationship is governed by Eq. 4 [40].

$$D = 0.5\lambda/NA$$

(Eq. 4)

Here, NA, a dimensionless value, characterizes the range of angles at which the lens within the diode laser can emit or accept reflected light as seen in Fig. 4 [40]. As the spot size is inversely proportional to NA, a higher NA results in a smaller spot size, consequently leading to a larger areal data capacity on the disc [40].

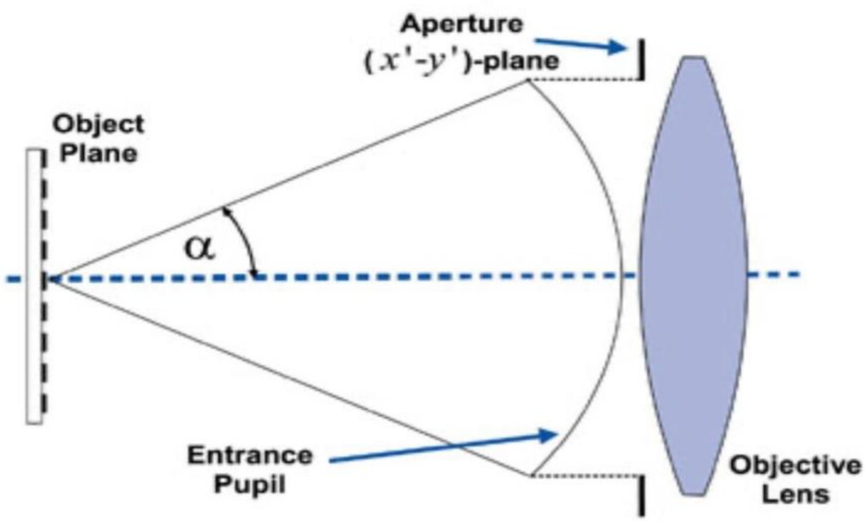

Fig. 4: Numerical aperture of a lens [51]

For Blu-ray, with a NA of 0.85 and a wavelength of 405 nm, the spot size is approximately 240 nm as shown in Eq. 5 [37]. For a DVD, with a NA of 0.6 and wavelength of 650 nm, the spot size is 540 nm as shown in Eq. 6 [37]. The spot size of a DVD is 2.25 times larger as shown in Eq. 7.

$$Blu-ray: D = 0.5 \times 405\, nm / 0.85 \approx 240\, nm$$

(Eq. 5)

$$DVD: D = 0.5 \times 650\, nm / 0.60 \approx 540\, nm$$

(Eq. 6)

$$D_{DVD} \ / \ D_{Blu-ray} \ = \ 540\,nm \ / \ 240\,nm \ = \ 2.25$$

(Eq. 7)

Storage Capacity.

The storage capacity of the disc hinges on the dimensions of the laser spot and is captured by Eq. 8 [40].

$$Storage\ Capacity \ = \ nA/\ D^2$$

(Eq. 8)

where A signifies the disk's area, and n represents the number of bits discernible within the spot [40]. The inverse dependence of storage capacity on the square of the spot size elucidates why a single-layer Blu-ray boasts a 25 GB storage capacity, surpassing a DVD, which holds a mere 4.7 GB, over a fivefold difference as shown in Eq. 10 [37].

$$Storage\ Capacity \ \propto \ 1/D^2$$

(Eq. 9)

$$D_{DVD} \ / \ D_{Blu-ray} = \ 2.25$$

(From Eq. 7)

$$Storage\ Capacity_{Blu-ray} \ / \ Storage\ Capacity_{DVD} \ \propto \ (D_{DVD} \ / D_{Blu-ray})^2 \ = \ 2.25^2 \approx 5$$

(Eq. 10)

The storage capacity of a Blu-ray disc is over five times larger than that of DVD, stemming from the fact that storage capacity is inversely proportional to the square of the spot size. Because the spot size of a DVD is 2.25 times larger than a Blu-ray, the storage capacity of a DVD is less than ⅕ of Blu-ray. In other words, the storage capacity of the Blu-ray is five times higher than a DVD.

Folio Photonics' DFD.

Folio Photonics, a pioneering company in optical storage solutions, employs cutting-edge technology, the Datafilm Disc, to enhance data storage capabilities. Much like Blu-ray technology, Folio Photonics utilizes a polycarbonate substrate crafted with spiral tracks, with a proprietary reflective layer on top, providing a foundation for efficient data encoding [27]. What sets them apart is their innovative approach of adhering a multilayer polymeric film to this reflective surface [27]. This multilayer film serves as the active data storage medium [27]. A blue laser is used to write data to these active layers, capitalizing on the advantages of a shorter wavelength for higher data density [27]. As stated previously, a smaller wavelength corresponds to a smaller spot size, thus producing a higher data density.

Notably, Folio Photonics also incorporates a buffer layer strategically positioned between the active layers to mitigate cross talk between active layers, demonstrating a commitment to optimizing data integrity [27].

In their pursuit to a TB capacity, the DFD will feature two 8-layer films adhered to both sides of the substrate [27]. Similar to a Blu-ray, each layer will hold approximately 25 GB, thus reaching an 800 GB capacity as shown in Eq. 11 [27]. The company plans to expand the number of layers in subsequent years to surpass the single-TB benchmark [28].

$$DFD\ Storage\ Capacity = 25\ GB/layer \times 16\ layers/film \times 2\ films/disc = 800\ GB/disc \quad \text{(Eq. 11)}$$

VII. Early Adopters

Definition of Early Adopter.

An early adopter embodies an individual person, business, or company that embraces cutting-edge technology or innovation at its inception, positioning themselves ahead of the curve before widespread adoption [41]. Often willing to pay a premium for this pioneering move, they recognize the potential impact of implementing this new technology to trim costs, optimize processes, or elevate the value of their product or service, thus justifying the initial investment [41].

Beyond financial gains, being an early adopter can confer a unique social status as they integrate new technologies into their overall framework [41]. This move showcases their forward-thinking approach and enhances their reputation as an innovator and trendsetter in their own domain [41].

Early Adopters of Folio Photonics' DFD: Amazon Web Service.

Amazon Web Services (AWS) comes to the forefront as an early adopter of Folio Photonics' pioneering technology because of the sheer volume of data that AWS stores. Folio Photonics' Datafilm Disc outpaces its competitors in more than one way including competitive pricing, sustainability advantages, and lifetime performance, making it a worthwhile option for AWS to adopt.

To gain insight into why AWS accumulates such vast amounts of data and the implication of storing so much data, a closer examination of the company is necessary. According to AWS, they are the world's most broadly adopted cloud platform with over 200 data centers worldwide, serving millions of customers every day including growing start-ups, large enterprises, and governmental agencies around the world [42]. AWS incorporates many features and services that other cloud providers may not include, like access to large databases, machine learning, artificial intelligence, data lakes and analytics, and IoT [42]. These advanced analytics are rapidly growing, but they also require an immense amount of data storage to be deployed. This presents a challenge for the global cloud platform provider as the need for the data continues to rise substantially.

Storing data is costly from a media vantage. As you accumulate data, this necessitates a need for more media to be deployed. Folio Photonics has a roadmap reaching $1/TB in upcoming generations of their technology, but the company currently projects the first generation to achieve $3/TB, well below other media prices, making this media a viable option for AWS to pursue [43].

While the media price point is important to consider, there are other aspects of the Datafilm Disc that make it competitive in the data storage industry, specifically its energy savings in data centers. According to Data Centre Magazine, AWS is considered a hyperscale data center, meaning they build “massive” facilities to run data-processing operations and fulfill data storage needs from their customers [44]. They also mention that these data centers require intense cooling measures to maintain temperature and humidity within their data centers and deploy renewable energy technologies where they can, implying their need for large electricity generation [44]. This presents another challenge for AWS as the global economy enters into an energy crisis. Folio Photonics presents an energy-efficient solution that not only reduces energy consumption during operation, but also eliminates the need for additional cooling measures to maintain functionality.

Lastly, Folio Photonics' DFD also boasts a minimum lifetime of 50 years, but could even reach 100 years compared to an average range of 4 to 10 years for HDD, SSD, and magnetic tape [19, 20, 28]. According to their site, AWS deploys both HDD and SSD in their data centers, highlighting their reliance on short-lived storage media [45]. This exemplifies the need for media with longer lifetimes to reduce data migration and overall media costs.

A major enterprise like AWS is poised to be an early adopter of TB-capacity optical due to its cost-effectiveness in both the near and long-term, lower energy requirements to maintain and operate, and extended lifespan compared to incumbent technologies.

VIII. Competition

Hard Drives.

Serving as one of the leading storage solutions, hard drives pose an ever present challenge to Folio Photonics because they continue to remain a prominent storage media in data centers with high capacities.. The advancements in the HDD industry, specifically helium hard drives and MAS-MAMR hard drives, ought to be investigated to determine whether the improvements of this incumbent technology will outpace the Datafilm Disc.

Helium Hard Drives.

Helium hard drives stand out as one of the most significant improvements seen in data center technology in the past decade. Helium is used in place of air in conventional hard drives because helium has an atomic weight of 4 compared to air with 29, making it seven times less dense than air [46]. This decrease in density minimizes the drag and turbulence that a spinning platter in an HDD would otherwise feel as it rotates [46]. This leads to a lower operating temperature, which results in less heat generation [46]. Furthermore, the power consumption of a helium HDD is 28% and 19% lower in idle and active modes respectively [46].

In addition to lower operating temperatures and a decreased power consumption, helium HDDs also have the potential to boost storage capacities [46]. A standard hard-drive drive has a 1" form factor, so the number of discs is limited by this parameter [46]. Helium-filled hard drives reduce the vibrations within the hard drive, and this reduction in vibrations enables closer stacking of discs, allowing for a higher data density [46].

The challenge in this technology, however, lies in sealing the HDDs to prevent helium from escaping or ambient air from entering [46]. Given that there are 10 openings on a hard drive, numerous techniques have been deployed including the use of gaskets, soldering, canning, and welding [46].

In 2021, HDDs manufacturer's officially launched a 10-disc HDD [46]. However, if HDDs want to remain competitive with other cost effective solutions like magnetic storage and optical, they will need to surpass a 10-disc hard drive [46]. What this really means is the helium hard-drives were not the breakthrough the HDD industry needed to remain competitive. It certainly showcases, however, the continued improvement and ongoing efforts of HDD manufacturers.

MAS-MAMR Hard Drives.

Toshiba has pioneered a new type of HDD with improved data density capacities, reaching up to 30 TB; this poses challenges for incumbent technologies with limited storage capabilities, such as optical storage, which already face difficulties in achieving the current industry standards for capacity [47]. The company is utilizing Microwave Assisted Switching-Microwave Assisted Magnetic Recording (MAS-MAMR) technologies to increase the recording density [47].

Although the prospect of expanding HDD storage capabilities appears to be on the horizon, it does not come without complexity. Increasing HDD recording density poses a challenge known as the "trilemma", which involves balancing three conflicting objectives: miniaturizing magnetic grains on the recording medium, ensuring thermal stability of the medium, and maintaining sufficient recording performance through the read/write head [47]. While shrinking the magnetic grains used in writing enhances data density, it compromises the thermal stability of the medium, and thereby increases the risk of data loss [47]. Moreover, achieving higher thermal stability requires a greater coercivity, but this impedes the head's ability to generate the necessary magnetic field needed for reading and writing [47]. Addressing this challenge necessitates external energy assistance in the recording process, hence why Toshiba is focusing their R&D efforts on developing this technology [47].

While MAS-MAMR HDDs remain in development, this future advancement may pose challenges for Folio Photonics, as they strive to reach a technological equivalent to HDD storage capabilities today. If HDD manufacturers succeed in enhancing data storage capacities, it could lead to a reduction in the cost per TB, thereby diminishing the competitive pricing of the Datafilm Disc.

High Data-Density Optical Storage.

As Folio Photonics attempts to compete with existing media technologies commonly deployed in data centers with the aim of supplanting these incumbent technologies, it is worthwhile investigating researchers who may also be developing high data density optical discs.

In February of 2024, researchers from the University of Shanghai in conjunction with other research institutions published a scientific research paper on a novel way to create a 3D optical disc with petabit capacity, equating to over 100 TB of data capacity on a single disc [48]. Researchers discovered a way to create a 54-nm spot size, while shrinking down the track pitch to 70 nm [48]. While the spot size determines the capacity of each layer, the number of layers is also pertinent to the overall data storage capacity [48]. Researchers are projecting the ability to record data to 100 layers, furthering its storage capabilities [48].

This technology faces similar challenges to Folio Photonics' DFD in several ways, specifically the working distance of the objective and cross-talk among layers. According to the authors, the writing depth of the laser diode is limited by the working distance of the objective lens; the objective lens can only focus up to 100 microns into the disc, thus the minimum layer size needs to be less than 1 micron to achieve 100 readable layers on the disc [48]. Folio Photonics faces a similar setback in the working distance of their objective lens, creating a barrier to increasing data density. Moreover, cross-talk among active layers is a common issue in any 3D optical media, but researchers at the university showcased that the even and odd layers in their film exhibit alternating patterns, indicating that cross-talk may not be a current threat to their optical disc [48]. Not unlike this emerging technology, Folio Photonics has to consider cross-talk, as well, when building out their optical disc to ensure data integrity.

Bit error rate is yet another issue facing general data storage media. While Folio Photonics does not currently disclose their bit error rate, researchers at the University of Shanghai found theirs to be 0.33% [48]. According to two prominent companies in the Communication Technologies industry, FOSCO and Viavi Solutions, an acceptable bit error rate in data communications is 10^{-13}, leaving little room for error [49, 50]. This means this new technology may be far from commercialization, but nonetheless it may be a considerable threat to Folio Photonics' emerging DFD given the substantial increase in storage capacities it may offer.

Works Cited

[1] E. M. Hart, P. Barmby, D. LeBauer, F. Michonneau, S. Mount, P. Mulrooney, T. Poisot, K. H. Woo, N. B. Zimmerman, and J. W. Hollister, *Ten Simple Rules for Digital Data Storage*, PLOS Computational Biology **12**, e1005097 (2016).

[2] *What Is Data Storage? | IBM*, https://www.ibm.com/topics/data-storage.

[3] J. Wu, L. Ping, X. Ge, Y. Wang, and J. Fu, *Cloud Storage as the Infrastructure of Cloud Computing*, in *2010 International Conference on Intelligent Computing and Cognitive Informatics* (2010), pp. 380–383.

[4] M. C. Whitlock, M. A. McPeek, M. D. Rausher, L. Rieseberg, and A. J. Moore, *Data Archiving*, The American Naturalist **175**, 145 (2010).

[5] *Data Storage Market Size, Share & Growth Statistics [2032]*, https://www.fortunebusinessinsights.com/data-storage-market-102991.

[6] U. Nations, *Big Data for Sustainable Development*, https://www.un.org/en/global-issues/big-data-for-sustainable-development.

[7] *Big Data Analytics Market Size, Value & Share Analysis [2032]*, https://www.fortunebusinessinsights.com/big-data-analytics-market-106179.

[8] *Cloud Computing Market Size, Share, Value & Forecast [2032]*, https://www.fortunebusinessinsights.com/cloud-computing-market-102697.

[9] U. Nations, *Big Data for Sustainable Development*, https://www.un.org/en/global-issues/big-data-for-sustainable-development.

[10] *Dictionary.Com | Meanings & Definitions of English Words*, https://www.dictionary.com/browse/digital-footprint.

[11] A. Williams, *What Is IoT? – Internet Of Things Explained*, https://www.forbes.com/sites/technology/article/what-is-iot/.

[12] *Data Storage Market Size, Share & Growth Report, [2023-2030]*, https://www.snsinsider.com/reports/data-storage-market-2768.

[13] *Artificial Intelligence - Global | Statista Market Forecast*, https://www.statista.com/outlook/tmo/artificial-intelligence/worldwide.

[14] *Machine Learning - Worldwide | Statista Market Forecast*, https://www.statista.com/outlook/tmo/artificial-intelligence/machine-learning/worldwide.

[15] S. Research, *Internet of Things (IoT) Market Size, Outlook, Growth, Report to 2030*, https://straitsresearch.com/report/internet-of-things-market.

[16] I. Odun-Ayo, O. Ajayi, B. Akanle, and R. Ahuja, *An Overview of Data Storage in Cloud Computing*, in *2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS)* (2017), pp. 29–34.

[17] Data Storage Market, Fortune Business Insights, 2024, https://www.fortunebusinessinsights.com/data-storage-market-102991.

[18] J. M. MacDonald, *Commercial Sector - an Overview | ScienceDirect Topics*, https://www.sciencedirect.com/topics/social-sciences/commercial-sector

[19] M. Dimopoulou and M. Antonini, *Data and Image Storage on Synthetic DNA: Existing Solutions and Challenges*, J Image Video Proc. **2022**, 23 (2022).

[20] C. Chakraborttii and J. Boettner, *Leveraging Temporality of Data to Improve Failure Predictions for Solid State Drives in Data Centers*, in *2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)*(2023), pp. 1272–1278.

[21] D. Van Le, J. Zhou, R. Wang, R. Tan, and F. Duan, *Impacts of Increasing Temperature and Relative Humidity in Air-Cooled Tropical Data Centers*, IEEE Transactions on Sustainable Computing 1 (2024).

[22] M. A. B. Siddik, A. Shehabi, and L. Marston, *The Environmental Footprint of Data Centers in the United States*, Environ. Res. Lett. **16**, 064017 (2021).

[23] A. Capozzoli and G. Primiceri, *Cooling Systems in Data Centers: State of Art and Emerging Technologies*, Energy Procedia **83**, 484 (2015).

[24] D. Mytton, *Data Centre Water Consumption*, Npj Clean Water **4**, 1 (2021).

[25] *Worldwide Economic Outlook (April 2024)*, https://www.imf.org/external/datamapper/profile/WEOWORLD

[26] W. Yan, J. Yao, Q. Cao, and Y. Zhang, *LT-TCO: A TCO Calculation Model of Data Centers for Long-Term Data Preservation*, in *2019 IEEE International Conference on Networking, Architecture and Storage (NAS)* (2019), pp. 1–8.

[27] K. Singer and I. Shiyanovskaya, *Optical Storage Medium*, (15 February 2024).

[28] K. Singer, S. Ijaz, and S. Santamaria, A ROADMAP FOR COST CONTAINMENT IN DATA ARCHIVING, Folio Photonics, n.d.

[29] M. Gu, X. Li, and Y. Cao, *Optical Storage Arrays: A Perspective for Future Big Data Storage*, Light Sci Appl **3**, e177 (2014).

[30] Y. Mitsuhashi, Optical Data Storage – How Long Will It Last?, ISO, n.d.

[31] K. Wada, *Write Once Read Many*, in *Encyclopedia of Database Systems*, edited by L. LIU and M. T. ÖZSU (Springer US, Boston, MA, 2009), pp. 3570–3570.

[32] White Paper Blu-Ray™ Disc Format: Physical Format Specifications for BD-ROM, Blu-ray, 2015.

[33] *Data Storage: Optical Discs*, https://learn.saylor.org/mod/book/view.php?id=66673&chapterid=60176.

[34] *3. Disc Structure • CLIR*, https://www.clir.org/pubs/reports/pub121/sec3/.

[35] *Methods and Materials: CDs and DVDs | Ismail-Beigi Research Group*, https://volga.eng.yale.edu/teaching-resources/cds-and-dvds/methods-and-materials.

[36] *What Are Bits, Bytes, and Other Units of Measure for Digital Information?*, https://kb.iu.edu/d/ackw

[37] *Blu-Ray | Definition, Facts, & History | Britannica*, https://www.britannica.com/technology/Blu-ray.

[38] R. Hui, *Chapter 4 - Photodetectors*, in *Introduction to Fiber-Optic Communications*, edited by R. Hui (Academic Press, 2020), pp. 125–154.

[39] *Photodetector - an Overview | ScienceDirect Topics*, https://www.sciencedirect.com/topics/engineering/photodetector
[40] *Optical Disk - an Overview | ScienceDirect Topics*, https://www.sciencedirect.com/topics/physics-and-astronomy/optical-disk.
[41] *What Is an Early Adopter, and How Does It Work (With Examples)?*, https://www.investopedia.com/terms/e/early-adopter.asp.
[42] *What Is AWS? - Cloud Computing with AWS - Amazon Web Services*, https://aws.amazon.com/what-is-aws/.
[43] *Product | Folio Photonics*, https://foliophotonics.com/product.
[44] I. Ozsevim, *Top 10 Data Centre Hyperscalers*, https://datacentremagazine.com/articles/top-10-data-centre-hyperscalers.
[45] *SSD vs HDD - Difference Between Data Storage Devices - AWS*, https://aws.amazon.com/compare/the-difference-between-ssd-hard-drive/.
[46] A. Aoyagi, B. Stipe, R. Wood, S. Campbell, and X. Che, *Helium Sealed Hard Disk Drive*, Journal of Magnetism and Magnetic Materials **564**, 170146 (2022).
[47] *Toshiba Achieves World's First Demonstration of Substantial Improvements in HDD Recording Performance with Microwave Assisted Switching - Microwave Assisted Magnetic Recording (MAS-MAMR) | Toshiba Electronic Devices & Storage Corporation | Americas – United States*
[48] M. Zhao, J. Wen, Q. Hu, X. Wei, Y.-W. Zhong, H. Ruan, and M. Gu, *A 3D Nanoscale Optical Disk Memory with Petabit Capacity*, Nature **626**, 772 (2024).
[49] *Bit Error Rate | BERT | Understanding BER Testing in Fiber Optics | VIAVI Solutions Inc.*, https://www.viavisolutions.com/en-us/products/bit-error-rate-test-bert.
[50] *What Is BER (Bit Error Ratio) and BERT (Bit Error Ratio Tester)?*, http://www.fiberoptics4sale.com/wordpress/what-is-ber-bit-error-ratio-and-bert-bit-error-ratio-tester/.
[51] *The Numerical Aperture*, https://spie.org/publications/spie-publication-resources/optipedia-free-optics-information/fg06_p16_na.