Measuring the Cost of Reliability in Archival Systems

James Byron
University of California, Santa Cruz
jbyron@ucsc.edu

Darrell D. E. Long
University of California, Santa Cruz
darrell@ucsc.edu

Ethan L. Miller
University of California, Santa Cruz
Pure Storage
elm@ucsc.edu

***Abstract*—Archival systems provide reliable and cost-effective data storage over a long period of time. Existing technologies offer familiar and well-defined features, but uncertainties about future developments complicate decisions about selecting the best storage technology that will continue to scale in the future. Furthermore, inaccurate assumptions about the long-term reliability of each storage technology can result in the use of suboptimal storage technologies for an archival system or the unwanted loss of data. Prospective storage technologies like archival glass and synthetic DNA may deliver much greater capacity and reliability than do existing technologies, yet their availability and exact features remain uncertain. As each storage technology develops and changes over time, its reliability may also change and give rise to further uncertainties about the long-term cost of highly reliable archival systems. We present results of simulations that explore the effects of various technology developments upon the cost of constructing archival systems that meet various levels of reliability against data loss. We show that storage density more than device reliability dominates the cost of constructing and maintaining reliable archival storage systems, and innovations to increase storage density—even at the expense of individual device reliability—can reduce total archival system cost. We also explore the advantages of prospective over existing archival storage technologies, and we present estimates of the extent to which their availability will affect the cost of long-term data storage.**

***Index Terms*—reliability, archival storage, simulation, performance**

I. Introduction

The need for reliable data storage often guides the design and deployment of archival storage systems. As archival systems grow in both size and number to store ever-increasing amounts of data, so does the importance of reliability as a central design feature for storage systems. The risk of losing data within an archival storage system increases with the size of the system and the length of time over which it stores data. For this reason, numerous techniques exist to reduce the probability of data loss.

Replication, RAID, and the use of highly reliable storage technologies can serve to increase the reliability of a storage system. Archival systems, which typically serve low-intensity workloads over a long period of time, must also integrate available storage technologies and organize them into a design that offers the needed storage capacity and reliability while minimizing the acquisition and operating costs of the archive. Balancing such requirements is a significant challenge for archival system designers. Choosing the best storage technology and system design to meet a given threshold for reliability can prove challenging due to the variability of device performance and reliability over time as well as the unique features of each storage technology within a large archival system. Furthermore, predicting the long-term cost of an archive that meets its designer's needs for reliability further complicates the design and implementation of archival systems.

We have implemented an archival storage simulator that measures and compares the cost of acquiring and operating a storage system while simultaneously assessing the level of storage reliability that it offers. For the purposes of this work, we include existing storage technologies–tape, optical disc (ODD), hard disk (HDD), and solid state disk (SSD)–as well as prospective storage technologies that are currently under development for archival storage systems–glass and synthetic DNA. Each storage technology offers unique features, limitations, rates of development, and expectations for long-term reliability. Our simulator calculates the cost of an archive and simulates its continued use over time to calculate both its reliability and total cost of ownership.

The main contributions of this paper include: 1) We analyze historical hard disk data, showing that hard disk reliability is increasing as its rate of development slows. 2) We show that technical developments for increasing the capacity of solid state disks will improve the competitive value of SSDs for archival storage relative to other technologies notwithstanding the possibility that advancements like storing more bits per cell of flash will have a negative impact on SSD reliability. 3) We describe the conditions under which prospective storage technologies like archival glass and synthetic DNA can reduce the cost of highly reliable archival storage, and we present recommendations for each technology to improve its utility for economical and reliable archival storage.

II. Related Work

Understanding and predicting storage system reliability has motivated their design and evaluation. Proposals to design highly reliable storage systems assume that storage devices are not sufficiently reliable for all use cases, particularly in large storage systems where many devices operate and present their

This research was supported in part by the National Science Foundation under awards IIP-1266400 and IIP-1841545 and also by the industrial members of the Center for Research in Storage Systems.

own independent probability of failure. Research on measuring and predicting the reliability of storage systems evaluates existing storage system designs to predict their total reliability given certain assumptions about the underlying reliability of the components within the system. Predictions of storage system reliability can be utilized to inform decisions about storage system design, storage technology manufacturing, as well as how best to utilize a storage system within a data center.

A. Reliability Models

A Redundant Array of Inexpensive Disks (RAID) is a storage system design to use inexpensive and relatively unreliable disks to match and exceed the reliability of expensive and reliable disks at a lower cost [1]. RAID exploits the statistical improbability of multiple simultaneous device failures to construct groups of data and parity drives that collectively offer configurable performance and reliability as needed for different use cases.

Increasing the number of parity drives or, alternatively, reducing the time required to repair a RAID group [2] can reduce the probability of data loss within a RAID group. Elerath et al. observed that in practice the failure probability over time for any storage device does not strictly follow a Poisson distribution. Instead, storage devices may fail with different probabilities throughout their lifetimes due to manufacturing defects, random failures, or from wearing out at the end of their useful lives [3]. Elerath proposed a reliability model for RAID systems based on the Weibull distribution of failure probability for each storage device. Storage devices may also fail suddenly in batches due to shared manufacturing defects. Designing RAID groups with spare drives to dynamically respond to failures [4] and distributing data across RAID groups from different drive batches or manufacturers [5] can mitigate the risk of data loss due to batch drive failures. External factors like natural disasters can also impact the reliability of storage.

Data replication and distribution are also frequently used in commercial storage systems to reduce the probability of data loss [6]. Li et al. observed that replicating data across different geographically-separated locations can significantly reduce the probability of data loss due to any localized event that may impact storage system availability [7]. Simulation has also been used to quantify the performance [8], [9] and mean time to data loss [10] of data replication and distribution schemes.

B. Storage System Models

The increasing need to store, use, and extract value from data over an extended period of time promotes the growing demand for archival storage systems. Hughes introduces an economic analysis of long-term data storage [11], observing that as the cost of storing each byte of data decreases and approaches zero, the quantity of information in storage systems will grow virtually without limit. Furthermore, growth in the amount of data will also increase the total value of the information stored in an archival system. Still, the viability of long-term storage depends on the technical developments and solutions that will ensure data remains available well into the future [12].

Numerous models have been implemented to measure the cost and performance of storage systems. ExaPlan is a storage system model that minimizes data access latency based on a given budget and workload using variably-sized tiers of storage devices [13]. Although ExaPlan includes parameters for cost and performance, it focuses on minimizing latency rather than exploring the effects on cost and overall system reliability of various alternative storage device designs and types.

Other previous work used a simulation model to evaluate the cost and performance of archival storage systems built with competing storage technologies [14]. Although the work we present in this paper utilizes a similar simulation model and parameters to those of the earlier work, we extend beyond the previous work with a consideration of reliability in archival systems and with the inclusion of prospective archival storage technologies.

Other models have measured the effects that different rates of development may have on the long-term viability of competing storage technologies like hard disk drives and Flash [15]. Other models have been used to compare the acquisition and operating costs among various storage technologies [16]. Existing models have not compared the total cost of archival systems based on the reliability of the devices and potential for future developments for each storage device type.

III. Simulator Design

Our simulator consists of several modules that we configure to model the functionality of archival systems. Here we present a summary of each critical component.

A. Events and Event Driver

We use an event-driven model to simulate actions that occur at specified time intervals or with a certain probability over time. Each action is represented by an Event object. Events include installing devices, reading and writing data, and replacing devices that have failed. Each Event acts upon the devices within the simulated archival system, and the status of the device changes accordingly to ensure that no conflicting actions utilize the same resource at the same time.

B. Time

Time within the simulator affects each of the Events that are acting upon the archival system. The passage of time in the simulator triggers new Events and adds them to a queue. The current time in the simulator also controls the expiration of events and values that change with time. Events to install devices, for instance, require a certain amount of time to be completed before the affected device can be utilized for another event. The simulation time affects values such as capacity, failure probability, and device features that change as a function of time.

C. Archival System and Devices

The Archival System class coordinates actions within the simulator and forwards actions to each individual storage device. The Archival System provides functions to calculate the total capacity, read and write throughput, reliability, and cost of the archival system. The Archival System class also calculates the total cost of meeting a threshold requirement of capacity and performance.

Devices within the simulator include drives and media, networking infrastructure, robots, and racks. We implement classes for each device type to model its unique behavior and features. Devices with removable media such as tape and optical disc use separate Device classes to represent the media and drives. Devices such as hard disk and solid state disk, which do not have separable media, are represented as one Device class. Glass and DNA storage, which feature separable media as well as separate devices to read and write data, are represented with classes for media, a drive for reading, and a drive for writing.

D. Configuration and Parameters

We use two types of configuration parameters to control the behavior of the simulator: archival parameters and device parameters. Archival parameters define the required capacity of the archive and its rate of growth, the performance and workload demand on the archive, the cost of electricity, and the number of years to simulate. The device parameters include the cost of each device, its performance and capacity as functions of time, and the device's probability of failure.

IV. Reliability Model

The reliability of a storage system depends upon that of the devices in the storage system, the organization of the storage system, and other events and conditions beyond the storage system. Events and conditions beyond the storage system include software errors, user errors, natural disasters that affect a data center, and electrical faults or surges. While external events are important to storage system reliability [7], their occurrence is not intrinsic to the design or implementation of any storage system since virtually any storage system can be affected by such events. We therefore focus on modeling the reliability of storage devices used in an archival system as well as the design of the archival storage system that can optimize it for reliable long-term operation.

In an archival storage system with many storage devices, each device presents its own probability of failure that typically changes over its lifetime. We define a device failure to be the condition of a device that either cannot reliably write or read data or that has diminished performance relative to its manufacturer's specification. A device failure may–but not necessarily–lead to data loss, particularly if the device begins to degrade its performance shortly before it fails completely. For the purposes of this work, we compare the reliability of storage devices within RAID groups in terms of their probability of failure resulting in data loss and in terms of the amount of data lost with a failure.

A storage system with n devices and no scheme for data redundancy has a total probability of data loss defined as

$$P(S,t) = 1 - \prod_{i=1}^{n}(1 - p(i,t)), \tag{1}$$

where $P(S,t)$ is the total probability of data loss due to a device failure in a storage system S over time t and $p(i,t)$ is the probability that device i will fail over the time period t. The total probability of data loss in large storage systems increases geometrically with the number of devices and the length of time over which the storage system is used to preserve data. Archival storage systems, which must store ever-increasing amounts of data, typically require many devices to store data for a long period of time. The large number of storage devices and longevity of archived data render archival storage systems especially prone to unplanned data loss resulting from device failure.

RAID systems may be configured to replicate data across some number of devices to decrease the probability that any combination of device failures will result in data loss. Various RAID levels combine Hamming codes with data distribution across multiple devices to offer configurable levels of performance and data reliability [1]. The probability of data loss in RAID systems depends on the probability that multiple devices within a RAID group will fail during the time required to recover from a failure. RAID levels with extra redundancy or parity can tolerate more near-simultaneous device failures, but they require more devices to store the same amount of data than RAID levels with less redundancy or parity. Archival system designers balance performance and reliability versus cost to suit their particular needs.

We define the failure probability of a RAID storage system as

$$P(S,t) = 1 - \prod_{n=1}^{g}\Big(\prod_{i=1}^{d_n+q_n}\big(1 - p(i,t) * \prod_{j=1}^{q}(p(j,r))\big)\Big), \tag{2}$$

where $P(S,t)$ is the total failure probability of a storage system consisting of g RAID groups, d is the number of data drives in a RAID group, $p(i,t)$ is the probability of failure for device i over time t, q is the number of parity drives in the RAID group, and $p(j,r)$ is the probability that parity device j will experience a failure in the time taken to reconstruct the data in the group, given as r. We use this model to simulate archival systems that use RAID levels 1, 5 , 6, as well those that have an arbitrary number of data and parity drives; however, we do not model storage systems that combine different RAID levels together, nor do we compare RAID with other schemes for increasing reliability such as distributed erasure coding.

We can use the failure probability of a storage system from Formula 2 to determine its reliability over time by evaluating

$$R(S,t) = 1 - P(S,t), \tag{3}$$

where R is the reliability of a storage system S over time t, given the function P for failure probability.

The probability of failure for any storage system should ideally decrease as the amount of data that would be lost from a failure increases. The potential amount of data lost depends on the amount of data in a RAID group, which itself depends upon how many data drives are used in the group as well as the capacity of those drives.

As the capacity of each storage device increases with time and development, so too does the capacity and rebuild time of each RAID group, assuming a fixed number of drives in each group. Using high capacity storage devices in a small number of large-capacity RAID groups introduces a greater risk for catastrophic data loss than many smaller RAID groups would. Large storage devices and RAID groups necessarily increase the amount of data that can be lost during any failure. By Formula 2, a small number of RAID groups may present a lower total probability of data loss due to a RAID group failure than a larger number of RAID groups would; however, reducing the number of RAID groups requires increasing each group's capacity to store a given amount of data. Rebuild times for larger-capacity groups are also longer, which offsets the reliability advantages of reducing the number of groups by using larger drives. Maximizing reliability in a storage system therefore requires balancing the use of many small RAID groups with fast rebuild times and larger RAID groups with lengthy rebuild times. Formula 2 also disguises the amount of data that would be lost as a result of any single failure, however rare that may be.

Storage systems and RAID groups should become more reliable as they store more data. Existing storage technologies continue to develop apace, yet the reliability of each device has not increased as quickly as its capacity. Prospective storage technologies like DNA and glass also promise large capacities, and their reliability remains uncertain. High capacity storage technologies allow increasing amounts of data to be concentrated within each RAID group; however, as the probability of failure $P(S,t)$ decreases with fewer RAID groups that each have greater capacity, the probability of failure relative to capacity may increase. We define the probability of failure relative to capacity with the formula

$$F(s,t) = p(s,t) * C_s, \tag{4}$$

where $F(s,t)$ is the probability of failure relative to capacity, $p(s,t)$ is the probability of storage device or RAID group s failing over time t, and C_s is the storage capacity of device or group s. By Formula 4, a RAID group that stores $1.000\,\mathrm{TB}$ should offer a lower probability of failure p by a factor of at least 100 than a RAID group with a capacity of $10\,\mathrm{TB}$ in order to compensate for its larger capacity. The *blast radius* of a RAID group is equal to its failure probability relative to its capacity as given in Formula 4. We calculate the total blast radius of a storage system with the formula

$$B(S,t) = \sum_{n=1}^{g} \Big(\big(1 - \prod_{i=1}^{d_n+q_n} (1 - p(i,t) * \prod_{j=1}^{q} (p(j,r)))\big) * C_n \Big), \tag{5}$$

where $B(S,t)$ is the cumulative blast radius for storage system S over time t, and C_n is the total capacity of RAID group n.

V. Parameters

We simulate the reliability, performance, and cost of archival storage systems using parameter values that describe candidate archival storage technologies. Our simulator uses configuration parameters to define the performance, capacity, reliability, and cost of storage devices that may be used within a storage system. The output of our simulator therefore depends upon the parameters that we use for each type of archival storage device. In this section, we present details of each storage technology's reliability and prospect for future development.

A. Archival Tape

Tape is a popular archival storage medium due to its high capacity, reliability, stability when stored for long periods of time, and good performance on sequential workloads. Its weaknesses are its poor random access performance, the high cost of tape drives compared with tape media, and the length of time it can require to access information.

1) Tape Reliability: One of the main advantages of tape as an archival medium is its ability to cost-effectively store large amounts of data reliably and with minimal need for ongoing maintenance. A 2012 study of the tape archive at the National Energy Research Scientific Computing Center (NERSC), which consisted of 40,000 tape cartridges that were between two and 12 years old, showed a reliability rate of 99.9991% when reading data [17]. NERSC relied upon a single copy of data within its archive, a choice facilitated by tape's high sequential read and write speeds as well as its high reliability as observed in the NERSC archive. The workload on the NERSC archive included a 30% daily read rate, which is much greater than many other archival workloads. While such a workload requires a tape archive to perform as though it were primary storage, it also serves the purpose of quickly discovering any problems that arise in the archive by continuously scrubbing or verifying the archive's data during each read operation, and continuous scrubbing preserves the archive's reliability by verifying that data is readable and uncorrupted. We use the NERSC study for our optimistic tape experiments with a failure rate of 0.0009% over 12 years or 0.000075% annually.

Another study of over 1 million tape cartridges shows that nearly 5% have at least one unrecoverable bit error during their lifetimes while 0.3% have at least 10 unrecoverable bit errors [18]. In our experiments, we set the pessimistic tape reliability to have a 0.3% annualized failure rate. The study finds that removing the least reliable 3% of tape cartridges could significantly improve the reliability of the tape-based archive as a whole. Tape as a storage medium is particularly sensitive to the environment in which it is stored and used; work continues on studying the impact of environmental pressures on the reliability of tape and how environmental conditions should inform the design of tape-based archival systems [19].

2) Tape Development: Tape has been developing consistently since its first introduction as a storage medium. The popular LTO-8 format of tape cartridges features 12 TB of storage capacity and approximately 300 MB per second of read and write throughput [20]. New generations of tape become available every two to three years with each generation of tape drive being able to read tape cartridges that are one or two generations older than it. The LTO Ultrium consortium has published a roadmap for increasing tape cartridge capacity to 192 TB [21] within 10 to 15 years.

B. Hard Disk

Hard disk drives (*HDDs*) are a popular medium for long-term archival storage due to their high capacity, widespread availability, and adaptability to sequential and random-access workloads. The difficulty of using them within archival systems includes their lower reliability compared with tape and recently their lower rates of development for capacity and performance.

1) HDD Reliability: We examined the reliability of hard disk drives using the Backblaze hard drive data set [22]. We analyzed six years of the dataset to measure the observed reliability and life cycle of hard drives in an online backup setting. Our goal for analyzing the Backblaze data is to gather insight on how hard disk reliability may be changing over time. We also observe the way in which trends in hard drive developments affect decisions about device retirement and replacement within Backblaze's server infrastructure. We apply these insights to our simulation model.

We began our data analysis by importing the hard drive statistics available from Backblaze. Next, we removed from the dataset of all of the drives that were active within the Backblaze data center on the first day for which data is available: April 10, 2013. We removed all of these drives from our analysis because we do not know from the information available how many drives failed or were removed before that first day in the dataset, and including them would introduce a bias to our observations. We measured all drives by the date that they were added into the Backblaze storage system, and we counted how many days each drive was active in the system until it either failed or was retired for another reason. A drive was said to have failed when the dataset's marker for failure became true. A drive was said to have been retired after the last day it appeared in the dataset and if the drive was not already marked as failed. Next, we determined how many drives survived for a given number of days. We also calculated how many drives failed or were removed from Backblaze's system after a given number of days. Finally, we determined the daily failure and retirement rates for all hard drives as shown in Figure 1. We also show in Figure 2 the cumulative portion of hard drives that were active, retired, or failed over their lifetimes and based on the calendar year in which the drive was first added to the storage system.

Figure 1 shows the 30-day trailing moving average of the daily hard drive failure and retirement rates. We observe that the failure rate for all drives combined remains stable until

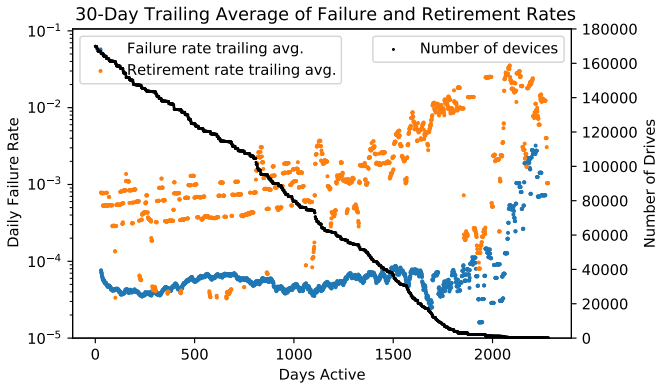

Fig. 1. The first day on the x-axis begins for each drive when it was added to the Backblaze storage system. Failure and retirement rates are shown in percentages. The number of drives over time shows the number that survived in the data center as drive age increased.

TABLE I
NUMBER OF OPERATIONAL DAYS BEFORE REACHING HDD FAILURE RATES

	Drives Added During Year					
%	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
1%	67	143	168	268	332	252
2%	208	332	357	565	589	440
3%	394	503	554	948	745	582
4%	491	684	721	1349	1104	-
5%	618	816	941	-	-	-
10%	1195	-	-	-	-	-
15%	1793	-	-	-	-	-

after the drives have reached approximately 5 years of age. The failure rate begins to increase after five years as the drives reach the end of their warranty periods. In Figure 2, we observe that hard disk failures grew more quickly in 2013 and 2014 than they did in 2017 and 2018. Table I shows the number of days taken to reach different failure rates and separated by the year in which the drive was added to the Backblaze data center. Drives added in 2013 took slightly over two months to reach a 1% failure rate, but drives added in 2018 took approximately eight months to reach the same 1% failure rate. We attribute this observation to the improving reliability of hard disk drives in recent years. We also observe that failure rates remain mostly consistent with time up to five years for drives added within each calendar year. Nevertheless, retirements rather than failures are the dominant reason for the removal of hard disks from the Backblaze data center.

We considered three possible reasons why a hard drive would be removed without failing. First, the hard drive may be part of a model or batch of drives that are likely to fail in the near future. In order to preempt multiple simultaneous drive failures that could result in data loss, system administrators may choose to replace the faulty drives with a more reliable model; however, as shown in Figure 1, we did not observe a high rate of premature drive failures, nor did the pattern of drive retirements follow any sudden increases in drive failures.

A second possible reason for drive retirement is the increasing likelihood of failure as each drive ages, particularly after approximately five years of operation as shown in Figure 1. Even though Backblaze's storage system may have enough redundancy to survive a high rate of drive failures, system administrators may prefer to control the timing of data migration between new and old drives rather than repairing drive failures as they occur. A third possible reason for early drive retirement is the availability of higher capacity drives which, if used to replace older drives, offer much greater capacity while using the same or less power.

2) HDD Development: The ever-increasing capacity and performance of hard drives helps to accelerate the replacement cycle of old drives; however, a reduction in the pace of hard drive development will also reduce the benefits of replacing older drives with new ones. Figure 2 shows a comparison of the portion of active, retired, and failed drives throughout their lifetimes and separated by the year in which each drive was added to the Bacbblaze data center. The portion of active, retired, and failed drives varies over time depending upon which year the drive was added to the storage system. Drives that were added in 2013 and 2014 began to be removed rapidly as they reached 1500 and 1200 active days, respectively; however, the cumulative total of drive failures did not increase dramatically during that time. Drives added in 2015, on the other hand, have been retired much more slowly as they age through 1500 active days compared with those added in 2013 and 2014. Hard drives available in 2018 and 2019 do not offer as compelling of a reason to replace drives from 2015 due to the slow pace at which hard drives have developed between 2015 and 2019; furthermore, the improvement in hard drive reliability supports their continued use as they approach five years of age. We conclude that a slowing growth rate of hard disk capacity will lead to fewer hard disk retirements. Instead, the reliability of hard disks at and beyond five years of age will become increasingly important for disk-based storage systems.

When describing their choices for removing older model hard drives, Bcckblaze confirmed that $8\,\mathrm{TB}$ drives replaced $2\,\mathrm{TB}$ drives [23] and $12\,\mathrm{TB}$ drives replaced $3\,\mathrm{TB}$ drives [24] that were more common in 2013 and 2014 than in 2015. The availability of higher capacity drives and the need for more data capacity within the same data center motivate decisions to upgrade hard drives from older, lower capacity models to newer, higher capacity models. For this reason, the pace of hard disk drive development will, to a large extent, determine the demand for drives within backups and other cold storage systems such as archives.

The International Disk Drive Equipment and Materials Association published a roadmap in 2016 indicating that hard disk drive capacity will continue to scale for years to come. Capacity will increase as manufacturing techniques improve and as existing technologies for increasing density mature [25]. New technologies like Heat-Assisted Magnetic Recording (HAMR) and Heated-Dot Magnetic Recording (HDMR) will improve HDD capacity when they become available; however, since new technologies can be challenging to manufacture reliably at scale, their emergence and the capacity they promise may prove to be uneven and prolonged. As we observe in the Backblaze dataset, significant increases in hard drive capacity motivate the adoption of newer drives. If the availability of higher capacity hard drives becomes increasingly terraced in years to come, we can expect the adoption of new generations of hard drives also to follow an increasingly uneven pattern following the availability of new technologies that increase hard drive capacity.

C. Solid State Disk

SSDs are ideal for demanding workloads due to their low latency and high throughput. SSDs have become prominent for primary data storage where performance is critical and where cost per gigabyte is not a primary concern. SSDs have also benefited from advances in manufacturing and design that lead to ever greater capacity, improving performance, and excellent reliability.

1) SSD Reliability: The reliability of SSDs has been studied within the context of demanding data centers. Meza et al. show that SSD failure rate increases non-monotonically with time and with the amount of data written to the device due to multiple different failure modes that dominate at different times during the lifetime of an SSD [26]. Reliability also varies widely with SSD model and the workload on the drive. Schroeder et al. found that the rate of unrecoverable errors grows linearly with the number of program-erase (PE) cycles across multiple SSD models. Furthermore, newer SSD drives offer similar or better reliability compared with older SSD models notwithstanding the smaller lithographies and additional bits per cell in newer drives [27]. SSDs can trade PE endurance for capacity by increasing the number of bits stored per cell of flash. We assume in our simulations that SSDs have an annual failure rate of 0.58% based on figures published in SSD data sheets [28].

2) SSD Development: SSDs are increasing in capacity over time as their development continues. Over time, declines in the cost of manufacturing each byte of storage dominate the total cost of data storage. SSDs have increased in capacity due to smaller lithographies, by stacking multiple layers of flash cells to form a three-dimensional flash chip, and by increasing the number of bits stored in each cell of flash. Recent additions to the number of bits per cell [29] and the number of layers in each flash chip [30] promise to extend the development of flash-based SSDs into the future.

D. Optical Disc

Optical disc (ODD) is a mature technology that promises durable and scalable archival storage. Optical disc is less sensitive to its long-term storage environment than other archival technologies such as tape [31]. Disc offers a 50 year lifetime for write-once archival storage media [32], and each new generation of optical drives remains compatible with every previous generation of optical media. Future generations of optical disc will triple capacity from $300\,\mathrm{GB}$ to $1\,\mathrm{TB}$ per disc [33]. Still, the slow rate of development, limited number

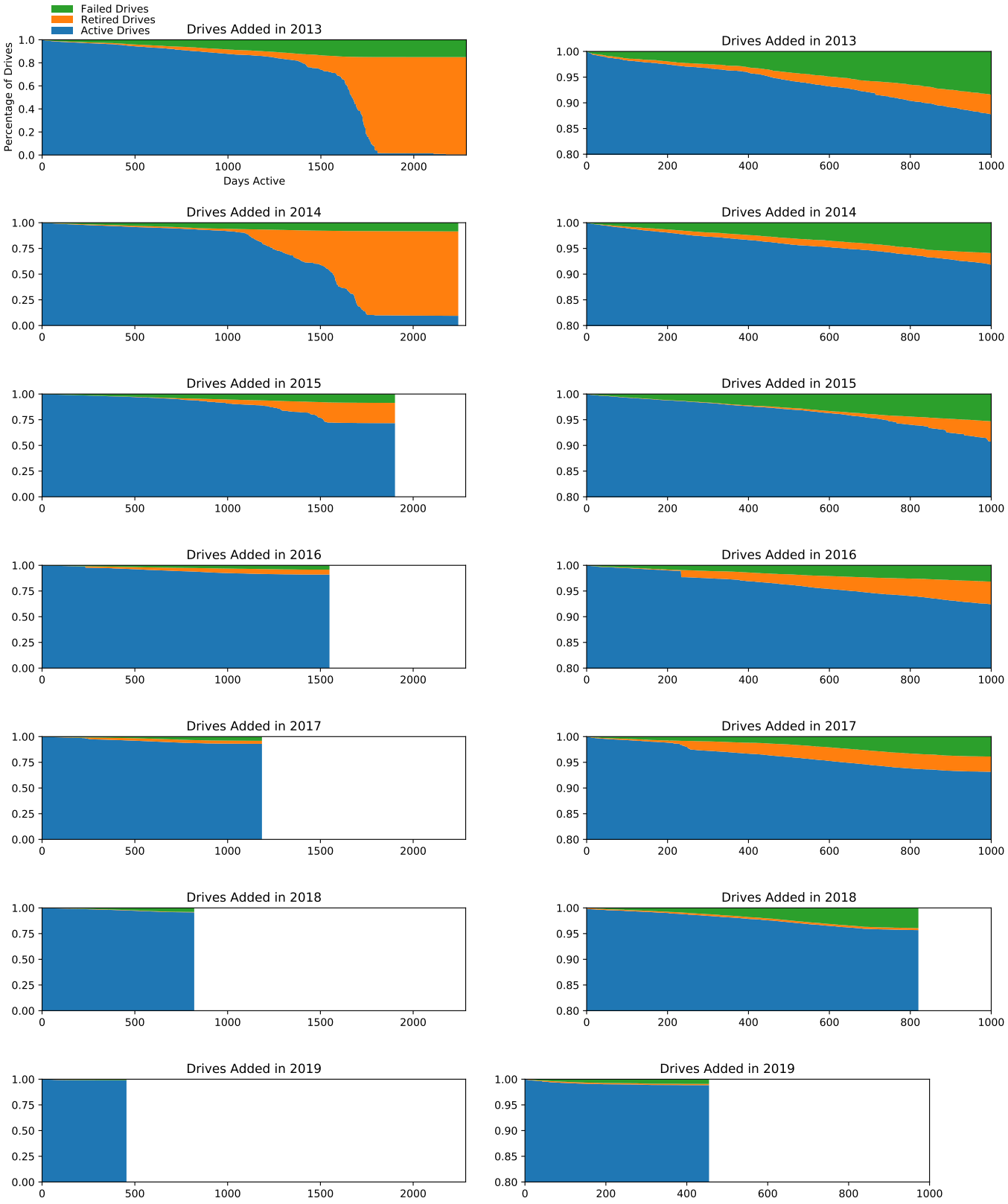

Fig. 2. Figures on the left show data for all drives that were added to the Backblaze data center within the specified calendar year. Figures on the right show the first 1,000 days and top 20% of the data. The x-axis corresponds to the lifespan for each drive.

of vendors, and the shortage of detailed information about long-term cost and reliability present ongoing challenges to optical disc as an archival storage technology. We assume for the purposes of our simulations that optical disc has a reliability equal to that of optimistic tape.

E. Archival Glass

Glass has been proposed as a novel long-term storage technology for use in data centers and archival systems [34]. Glass-based storage utilizes femtosecond lasers to encode data into a multi-dimensional pattern within a small plate of glass. Glass could offer good potential as a storage technology due to the low cost of manufacturing the storage medium, high data density, and excellent reliability. Ongoing work is refining the technology by increasing density and throughput for both read and write operations. Unlike tape and optical disc that have separate storage media and drives, glass-based storage requires separate media, drives for reading, and drives for writing.

F. Archival DNA

DNA has been envisioned as a high-capacity medium for long-term archival data storage [35]. DNA promises storage density that is orders of magnitude greater than existing storage technologies [36]; however, there remain many challenges to implementing a functional storage system that uses DNA as the storage medium [37]. Synthetic DNA requires a synthesizer to encode data within DNA molecules, a repository to store the DNA over a long period of time, and a sequencer to read data from the DNA molecules. Takahashi et al. recently demonstrated an end-to-end storage system with an approximate total cost of $10,000 that synthesizes, stores, and retrieves data using DNA molecules [35]. The principal challenges of utilizing DNA for archival storage remain a high latency for read and especially write operations, the difficulty of encoding data into the language of DNA, the reliability of the storage system, and the cost of materials and equipment.

DNA could potentially store up to exabytes of data per mm^3 [37], potentially over thousands of years [38] if the storage system uses enough redundancy and effective protection from contamination and degradation. Unlike other storage technologies, DNA has been used by biologic organisms to store and transmit information throughout history. DNA does not require migration of data from older to new generations of storage media as time passes because the underlying storage medium remains the same, assuming that the encoding scheme for binary information stored in DNA remains accessible into the future. Given some assumptions about its performance and cost, DNA can be compared with other storage technologies on the basis of its cost to store data over time while achieving a needed amount of reliability.

We expect that DNA-based storage systems will remain in development for years before they become commercially viable; however, we begin our simulations for DNA at the current year in order to easily compare it with other technologies while demonstrating how DNA's cost changes with the target reliability of the storage system.

TABLE II
STORAGE MEDIA CAPACITY AND RELIABILITY

Medium	Capacity	Roadmap	AFR
Tape [17], [18]	12 TB	192 TB	0.000075-0.3%
HDD [22]	10 TB	100 TB	1%
SSD [28]	4 TB	100 TB	0.58%
ODD [32]	0.3 TB	1 TB	0.000075%
Glass [34]	100 TB	360 TB	0.01%
DNA [35]	1 TB	-	1%

VI. EXPERIMENTAL SETUP

We designed our experiments to measure the total cost of using a variety of storage technologies arranged in a variety of RAID configurations over 25 years of operation. We varied the number of data drives in each RAID group, the number of parity drives, and the maximum age of the storage devices before they were retired from the storage system. The values for data drives, parity drives, maximum age, and the storage technology remained unchanged during each simulation, and separate simulations tested different combinations of values and storage devices. We measured the total expected reliability of each storage system by the number of *nines* of reliability that it would not fail over the next year of operation. We used the minimum reliability and maximum blast radius found during each simulation to represent that simulation's RAID configuration, and the cost of each configuration, measured in US dollars, is the cumulative acquisition and energy cost after 25 years. For each storage technology, we plot the minimum cost to reach a minimum of zero through 16 nines of reliability, including RAID configurations that offered more nines of reliability at a lower cost. We calculated cumulative acquisition cost as the sum total cost of all storage devices, read and write drives, rack mounting infrastructure, and any robotics that are needed to operate the archive with a given storage medium. Electricity cost was $0.11 per kilowatt hour, which increased by 1.3% annually. We simulated each storage technology for an archival system with a minimum initial capacity of 1 PB that grew by 30% each year.

We choose reliability and capacity values from manufacturers and other published sources [14]. We assume that the capacity for each storage technology will develop at a rate consistent with its historical rate of development, but we also assume that capacity growth will slow once each technology reaches the end of its development roadmap. We list the baseline values for each storage technology in Table II together with the roadmap capacity for each technology after which subsequent increases in capacity are likely to become more difficult to achieve and hence less significant. Parameters for glass and DNA storage are estimations.

Figure 3 shows the expected growth of each storage technology's future capacity that we use in our simulations; values are normalized to the baseline capacity for each technology as given in Table II. Hard disks and solid state disks frequently increase their capacities as new models incorporate developments and increases in areal bit density for their

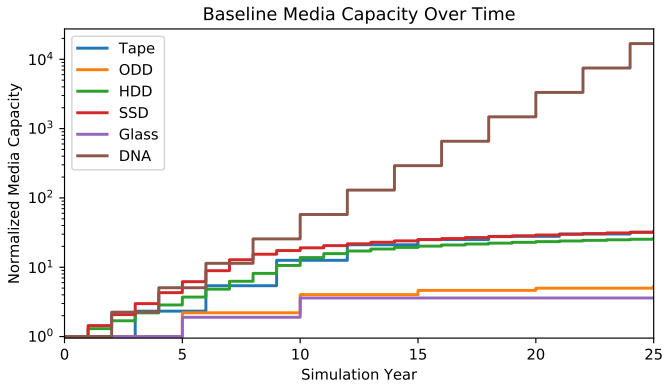

Fig. 3. We model the growth of device capacity as a step function of time. We present here the baseline growth trajectories for each technology, normalized to the starting capacity for each medium. The rates of capacity growth slow once each technology reaches the end of its developmental roadmap.

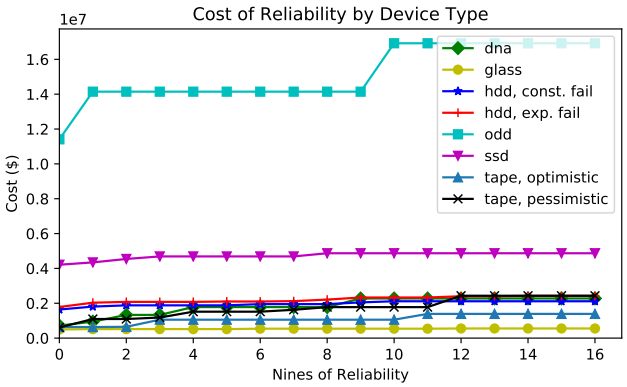

Fig. 4. The minimum cost of achieving different levels of reliability varies by the type of storage used in an archival system. The cost values are expressed as cumulative total for the archive after 25 years of operation.

TABLE III
SUMMARY OF BASELINE PARAMETERS FOR STORAGE DEVICE COST

Type	Media	Drive (Read / Write)	Library	Other
Tape	$150	$8,000	$7,000	$1,000
HDD	-	$200	$10,000	$2,500
SSD	-	$500	$10,000	$2,500
Disc	$10	$10,000	$15,000	$1,000
Glass (est)	$1	$1,000 (r) / $10,000 (w)	$1,000	$1,000
DNA (est)	$100	$1,000 (r) / $9,000 (w)	$1,000	-

storage media. We model their capacities in our simulator with annual increases. Tape and optical disc generally offer more infrequent upgrades in part because new generations of storage media require new and often expensive drives. Tape and optical disc have 3 and 5-year upgrade cycles in our simulations, respectively. We expect that archival glass will provide an upgrade trajectory similar to that of optical disc. We model the capacity growth of DNA without the constraints of other storage mediums that are constrained by their manufacturing and scalability limitations. Archival DNA's capacity will grow with the developments of the technologies used to sequence and synthesize DNA molecules, and therefore, we model a 2-year upgrade cycle for DNA without tapering the pace at which its capacity grows.

VII. EXPERIMENTAL RESULTS

A. Reliability Cost Inflation

Our baseline experiments as shown in Figure 4 were based on the values in Tables II and III. Flat lines between different nines of reliability for each technology indicate that more reliable RAID configurations cost less than some of the less reliable configurations among those that we simulated. We also include separate simulations for hard disks with constant failure rates and exponential failure rates. A constant failure rate for hard disks is an unchanging annual failure rate (AFR) of 1%. Experiments for hard disks with exponentially growing failure rates have a constant AFR of 1% until five years of operation within the archival system, and the failure rate doubles each year thereafter. As we discussed in Section V-A1, experiments for optimistic tape use an AFR of 0.000075% that does not grow over time while our pessimistic experiments for tape use a 0.3% AFR that grows 10% each year.

We observe that optical disc (ODD) costs the most of all storage technologies due to its limited road map of future developments. We study ODD further in Section VII-H. SSDs also have a high cost across the entire range of reliability values because of their high cost per byte of storage relative to other technologies. We explore the cost and reliability of SSDs further in Section VII-E.

Our pessimistic experiments for tape and hard disk both return similar results for the highest levels of reliability. We conclude from this that hard disk and tape are competitive in terms of cost for reliability in archival storage. The low AFR of optimistic tape requires less RAID parity and therefore fewer tape cartridges to reach 16 nines of reliability than pessimistic tape, and the fewer number of tape cartridges and other hardware like tape drives needed to support them results in a total cost that is 43% lower for our optimistic tape results. If the actual reliability of future tape media is better than our pessimistic AFR of 0.3%, we expect that tape will cost less than hard disk at every level of reliability.

Our experiments for archival glass show that it could become a highly cost-effective storage medium, granted that our assumptions about the cost of glass media and drives prove accurate. We explore other possibilities for glass in Section VII-F. Synthetic DNA, on the other hand, is only competitive with existing technologies due in large part to the high cost of materials for each DNA molecule. We further explore DNA in Section VII-G.

We use Formula 5 from Section IV to calculate the blast radius for each storage device. We defined the blast radius to be a function of the failure probability of a RAID group relative to its capacity. A large blast radius indicates a high average probability of losing data when a RAID group fails. As shown in Figure 5, the blast radius varies widely by storage technology and cost. In this experiment, we present the blast

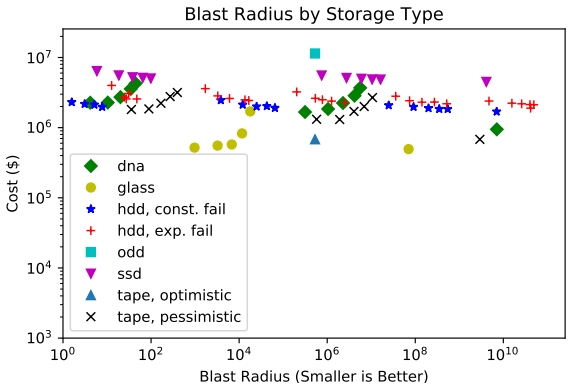

Fig. 5. The total blast radius for an archival system depends on the reliability of the storage devices, the amount of parity in each RAID group, and the capacity of each device.

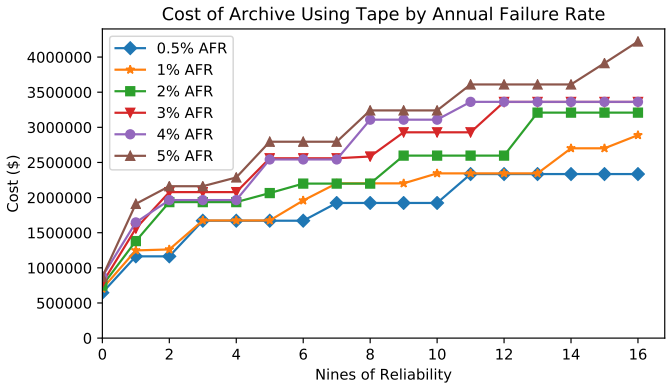

Fig. 6. The cost of reliability for tape increases marginally with the AFR of tape media.

radius for devices with a maximum age of 10 years. Each storage technology has multiple data points since different RAID configurations result in different values for cost, RAID group capacity, and RAID group failure probability.

We observe that archival glass has a relatively large blast radius due to the high capacity of each storage device. For many of the storage technologies like DNA, HDD, and tape, blast radius can be minimized without much additional cost. DNA and optimistic HDD offer both moderate costs and a low blast radius. Increasing the capacity of either HDD or DNA would necessarily increase the blast radius; however, we propose changes to their designs in Section VII-D and Section VII-G that could further reduce their costs while preserving their low blast radius.

B. Cost and Reliability of Tape

We showed in Section VII-A that reducing the AFR of tape media can have a large effect on the total cost of an archival system across a range of reliability values. If, however, the AFR of tape storage media increases as its capacity continues to grow with time, how much more will tape archival systems cost while meeting the same reliability goals? We ran simulations with increased AFR values for tape media from 0.5% to 5%. Figure 6 shows that the total cost of tape-based archival storage grows with higher AFR values for tape media at each level of reliability; however, a tenfold increase in AFR results in less than a doubling of total cost for 16 nines of reliability. We see an 81% higher total cost with a 5% AFR compared with a 0.5% AFR. We therefore observe that the cost of a highly reliable archive using tape increases more slowly than the AFR of tape media. The large difference between our optimistic and pessimistic experiments for tape in Section VII-A reflect the impact on cost of an increasing AFR as devices age. With all other tape experiments using an AFR that grows 10% annually with device age, the optimistic experiments show that a stable storage medium significantly reduces the cost of highly reliable storage because stable old storage devices can remain in the archive without dramatically

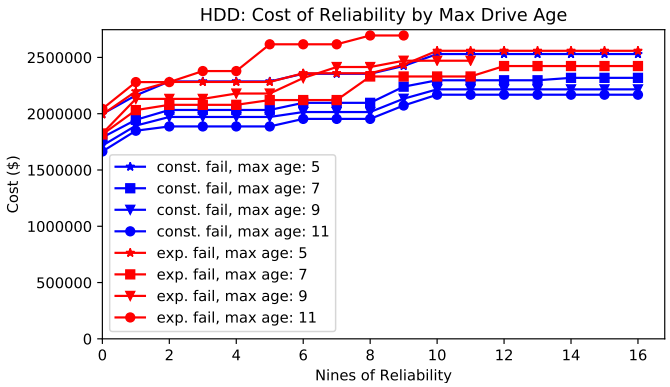

Fig. 7. Hard disks with exponentially growing failure rates cost more to use than drives with constant and unchanging failure rates. Lines that do not extend across the entire x-axis indicate that we found no RAID configuration to reach those higher levels of reliability.

increasing the probability of data loss. We conclude that the stability of the AFR for tape can have a large impact on the cost of reliably storing data over the long term.

C. Hard Disk Reliability

Figure 7 shows results of simulations using two models for hard disk reliability. We compare the constant failure rate with the exponential failure rate for hard disks as described in Section VII-A while also comparing different maximum ages for the drives in the archive. The *max age* is the age at which drives are retired from the storage system and replaced with new drives.

We observe as expected that growing failure rates result in higher costs overall compared with the optimistic case of hard disks that fail with an unchanging AFR of 1%. The most economical option for all levels of reliability with drives that have constant AFRs is to keep the drives for as many years as possible because such old drives would not fail with any greater probability than new drives. For drives with failure rates that increase after five years of operation, keeping the drives for 11 years proves to be the most expensive option.

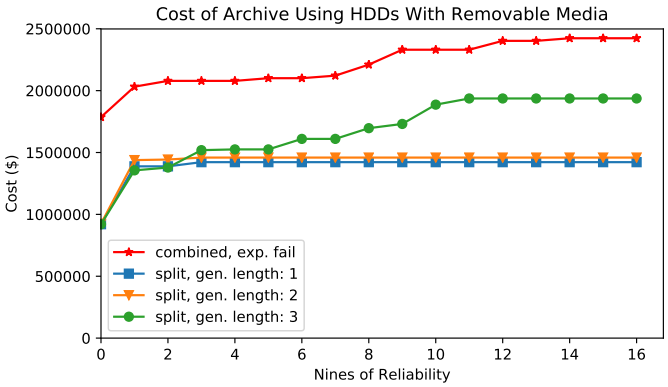

Fig. 8. The cost of hard disks with separable platters in archival systems depends on how often the technology is updated with increased capacity and performance.

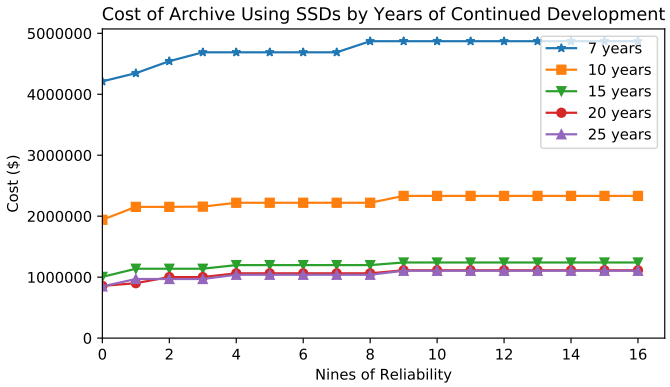

Fig. 9. SSD capacity and development dramatically affect the cost of reliability in archival systems using SSDs.

Instead, it is best to keep drives for approximately seven to nine years in order to minimize the cost of the archive across a range of reliability values while simultaneously extracting as much useful lifetime from each drive as feasible. Keeping the drives for longer reduces the storage system's total reliability so that additional parity must be used to compensate for the increasing failure rate of old drives, and adding more parity drives causes the cost of the system to increase. Constant failure rates for hard drives reduce the cost of an archive with 16 nines of reliability by 10%. We conclude that if hard disks could be made to last at least 10 years instead of five, the cost of constructing a reliable archive using hard disks would decrease accordingly.

D. Hark Disks With Removable Media

Hard disk drives currently have physically combined platters and recording devices. We designed experiments to explore the possible benefits to the cost of reliability in archival storage of separating platters from the HDD recording mechanism. In these experiments, we use the same capacity and failure rate as our other experiments with hard disks. We assume that the platters of the drive by themselves will cost 75% of what a typical hard disk costs and that the archive will use similar mounting infrastructure to a tape library system. Finally, we assume that the read and write mechanism for the removable platters will cost more than a traditional hard drive but less than a tape drive. We use the estimate that the read and write mechanism will cost 10% of what a tape drive costs.

Figure 8 compares the cost of reliability as we vary the time between successive generations of hard disks with removable platters. We also compare traditional hard disks as described in Section VII-A. We observe that separating hard disk platters from their read and write mechanism could cost significantly less than traditional hard disks in highly reliable archival systems, but the amount of the savings depends on how frequently the hard disk technology is updated. Updating the technology every one to three years could save between 42% and 20% compared with the cost for traditional hard disks.

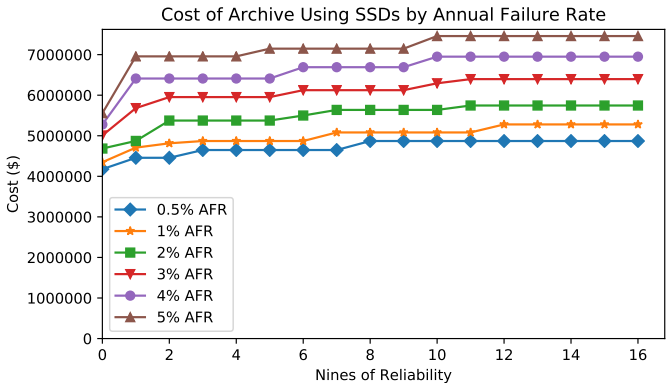

Fig. 10. Higher AFR values have a marginal impact on the cost of reliability in SSD-based archival storage.

We conclude that exploring alternative design possibilities for established technologies like hard disk drives could result in meaningful savings for demanding and reliable archival systems.

E. SSDs for Reliable Archival Storage

Results in Section VII-A showed that SSDs cost more than most other technologies for archival data storage. We explore the effects of increasing SSD capacity by considering the possibility that the current pace of SSD developments will continue further into the future and by examining the effects of increased AFR on the cost of reliable archival storage using SSDs. Figure 9 compares the cost for reliability in archival storage if the development of SSDs continues apace for seven to 25 years. We observe that, as expected, a longer development roadmap, which would result in a lower cost per byte of SSD storage, reduces costs for SSD-based archival storage relative to a shorter development roadmap. We also observe that relatively short extensions of the SSD development roadmap can dramatically decrease the cost of SSD-based archival storage. Extending the development of SSDs apace for 10 to 15 years can reduce the cost of archival storage with 16 nines of reliability by 52% and 75%, respectively.

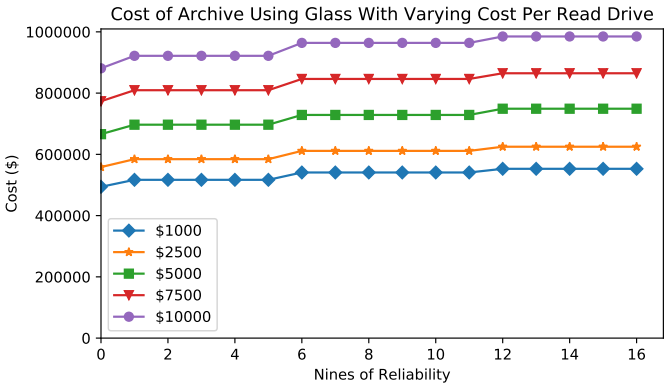

Fig. 11. The cost of storing data in glass increases marginally with the cost of a reader drive.

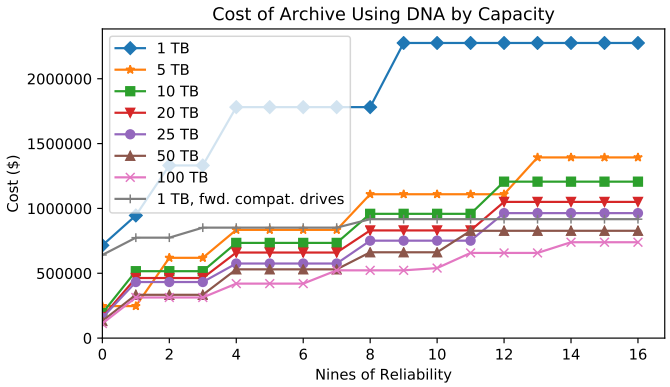

Fig. 12. The cost of reliability in DNA-based archival storage depends upon the capacity and cost of each DNA molecule as well as the forward compatibility of DNA sequencers and synthesizers.

The recent emergence of QLC flash along with continued scaling and stacking of flash layers have increased capacity and reduced the cost of data storage in SSDs, yet such changes may come at the expense of SSD reliability. Some have argued that the lower endurance of novel SSD technology such as QLC outweighs its cost advantage and renders it unsuitable for archival storage [39], but the emergence of denser SSD technology may prove to offer cost advantages over less dense and, by extension, more reliable SSD technologies if the increased density does not prevent archival systems from offering a similar level of reliability at a lower total cost.

What effect would lower SSD device reliability have on the cost of archival system reliability over the long term? Figure 10 shows that large increases in AFR have a relatively small impact on the overall cost of reliability for archival storage with SSDs. Doubling the AFR from 0.5% to 1% increases the cost of archival storage by 8% over 25 years. Even if future developments to SSDs come at the expense of some device reliability, we predict that the increased capacity of such SSDs will notwithstanding make them ever more suitable for archival storage.

F. Archival Glass

Archival glass promises to be a highly reliable storage medium, yet the exact cost of the hardware needed to read and write data into glass remains unknown. We designed experiments to measure the effect of increasing the cost of a drive to read glass from our baseline of $1,000 to $10,000, which is also the cost of the drive to write data in our experiments. We set the cost of media at $1 and its AFR at 0.01%.

As shown in Figure 11, increasing the cost tenfold of a drive for reading data from glass increases the total cost of archiving data in glass by 78% for 16 nines of reliability. The total cost of reliability in glass storage thus increases only modestly with the price of its drive for reading because the intrinsic reliability of the glass medium requires only minimal parity to achieve high levels of reliability. We conclude that glass has an advantage over other technologies due to its low cost and high reliability as a storage medium.

G. Synthetic DNA for Reliable Archival Storage

Synthetic DNA is currently in development, and we do not yet know how its development will proceed or if it is likely to provide the features that we have modeled. Although it is not our goal to predict the cost of individual DNA components and storage devices, we present these results to provide a baseline of performance and cost against which DNA-based archival storage systems can be compared and developed as time passes. We also leave it to continuing work to assess the real performance, cost, and reliability characteristics of DNA-based archival storage systems should they become commercially available.

Our previous experiment in Section VII-A calculated the cost of DNA storage using a baseline of 1 TB per DNA molecule. We explore the effect on cost for reliability of increasing the capacity per molecule and, alternatively, envisioning DNA sequencers and synthesizers that can read and write any molecule of DNA produced in the future. We assume that each DNA molecule costs $100 in materials with an AFR of 1%, the sequencer costs $1,000, and the synthesizer costs $9,000.

Figure 12 shows that cost for each level of reliability decreases as the capacity of DNA increases. Cost for 16 nines of reliability decreases by 39% as capacity increases from 1 to 5 TB and by 68% with a capacity of 100 TB; however, enabling sequencers and synthesizers to read and write DNA molecules created with future generations of DNA technology reduces the total cost by 60% compared with our baseline that does not support forward compatibility. We conclude that flexibility in the design of DNA storage systems can help to dramatically reduce their cost for reliable archival storage.

H. Cost of Preserving Fixed Amount of Data

The demand for new advancements in hardware reflects the presumption that the demand for data storage is growing. If an archival system stores a fixed amount of data over a

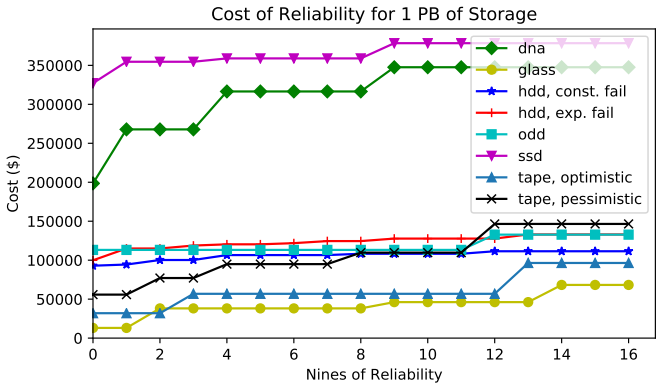

Fig. 13. The cost of reliably storing 1 PB of data favors devices that offer high reliability.

long period of time, then the controlling factor in the cost of the archival system becomes the initial acquisition cost of the system and the reliability of the storage devices within it. Figure 13 shows the total cost of an archival system implemented with each storage technology to preserve the same data over 25 years without adding to or modifying the data.

In this experiment, optical disc proves to be competitive with both tape and hard disk, particularly if we assume that hard disks exhibit an increasing probability of failure as they age. The stability and reliability of optical media also offsets its limited prospects for development. Synthetic DNA and glass differ dramatically in terms of cost because of the disparity in our assumptions about the costs of their storage media, and our assumptions about the higher AFR of DNA compared with glass cause DNA to increase in cost more than glass as the storage system provides more nines of reliability.

VIII. Conclusion

As existing storage technologies continue to develop and new storage technologies emerge, archival systems must adapt to both the demands of the storage market and to the availability of storage technology. Existing storage technologies like hard disk, solid state disk, and tape may provide good reliability and capacity scaling, but prospective technologies like archival glass and synthetic DNA seem necessary in order to economically preserve ever-increasing amounts of digital information.

We have run simulations that show how the cost of archiving data changes using existing and prospective storage technologies while planning for a range of different reliability values. We showed that increasing the lifetime of hard disks can reduce archiving costs by as much as 10%. Hard disks with separable platters could also reduce the cost for a highly reliable archive by 20% to 42% relative to traditional hard disks. SSDs, which today cost more per byte of storage than other technologies, can become more economical for use in reliable archival systems by increasing capacity, even at the expense of some reliability as they develop. Tape can also remain viable for archival storage even if its failure rate increases with successive generations of tape technology; however, to compete effectively with glass for highly reliable archival storage, tape must continue to prove its reliability and longevity as it develops. We showed that archival glass is a promising and likely cost-effective archival technology due to the stability and low cost of glass as a storage medium. Finally, we showed that flexibility in the design of DNA sequencers and synthesizers can lead to as much of a reduction in the cost of reliable DNA archival storage as a large increase in the capacity of each DNA molecule.

Acknowledgments

We would like to thank the anonymous reviewers for their helpful comments on this paper. We also thank our colleagues in the Center for Research in Storage Systems for their insightful contributions and support.

References

[1] D. A. Patterson, G. Gibson, and R. H. Katz, "A case for redundant arrays of inexpensive disks (RAID)," in *Proceedings of the 1988 ACM SIGMOD International Conference on Management of Data*. ACM, 1988, pp. 109–116. [Online]. Available: http://www.ssrc.ucsc.edu/PaperArchive/patterson-sigmod88.pdf

[2] J. Pâris, S. J. T. Schwarz, and D. D. E. Long, "Improving disk array reliability through faster repairs (extended abstract)," in *2016 IEEE 35th International Performance Computing and Communications Conference (IPCCC)*, Dec. 2016, pp. 1–2.

[3] J. G. Elerath and M. Pecht, "Enhanced reliability modeling of RAID storage systems," in *Proceedings of the 2007 Int'l Conference on Dependable Systems and Networking (DSN 2007)*. IEEE, Jun. 2007, pp. 175–184. [Online]. Available: http://www.ssrc.ucsc.edu/PaperArchive/elerath-dsn07.pdf

[4] J.-F. Pâris, T. Schwarz, A. Amer, and D. D. E. Long, "Protecting RAID arrays against unexpectedly high disk failure rates," in *Proceedings of the IEEE 20th Pacific Rim International Symposium on Dependable Computing (PRDC '14)*, Nov 2014, pp. 68–75.

[5] J.-F. Pâris and D. D. E. Long, "Using device diversity to protect data against batch-correlated disk failures," in *Proceedings of the Second ACM Workshop on Storage Security and Survivability*. New York, NY, USA: ACM, 2006, pp. 47–52. [Online]. Available: http://doi.acm.org/10.1145/1179559.1179568

[6] S. A. Weil, S. A. Brandt, E. L. Miller, D. D. E. Long, and C. Maltzahn, "Ceph: A scalable, high-performance distributed file system," in *Proceedings of the 7th Symposium on Operating Systems Design and Implementation (OSDI '06)*, Nov. 2006. [Online]. Available: http://www.ssrc.ucsc.edu/Papers/weil-osdi06.pdf

[7] Y. Li, D. D. E. Long, and E. L. Miller, "Understanding data survivability in archival storage systems," in *Proceedings of the 5th Annual International Systems and Storage Conference (SYSTOR '12)*, Jun. 2012.

[8] P. Carns, K. Harms, J. Jenkins, M. Mubarak, R. Ross, and C. Carothers, "Impact of data placement on resilience in large-scale object storage systems," in *2016 32nd Symposium on Mass Storage Systems and Technologies (MSST)*, May 2016, pp. 1–12.

[9] H. Zhu, P. Gu, and J. Wang, "Shifted declustering: a placement-ideal layout scheme for multi-way replication storage architecture," in *Proceedings of the 22nd Annual International Conference on Supercomputing, ICS 2008, Island of Kos, Greece, June 7-12, 2008*, P. Zhou, Ed. ACM, Jun. 2008, pp. 134–144. [Online]. Available: https://doi.org/10.1145/1375527.1375549

[10] V. Venkatesan, I. Iliadis, C. Fragouli, and R. Urbanke, "Reliability of clustered vs. declustered replica placement in data storage systems," in *2011 IEEE 19th Annual International Symposium on Modelling, Analysis, and Simulation of Computer and Telecommunication Systems*, Jul. 2011, pp. 307–317.

[11] J. P. Hughes, "Economics of information storage: The value in storing the long tail," in *2019 35th Symposium on Mass Storage Systems and Technologies (MSST)*. IEEE, 2019, pp. 185–192.

[12] D. S. H. Rosenthal, D. C. Rosenthal, E. L. Miller, I. F. Adams, M. W. Storer, and E. Zadok, "The economics of long-term digital storage," in *The Memory of the World in the Digital Age: Digitization and Preservation*, Sep. 2012. [Online]. Available: http://www.ssrc.ucsc.edu/Papers/rosenthal-unesco12.pdf

[13] I. Iliadis, J. Jelitto, Y. Kim, S. Sarafijanovic, and V. Venkatesan, "ExaPlan: Efficient queueing-based data placement, provisioning, and load balancing for large tiered storage systems," *ACM Transactions on Storage*, vol. 13, no. 2, pp. 17:1–17:41, May 2017. [Online]. Available: http://doi.acm.org/10.1145/3078839

[14] J. Byron, D. D. E. Long, and E. L. Miller, "Using simulation to design scalable and cost-efficient archival storage systems," in *Proceedings of the 26th IEEE International Symposium on the Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS 2018)*, Sep. 2018.

[15] P. Gupta, A. Wildani, D. Rosenthal, E. L. Miller, I. Adams, C. Strong, and A. Hospodor, "An economic perspective of disk vs. flash media in archival storage," in *Proceedings of the 22nd International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS '14)*, Sep. 2014. [Online]. Available: http://www.ssrc.ucsc.edu/Papers/gupta-mascots14.pdf

[16] D. Rosenthal, "Economic models of long-term storage," Feb. 2019. [Online]. Available: https://blog.dshr.org/2019/02/economic-models-of-long-term-storage.html

[17] Active Archive Alliance, "NERSC exceeds reliability standards with tape-based active archive," Feb. 2012. [Online]. Available: https://www.nersc.gov/assets/pubs_presos/AAA-Case-Study-NERSC-FINAL2-6-12.pdf

[18] MP Tapes, Inc., "Reliability of magnetic data tape," Nov. 2017. [Online]. Available: https://mptapes.com/Reliability/reliability.html

[19] J. Adrian and G. Goins, "Accellerated tape durability testing," Oct. 2018. [Online]. Available: https://drive.google.com/file/d/1pPTXKDo7_1tvG_qLC1UmhbxIKsAdAMQj/view

[20] Hewlett Packard Enterprise, "HPE LTO ultrium cartridges," 2018. [Online]. Available: https://psnow.ext.hpe.com/doc/PSN34648USEN.pdf

[21] LTO Consortium, "LTO program outlines generation 8 specifications and extends technology roadmap to 12th generation," Oct. 2017. [Online]. Available: https://www.lto.org/2017/10/lto-program-outlines-generation-8-specifications-extends-technology-roadmap-12th-generation/

[22] Backblaze, Inc., "Hard drive data and stats," Jun. 2019. [Online]. Available: https://www.backblaze.com/b2/hard-drive-test-data.html

[23] A. Klein, "Hard drive stats for Q3 2016: Less is more," Nov. 2016. [Online]. Available: https://www.backblaze.com/blog/hard-drive-failure-rates-q3-2016/

[24] ——, "Hard drive stats for Q3 2018: Less is more," Oct. 2018. [Online]. Available: https://www.backblaze.com/blog/2018-hard-drive-failure-rates/

[25] The International Disk Drive Equipment and Materials Association, "ASTC technology roadmap," 2016. [Online]. Available: http://idema.org/wp-content/plugins/download-monitor/download.php?id=2456

[26] J. Meza, Q. Wu, S. Kumar, and O. Mutlu, "A large-scale study of flash memory failures in the field," in *Proceedings of the 2015 SIGMETRICS Conference on Measurement and Modeling of Computer Systems*, Jun. 2015. [Online]. Available: http://www.ssrc.ucsc.edu/PaperArchive/meza-sigmetrics15.pdf

[31] Sony Corporation, "Optical disc archive, generation 2: White paper," Apr. 2016. [Online]. Available: http://assets.pro.sony.eu/Web/ngp/pdf/optical-disc-archive-generation-two.pdf

[27] B. Schroeder, R. Lagisetty, and A. Merchant, "Flash reliability in production: The expected and the unexpected," in *Proceedings of the 14th USENIX Conference on File and Storage Technologies (FAST '16)*, Feb. 2016, pp. 67–80. [Online]. Available: http://www.ssrc.ucsc.edu/PaperArchive/schroeder-fast16.pdf

[28] Samsung Electronics, "Samsung v-nand ssd: 860 evo," Dec. 2017. [Online]. Available: https://www.samsung.com/semiconductor/global.semi.static/Samsung_SSD_860_EVO_Data_Sheet_Rev1.pdf

[29] C. Mellor, "Qlc flash is tricky stuff to make and use, so here's a primer," Jul. 2016. [Online]. Available: https://www.theregister.co.uk/2016/07/28/qlc_flash_primer/

[30] ——, "Wd shoots out 96-layer embedded flash chips," Oct. 2018. [Online]. Available: https://www.theregister.co.uk/2018/10/18/wds_96layer_embedded_flash_chips/

[32] Sony Corporation and Panasonic Corporation, "White paper: Archival disc technology," Jul. 2015. [Online]. Available: https://panasonic.net/cns/archiver/pdf/E_WhitePaper_ArchivalDisc_Ver100.pdf

[33] D. Robb, "Is optical disc an illusion?" Jan. 2016. [Online]. Available: https://www.enterprisestorageforum.com/storage-technology/is-optical-disc-an-illusion.html

[34] P. Anderson, R. Black, A. Cerkauskaite, A. Chatzieleftheriou, J. Clegg, C. Dainty, R. Diaconu, A. Donnelly, R. Drevinskas, A. Gaunt, A. Georgiou, A. G. Diaz, P. G. Kazansky, D. Lara, S. Legtchenko, S. Nowozin, A. Ogus, D. Phillips, A. Rowstron, M. Sakakura, I. Stefanovici, B. Thomsen, and L. Wang, "Glass: A new media for a new era?" in *10th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 18)*. USENIX Association, July 2018, pp. 1–6. [Online]. Available: https://www.microsoft.com/en-us/research/publication/glass-a-new-media-for-a-new-era/

[35] C. N. Takahashi, B. Nguyen, K. Strauss, and L. Ceze, "Demonstration of end-to-end automation of dna data storage," *Nature Scientific Reports*, vol. 9, March 2019, article number: 4998 (2019). [Online]. Available: https://www.microsoft.com/en-us/research/publication/demonstration-of-end-to-end-automation-of-dna-data-storage/

[36] S. Newman, A. Stephenson, M. Willsey, B. Nguyen, C. N. Takahashi, K. Strauss, and L. Ceze, "High density dna data storage library via dehydration with digital microfluidic retrieval," *Nature Communications*, vol. 9, April 2019, (2019)10:1706. [Online]. Available: https://www.microsoft.com/en-us/research/publication/high-density-dna-data-storage-library-via-dehydration-with-digital-micro%ef%ac%82uidic-retrieval/

[37] L. Ceze, J. Nivala, and K. Strauss, "Molecular digital data storage using dna," *Nature Reviews Genetics*, May 2019. [Online]. Available: https://www.microsoft.com/en-us/research/publication/molecular-digital-data-storage-using-dna/

[38] W. D. Chen, A. X. Kohll, B. Nguyen, J. Koch, R. Heckel, W. J. Stark, L. Ceze, K. Strauss, and R. N. Grass, "Combining data longevity with high storage capacity—layer-by-layer dna encapsulated in magnetic nanoparticles," *Advanced Functional Materials*, May 2019. [Online]. Available: https://www.microsoft.com/en-us/research/publication/combining-data-longevity-with-high-storage-capacity-layer%e2%80%90by%e2%80%90layer-dna-encapsulated-in-magnetic-nanoparticles/

[39] H. Smith, "Using QLC for cold storage is a fool's errand," Sep. 2019. [Online]. Available: https://blocksandfiles.com/2019/09/27/using-qlc-for-cold-storage-is-a-fools-errand/