Effects of Prolonged Media Usage and Long-term Planning on Archival Systems

Preeti Gupta*, Avani Wildani†, Ethan L. Miller*, David S.H. Rosenthal‡, and Darrell D.E. Long*
*University of California, Santa Cruz †Emory University ‡Stanford Libraries

***Abstract*—In archival systems, storage media are often replaced much earlier than their expected service life in exchange for other benefits of new media, such as higher capacity, bandwidth, and I/O operations per second, or lower costs. In an era of decreasing media density growth rates, retiring media early by considering only short-term benefits while discarding potential long-term cost benefits could have a negative long-term impact on an archival system's economics. To extend an archival system's life, at low cost, while limiting performance degradation, we suggest extending media lifetime past manufacturer recommendations as well as increasing the horizon for planning and provisioning future media purchases.**

We present a cost-benefit analysis of the impact of prolonged media usage and long-term planning. Through Monte Carlo simulation, we simulate the behavior of an archival system using tapes, hard disk drives (HDDs), solid state devices (SSDs), and Blu-ray discs. We show that leaving older media in the archival system makes economic sense for SSDs without significantly affecting reliability; we show cost improvements of approximately 10% for SSDs for a low annual media density growth rate, such as 5%, which would have been a loss of 35%, for a high annual media density rate, such as 20%. We show that, for SSDs and hard disks, the optimal planning time of an archival system is at least as long as the media service life. Combining prolonged media usage with an extended planning horizon reduced costs by 15% for a system using SSDs.

I. Introduction

In our current "big data" environment, data is growing by almost 60% per year [1]. The community is finding new incentives, such as longitudinal studies and compliance, to retain data indefinitely and realizing the value that can be extracted from the archived data. However, organizations find it extremely hard to finance storing this data in an efficient way [1], particularly due to regular expenses incurred on maintenance. Maintenance costs include the costs of replacing failed devices, integrity checking, purchasing new media, labor, and other data center costs. These difficulties, in addition to various uncertainties involved with long-term financial calculations, such as unpredictable media density growth rates, media failure, interruptions in the supply chain, and interest rate fluctuations, lead system designers to look for short-term benefits. Real life examples of archives with extended media retention policies or long term planning horizons are rare. This work focuses on exploring the economic and performance impact of long-term planning and prolonged media usage.

An accurate model of storage costs is essential when planning for the long-term survival of data. As devices, workloads, and data center conditions, shift over time, the model must coarsely predict these changes to determine the initial and ongoing costs of the system. In this paper, we use a Monte Carlo model to project the cost of a realistic archival system that includes a replication and failure model and stores a fixed size data set of 1 PB. We show the effect of varying the planning horizon and prolonged media usage on cost, bandwidth, energy consumption, and the reliability of the system. We considered SSD, hard disk, Blu-ray, and tape to evaluate the effect of extending the planning horizon for all media primarily used or suggested for archival purposes.

Storage costs are approximately one third of the total system's cost [2], [3], and they are going to be more important in the future as the annual media density growth rate (Kryder's rate, or K_r) continues to decrease [3], [4]. In times of highly increasing media density growth, it was cost-effective to replace media before the manufacturer-supplied end-of-life, and not much thought was given to the planning duration or using media for an extended period of time. However, as media reach their technical limits, and media density growth slows down, or stops [3], [4], the total costs of preservation would increase further. Given the low cost or capacity benefit new media will give, and the high value of replaced media being discarded, frequent media replacement may not be cost-effective. There is a need to revisit the existing approach for media retirement and selection.

Though new devices are almost always more efficient in an important dimension, we argue that there may be some economic advantage in leaving older devices in the system. As media is used past the point of maximal durability and planning horizons are extended, device reliability will decrease. However, decreased system reliability may be an acceptable trade-off for a lower total cost or greater risk tolerance in our economic model. For this work, we expand the traditional definition of *data loss* beyond technical failures to include the loss of availability that results when it is no longer economically viable for a system to maintain the data. There have been several instances of archives closing [5]–[7], or storage providers changing the storage pricing model in response to unforeseen conditions [8].

Storage media typically has a manufacturer-recommended *service life*. Consumer disks and SSDs average five years. They can typically last longer, but they are not guaranteed to do so. These numbers are unsurprisingly conservative; decades after their introduction, the manufacturer-suggested-service-life of hard disks, tapes, and optical disks has remained constant. *Prolonged media usage* means using media beyond their manufacturer-recommended service life.

The *planning horizon* represents the time across which total costs of maintaining the archival system (including purchase, operational, and maintenance costs) are calculated [9]. Selection of the planning horizon is important, because it is the width of the window through which the economic

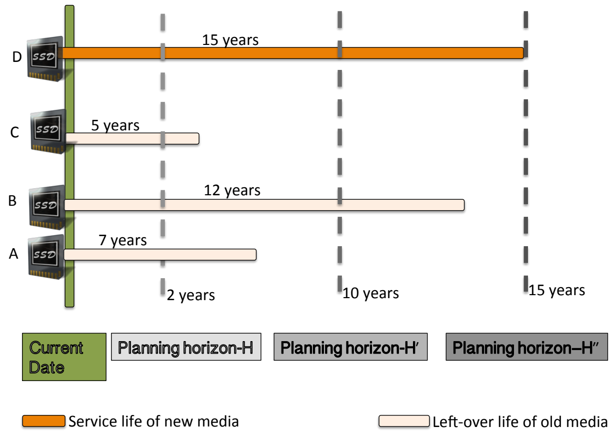


Fig. 1: Example to show the effect of the planning horizon. D is the new device, while A, B, and C are old devices. Depending on the cost-per-byte over the planning horizon, the old device may or may not get replaced by the new device. If the devices are SSDs, all A, B, and C would be replaced by D for H (the decision would be different for H′ and H″), given the high residual value a new device would give to the system (discussed in more detail in Section III-C).

performance of alternatives should be considered. In storage, a short planning horizon means considering a device's value for a short period of time; its leftover value is considered as residual value for the planning horizon. Considering short-term costs, or a short planning horizon, is a low risk situation, due to fewer uncertainties involved but could possibly lower the long-term benefits.

Not much attention has been paid to researching planning horizons in archival systems, because premature media replacement was traditionally thought to be cost-effective. Figure 1 shows a minimal example of the functionality of horizon. A planning horizon is widely used in many industries, such as manufacturing, finance, and insurance [9]. For example, in the wealth management industry, retirees operate at sub-annual planning horizons, whereas a planning horizon of five years makes sense for a younger person who is more likely to be looking for long-term capital gains and open to taking more risks.

Obtaining maximal value from the storage media is important for archival systems because of various hardships they face, such as limited funds. Previously, media densities were low, and transferring data from one device to another device did not cost much. Today, media densities are high, transferring data from one media to another incurs heavy costs, media density growth is slowing down, and archival media is getting popular. In such environments, it makes sense to consider the options of long-term media usage. Our argument is based upon the assumption that archival media will have a long service life, and that such media can be used in archival systems for a long period of time, without much impact on reliability. Anderson had suggested that designing archival hard disks having long service life is technically feasible [10] but increases manufacturing cost.

Using our model, we show that it is cost-effective to use SSDs beyond their suggested service life without significantly affecting reliability, power consumption, and bandwidth. We observed cost improvements of up to 10% with media life extensions of up to five years for SSDs for a low annual media density growth rate, such as 5%, which would have been a loss of 35% for a high annual media density growth rate, such as 20%.

Extending media service life also resulted in a 10× improvement in system bandwidth without increasing costs, which is highly required for active archives. Extended planning horizons were cost-effective for SSDs, saving almost 6% on total costs, which would have resulted in 9% higher costs for a 20% annual media density growth rate. For hard drives, we saw an almost 4% drop in total costs with extended planning horizons. Combining extended media life with long-term planning helped save 15% for SSDs with a five year service life extension. For Blu-ray and tape, we did not see significant cost benefits by planning long-term or extending media service life.

The contributions of the paper are fourfold:

- We present an argument for extending media usage and the planning horizon in archival systems.
- We show that prolonged media usage for SSDs could help to minimize the costs while significantly improving bandwidth, with no significant effects on the system's reliability.
- We show that extended usage of hard disks, Blu-ray, and tape does not help the archival system economically.
- We show that the optimal planning horizon for SSDs and hard disks is at least as long as the media service life.

The rest of the paper is structured as follows. In Section II, we discuss the background and related work. Section III talks about the methodology. Section V shows the default parameters for archival systems, devices, and media readers and writers, and experiments we did to support our hypothesis. Sections VI, VII, and VIII provide discussion, future work, and conclusion, respectively.

II. Background and Related Work

This section provides necessary background and discusses related work.

A. *Cost Modeling*

A substantial amount of work has been done exploring cost models for digital preservation. CMDP [30], Chapman [31], LIFE (Life Cycle Information for e-Literature) [32], [33], the Prestoprime Project [34], KRDS (Keeping Research Data Safe project) [35], TCP [36], and ENSURE [37] are some of the efforts to model long-term costs. These projects primarily focus on abstract cost models for digital preservation that do not consider storage costs or take into account the overhead of ingestion and migration.

Li provided a cost model and hybrid storage system (HDD and SSD) to study different workloads and evaluate their costs and performance on their storage system [29]. They did so with one SSD and one HDD device; though they said the system can be made to use more devices. They did not consider a planning horizon or using media beyond their service life.

B. System Failure Modeling

Schroeder et al. [11], Backblaze [12], and Pinheiro et al. [13], document annual failure rates of up to $11-16\%$ by the sixth year of a hard disk's life. Anderson et al. [14] showed that the longer the hard disk is running, the higher its annual failure rate. Archival systems need to be rescanned frequently to replace expired/failed media, decide whether devices are cost efficient or should be replaced with new devices, and perform data scrubbing. A study by Seagate [14] showed that hard drives can be engineered for an extended service life with small incremental costs.

For SSDs, since the damage is proportional to the number of reads and writes, the service life should be longer in archival system settings [15], [16]. Disk manufacturers typically warrant enterprise hard drives for a 5-year service life [24]–[27]. Unlike HDDs, the failure rate of SSDs remains fairly constant as they age [17]–[19], the bits on the platters persist for much longer [28], and they can be used for an extended period of time [20], [21].

A recent Facebook study [22] showed that SSDs using flash memory do not degrade monotonically as flash wears but go through several distinct periods, depending on when failures emerge and are detected, and the effect of read disturbance is not the predominant source of errors in SSDs. This study was done on the use of SSDs for hot data, which lends support to our claim that extending service life of SSDs should be cost efficient in an archival setting. Schroeder et al. [23] showed that SSDs' raw bit error rate grow at a much slower rate with wear out than the commonly assumed exponential rate, and as compared to hard drives. SSDs have much lower replacement rate in the field, however; they have a high rate of uncorrectable errors, requiring replication for data protection.

A study by Seagate [14] showed that engineering drives for an extended service life would impose small incremental costs but that the market would not bear those costs. Before 2010, 5-year-old drives consumed too much space and power to be economically viable, and would thus be replaced even if their service life had not expired. Thus, the service life is an economic rather than technical parameter. Moreover, even when designing more robust hardware aimed at the archival market, manufacturers are hesitant to guarantee longevity. For instance, Seagate's ST8000AS0002 [24] is designed for extended archival but carries only a 3-year warranty.

Hard disks and SSDs with a ten year service life are not available today, and some even argue that it is hard, if not impossible, to extend life of hard disks and solid state devices economically beyond five years [14], [20]. However, given the media density growth slowdown and the effort extended in the direction of designing archival media, it is important to explore the effect that keeping old media could have on the archival system.

C. The Planning Horizon

In the context of archival systems, the planning horizon represents the period we consider while choosing whether to replace an old device. The costs of supporting the archival system are calculated **only** over that duration. For example, consider a device with purchase cost p, annual operational costs of c, and a service life of y years. If the planning horizon of the archival system is h years, and $h \leq y$, the residual value of the device is $((y-h)/y) \times p$. If we set r as the discount rate [1] of future expenditures, the operational cost discounted over k years for this device is $c/((1+r)^k)$. Combining these concepts, we can derive the cost of buying and keeping a device over the planning horizon (Equation 1).

$$p + \sum_{k=0}^{h-1} \frac{c}{(1+r)^k} - \frac{(y-h)}{y} \times p \times \frac{1}{(1+r)^{(y-h)}} \quad (1)$$

For horizons greater than a device's service life, the device's repurchase cost needs to be considered as well. The leftover value of a device represents the value it leaves the system with, at the end of the planning horizon. The longer the planning horizon the riskier it is in business calculations due to the presence of multiple uncertainties, such as interest rate fluctuations and interruptions in supply chain. For these reasons, short planning horizons are preferred. Our work focuses on presenting the concept of a planning horizon in storage environments, discussing the need of extending it, implementing the methodology, and exploring the economic and performance effects of planning horizons in addition with prolonged media usage in a realistic environment.

III. Model

We built a Monte Carlo model to estimate the financial trade-offs in disk purchase and replacement in a fixed size storage system. The model tracks the money spent on devices using the selected initial state for each run. Inputs to the model are the amount of the data to be preserved, the duration of the preservation, the desired replication level, the planning horizon, the extension period of device service life in the model, media and drive characteristics, if required, annual discount rate, and the cost of power. The model outputs statistics of the archival system built, such as the total number of devices, total cost, bandwidth, power consumption, reliability, failed devices per year, and replaced devices. We explored the effect of media life extension and extending the planning horizon on the cost, energy consumption, bandwidth, and the reliability of the system.

A. System Design

The archival system maintains an inventory of all devices holding data. *Device groups* of identical devices are purchased annually and added to the inventory to meet the storage demand. Each year, the inventory is revisited to check the cost-effectiveness of existing devices. We are considering an archival system with homogeneous media (one media type per simulation) for this work. In the future, we plan to consider heterogeneous systems as well.

If a device has failed, a replacement device is purchased immediately, and data is rebuilt from other devices in the device group. Data read costs are paid for the devices being

[1] The discount rate is an economic measure of the amount of interest paid as a percentage of balance at the end of an annual period. Including discount rate in our calculations accounts for future fluctuations in monetary value.

read for this operation and writing cost for the replacement device. In the unlikely case that the number of overlapping failure events in a device group exceeds the replication level, the data cannot be rebuilt and is lost, and the device group is retired. If a device has expired, it gets replaced, and the data is transferred to new device. Reading (to get the data out) and writing costs (to wipe the data) are paid for the expired device, and writing costs are paid for the new device.

As device groups decrease in size from attrition, they are merged to give more protection against device failures, and read and write costs are paid for each device involved in this operation. If after merging the number of devices in a device group falls below the minimum system replication level, device groups are re-arranged until the replication level is met, and data is re-balanced amongst the devices in the device group. Writing costs are paid for the devices updated in the operation. Consolidating data on denser devices may cause an imbalance where throughput requirements are not met any longer, and future work in this area should investigate the tradeoff between reliability and system bandwidth.

For all media in the model, areal density increases by a function of the annual media density growth rate (Kryder's rate, or K_r) each year. Disk and optical-media bandwidth grows by the square root of K_r each year [38]. For SSDs and tapes, bandwidth grows linearly with K_r [38]. Our future-work would need to consider non-linear bandwidth improvement. Once purchased, a device takes a specific capacity and bandwidth. K_r is constant for one run. We did not use varying K_r to explore the effect of different policies on individual K_r.

B. Cost Model

The *capital cost of a device* (CC) represents the device purchase cost. *Operating costs* (OC) represent the cost of operating the device/drive, including cooling, power, and physical space. *Data refresh costs* (DR) represent the cost of migrating the data, and are calculated by the capacity, bandwidth, and power consumption at maximum throughput. The *total cost of ownership of a device* (TCO) represents the capital cost plus its operating cost over the service-life plus the cost to move data onto the device, shown in Equation 2. TCO for drives represents capital costs and operational costs over the service-life. Move out costs of a device are paid at the time of its retirement, while move in costs are paid at the time of purchase. To calculate migration costs for removable media such as tapes or optical disks, the power consumption and bandwidth of their drives is used. To calculate read/write costs, the active power of the devices/drives, mentioned in II, is used. For the rest of the year when device/drive was not being used for reading or writing, the idle power of the device/drive is used to calculate the operational cost.

$$TCO = CC + OC + DR \tag{2}$$

The annual cost of maintaining the archival system is obtained by adding the expenditure made on maintaining the devices and drives (purchase, operational, and data refresh costs). Annual costs for all the years are added, after discounting them to the present, to give the estimate of the cost of preserving the data for 100 years. This estimate could be used for comparisons between different operational and policy decisions that could be taken for data preservation.

C. Device Replacement and Planning Horizons

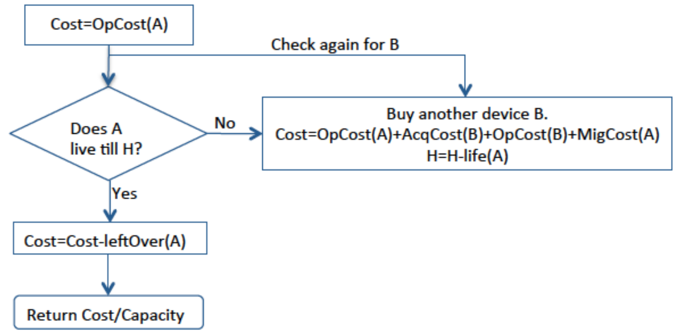


Fig. 2: The device selection policy. The costs of following *pOld* and *pNew* are calculated using this logic. The cost of each device is divided by the weighted average of the capacities of devices purchased on each path to give cost-per-byte.

Figure 2 shows the device selection policy. *OpCost(A)* represents operational costs of a device **A** calculated over the planning horizon or its service life, whichever is smaller. *MigCost(A)* represents the cost to migrate the data from device **A**. *LeftOver(A)* is the leftover value of **A** beyond the planning horizon. Leftover value is a credit for the archival system, since it shows the value the device can give to the system if used for the entirety of its lifetime. *AcqCost(A)* is the acquisition cost of **A**. It includes purchase cost and the cost to move the data onto **A**. If the path of getting a new device has more leftover value, meaning it would give more value to the system, it could be cost-effective to take the path of getting a new device over keeping the old device. A new device that has higher up front cost, but low cost-per-byte in the long-term should replace the old device. If the old device or the new device does not meet the horizon and needs to be repurchased, move out/move in costs need to be paid.

For example, consider **A**, purchased for \$100 two years ago. **A** has an annual operational cost of \$20, migration cost \$10, and a service life of five years. Lets say the planning horizon of the system is five years. The cost of supporting the system for the next five years with **A**, option of keeping the old device *pOld*, would be $20 \times 3 + 110 + 10 + 20 \times 2 - 100 \times \frac{3}{5} = 160$ where the first 20×3 accounts for operational costs paid for **A** for its remaining life (three years), 110 is the acquisition cost of new device **B**, 10 is the cost to move data out of **A**, 20×2 is the operational cost of **B** over its used life (two years), and $100 \times \frac{3}{5}$ is the residual value of **B**. The cost of getting new device **C** to support the next five years, *pNew*, would be $100 + 20 \times 5 = 200$. In this case, it was cost-effective to let **A** run for the rest of its life.horizon functionality.

If the old device is found to be less cost-effective than a new device, the old device gets replaced, and the appropriate migration costs are paid. In addition to replacement decisions, choosing an appropriate planning horizon is important to foreseeing the long-term impact of the media choices made at the time of purchase. The planning horizon is, in essence, the window across which the impact of these choices is evaluated. Each year, when evaluating if old devices justify their costs over the planning horizon *H*, the cost of two paths, keeping the existing device, *pOld*, and getting the new device, *pNew*, is calculated over the duration of the planning horizon or the old device's leftover life, whichever is smaller. The frequency that a system is re-evaluated will depend on pace of significant change in device capabilities as well as the fluidity of the endowment. For our model, we assume constant growth rates and evaluate the cost-effectiveness of the system annually. If the planning horizon exceeds device life, the current most cost-effective device is purchased to provide the coverage until the end of the planning horizon. The new device is considered cost-effective over the old device if the cost-per-byte calculated for *pNew*, accounting for acquisition, migration, operational costs, and leftover value, is lower than the cost per byte for the old device.

D. Prolonged media usage

Drives, usually, are covered by the manufacturer for the period shorter of their warranted service life and the period elapsed in writing the maximum data allowed [15]. We hypothesize that using media beyond its service life may be cost-effective over the long-term in archival systems, without compromising on performance.

We chose to test hard disk, tape, and Blu-ray characteristics against this hypothesis, since they are the primary archival storage media as of today. We also tested SSD characteristics since it had been shown that flash may be suitable for archival storage over time while keeping the costs comparable to traditional disk archives [4]; also there is increasing interest in designing flash to be used for archival purposes [16] and industry is looking at incorporating flash in their archival storage systems [38], [39].

Since the devices in our model are run past their warranted service life, failure rates are expected to increase. However, our goal was to explore the trade-off between survivability and reliability. The reliability of HDD, tape, and Blu-ray decrease steeply over time, and we expected that extending the service life for these media would not be cost-effective [11], [40]–[42]. SSD have been shown to have high reliability [17]–[19], and so we expected to see significant benefits by extending their life in archival systems.

IV. EXPERIMENTAL SETUP

Table I: Archival system settings

Duration	100 years	Storage demand	1 PB
Tapes per reader	10	Blu-ray per reader	100
K_r	5% (unless specified otherwise)	Replication level	2

Table II: Initial media settings

Media	Disk	SSD	Tape	Blu-ray
Purchase Cost	\$100	\$619	\$42	\$10
Capacity per device (TB)	4	1	6.25	0.300
Idle power (W)	7	.31	0	0
Power required when active (W)	9.45	3.5	0	0
Read bandwidth (MB/s)	186	540	0	0
Data center cost (\$/month)	1.40	0.70	0.40	0.05
Cooling power (W)	10	0.5	0	0
Service life (years)	5	5	30	15

We are simulating the cost of keeping 1 PB of data in the archival system for 100 years. The period of 100 years is chosen to simulate long-term preservation. 1 PB is used to ensure that the archive is big enough that, at the end, it is not on a single drive. All the future costs are discounted by a 3% annual discount rate to consider the inflation in the future. Tables I and II show the settings used for the test archival system and media. At least fifty simulations were run per configuration, at which point we found that further experiments did not significantly alter our results. We tried up to 150 runs. The error bars in graphs in Section V represent one standard deviation.

A. Media Characteristics

Though the model framework we propose in Section III can be applied to arbitrary device characteristics, and devices will of course change over time, we seed our simulation with current parameters for a variety of device types. There are two benefits to this. First, it provides a framework for understanding the current storage landscape with regards to archives. Second, and more importantly, it allows us to suggest a path for future hardware development to better design for efficient long-term data archiving.

Default values for hard disk have been taken from the `Seagate Constellation ES.3` and SSD defaults are from the `Samsung 840 Evo 1TB`. Blu-ray defaults are for the `Memorex BD-R` with the capacities of archival disks [43] and read with the `LG UH12NS30 Super Multi Blue` drive. Our example tape media is based on the `IBM LTO6 Ultrium` read with the `IBM LTO Ultrium 6 Full High Tape Drive`. All values are current as of early 2016.

Hard disk failure rates, in the model, follow a Weibull distribution, or "bathtub curve" [11]. We assume a constant annual media failure rate of 0.1% for SSDs [18], [19]. Due to limited data on media failure rates for tape and Blu-ray, we assumed a linearly increasing media fail rate [41], [42]. Annual media fail-rates are shown in Figure 3.

For hard disks and flash, we experimented with a five year service life. The Blu-ray service life is set at fifteen years. The Blu-ray reader and the tape reader service life is set at five years. Tape service life is set at thirty years. Hard disk and SSD's power consumption have been taken from benchmarks [44], [45].

B. Data Center Costs

Data center costs, in the model, represent the costs of computer room floor space. We are not considering networking and labor costs separately in this work, but we account for them in the cost of computer room floor space. Power, space, and cooling cost are considered to be constant for the entire duration of the simulation for all media. There will be variations in the power and space costs over 100 years, in reality, and that would affect the output of the model. As we exhaust non-renewable power sources, energy will be more expensive over time. This means that our model is a worst case: in reality, SSDs and other lower power devices will be even more critical.

Each tape library takes approximately ten square feet of computer room space [46]. Each disk rack takes approximately seven square feet. if computer room space is billed at a modest

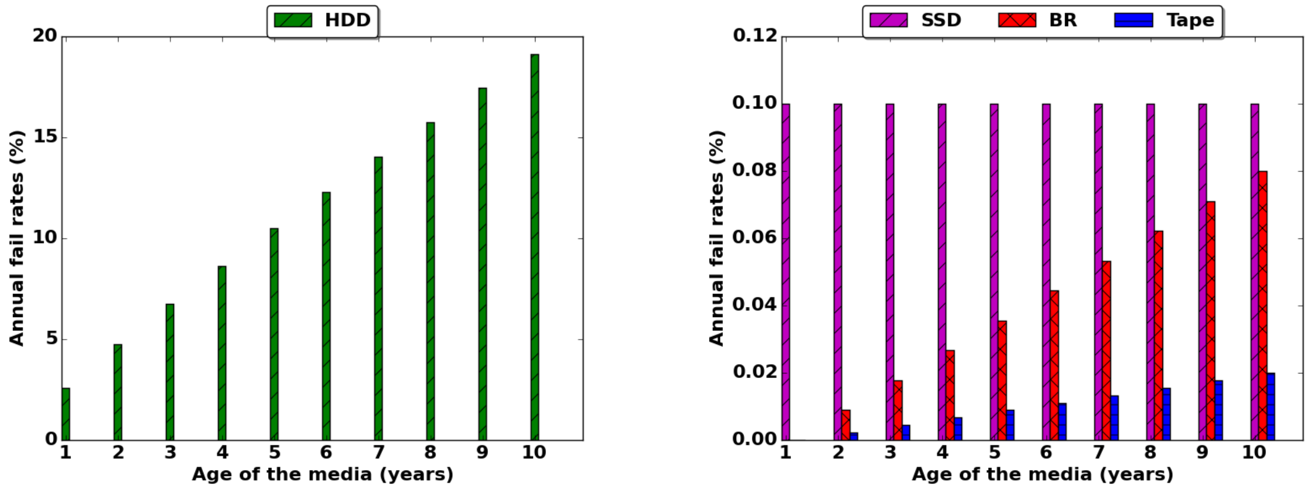


Fig. 3: Annual failure rates for hard disks (Weibull distribution) increase up to 20% by the end of their 10^{th} year. For tape and Blu-ray, failure rates increase linearly and for SSD, we assumed constant failure rates following the usage patterns in long-term preservation systems. Hard disk failure rates are much higher than those for SSD, tape, and Blu-ray.

rate of $\$20$ per square foot per month, and there are 500 tapes per library, one tape thus takes approximately $\$0.40$ per month worth of computer room space. Assuming each disk rack hosts 100 disk drives, each disk drive costs approximately $\$1.40$ per month [46].

There is limited data on the data center costs of Blu-ray and SSDs. Blu-ray is a low density medium, and requires library and external readers in data center environments [47]. Keeping in mind the demands of long-term preservation, we assumed the existence of one Blu-ray reader per 100 Blu-ray disks, and that one library can host up to 40 Blu-ray drives.

Assuming that a Blu-ray library takes ten square feet of space, the same as a tape library, each Blu-ray disk takes approximately $\$0.05$ per month worth of computer room space. Due to the small size of SSDs compared to hard disks, we assumed that one SSD takes half of a disk's data center cost worth of computer room space, which is $\$0.70$ per month. We purchase one tape drive per 10 tapes. The model does not account for tape and Blu-ray libraries.

V. Results

We ran a series of experiments to test the hypotheses that extending media service life or adjusting the planning horizon could reduce the cost of long-term archival storage. *BR* represents Blu-ray in the graphs.

A. Service Life

Prolonging media service lives leads to a system having relatively more low capacity devices. At the same time, K_r is increasing, so late replacement brings higher capacity devices than they would have had in case of early replacement and hence the cost goes down. In case of slow K_r growth, such as 5%, late replacement is cost-effective, but if the annual media density growth is high, such as 20%, early replacement would be cost-effective.

1) Cost: Extending media service life, as shown in Figure 4, helps keep the cost down for both SSDs and hard disks. This change is more prominent for SSDs because they are expensive, have low operational cost, high bandwidth, and low failure rates. We see approximately a 10% cost reduction, close to $\$500,000$ for a system having $1\,\text{PB}$ data, for SSDs, in the event of 5% density growth rates. For systems, storing a larger amount of data, the savings would increase. For hard disks, the failure rate increases up to 20% by the tenth year of their life leading to a high number of failures, and eventually more expenditure due to a higher number of purchases. Also, their operational costs are significant. Hence, we noticed a 7% improvement in total costs for hard disks. For tape and Blu-ray, media failure rates are high and media service life is high, and the dominating costs are due to drives and libraries' capital and operational costs. We do not see any benefit in using them beyond their suggested service life.

2) Reliability: We did not see any measurable drop in reliability due to prolonged usage of media. We maintain (8, 2) parity in the system which sufficiently counters our extrapolated annual media failure rates with a capacity overhead of 20%. We check for failures on a monthly basis in our system, which might be appropriate for an archival system given the large amount of data they hold, and the financial crunch they are in. Checking failures more often would increase costs though it might catch media failures early.

3) Energy Consumption and Bandwidth: Extending device service life leads to both an increased proportion of older, less energy efficient devices in the system, as well as a constant stream of new devices as these lower reliability devices fail. For SSDs, using media beyond its service life, for both high K_r, such as 20% and a low K_r, such as 5%, increase the system's energy consumption (as shown in Figure 5a), because keeping a high number of old low capacity devices increases the energy consumption. For hard disks, due to failure, old media gets replaced earlier than the period of extension and hence does not affect the energy consumption. For tape and Blu-ray, since all the energy consumption comes from drives and libraries, we do not see a change in the energy consumption as media is used past their suggested service life.

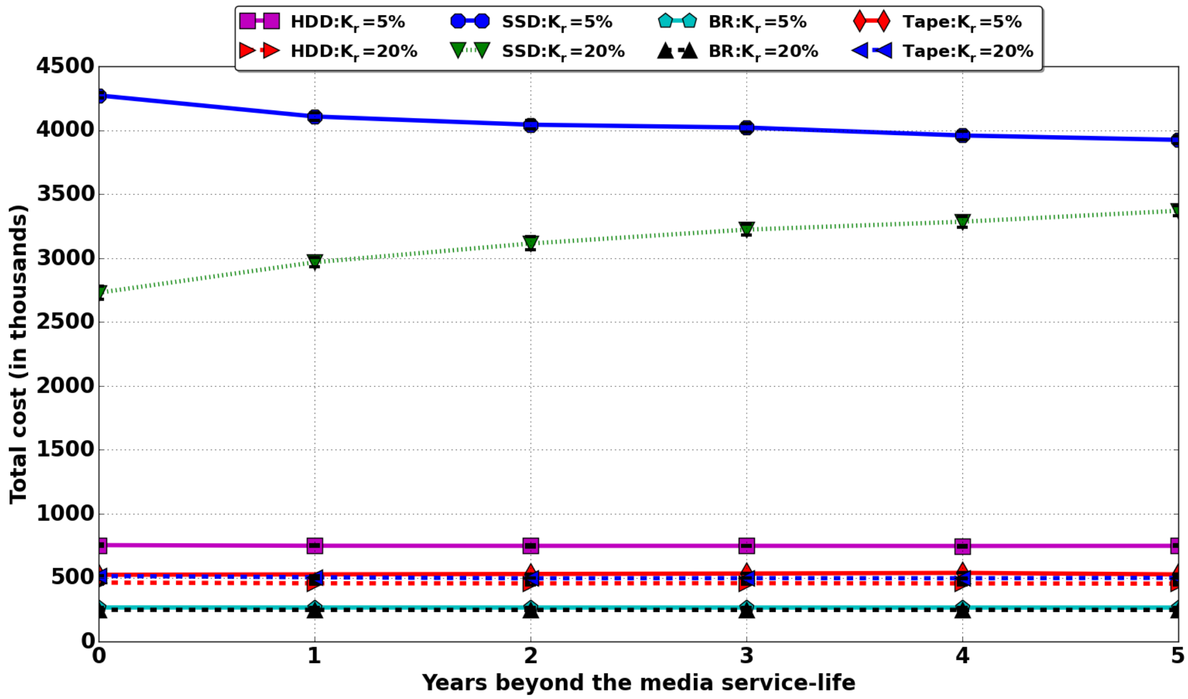
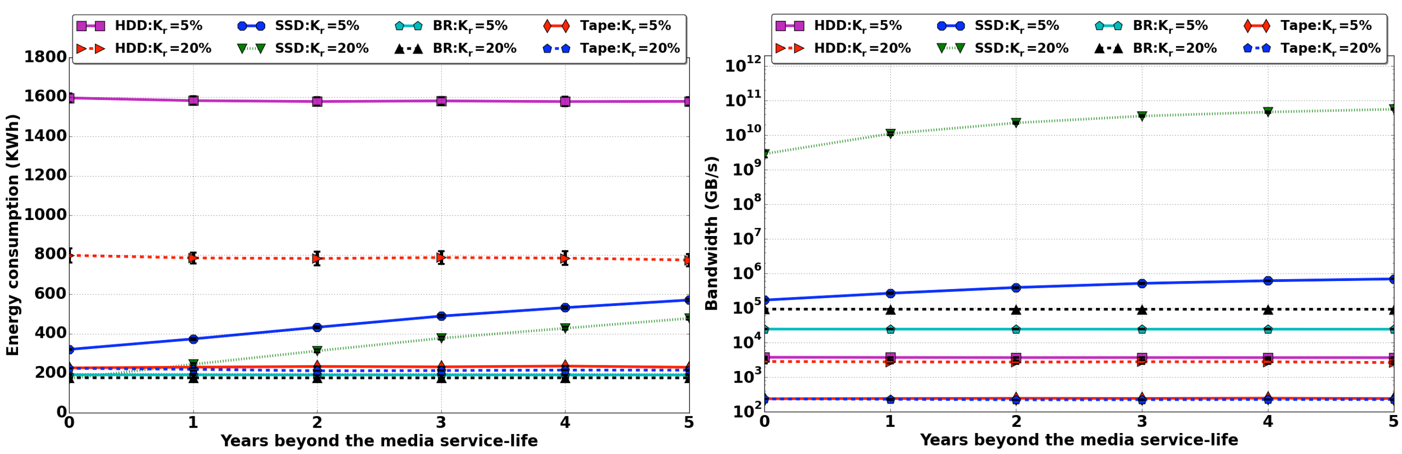


Fig. 4: Prolonged media usage helps keep costs low for both SSDs and hard disks, however, for SSDs, the impact is more (up to 10%) for low Kryder rates, such as 5%. Notice that, extending media usage, in the event of high Kryder rates, such as 20%, would have resulted in a cost increase, up to 35%, due to the system having a high number of low capacity drives rather than a low number of high capacity drives. Running hard disks beyond their service life resulted in a 7% cost savings over the duration of the simulation due to their high failure rates. For tape and Blu-ray, we did not notice any cost benefits by using them beyond their suggested service lives likely because most of their expense come from readers and libraries.



(a) (b)

Fig. 5: Due to prolonged media usage, a high number of older devices stay longer in the system, increasing power consumption and system bandwidth. For hard drives, high media failure rates cause devices to be replaced earlier, but for SSDs, due to their low failure rates, a high number of low capacity devices stay in the system for long, increasing the power consumption and supported bandwidth. Also, as period of extension increases, old media is replaced by higher density and high bandwidth devices than they would have if replaced early, hence even though there will be fewer devices in the system. We see almost a $10\times$ growth in the system bandwidth for SSDs for both 5% and 20% annual density growth rates.

As shown in Figure 5b, keeping a high number of old devices in the system would affect the system's bandwidth positively, since the old devices would get replaced by higher density devices than they would have if they were replaced early. For SSDs, bandwidth improves markedly in periods of rapid improvement in device manufacturing, pegged here to Kryder's rate. Notice that for SSDs, bandwidth grows linearly with density growth, while for hard disks, it is the square root of the density growth [38]. Also, if disks are denser, we need fewer of them, so overall bandwidth is lower.

B. Lengthening Planning Horizon

We calculate the cost of supporting the archival system, comparing old devices with their newer equivalents over the planning horizon. If the old device is found to be less cost-effective than a new device over the planning horizon, the old device gets replaced. A shorter planning horizon represents a low risk situation, since only short-term costs are considered. We found that the planning horizon has a strong effect on the cost of archival systems, especially in the time of low media density growth. This effect is significant for systems using SSDs, due to their high CAPEX, high reliability, and low OPEX, improving costs by as much as 6%, close to $\$300,000$ for 1 PB, but not so much for hard disks, tapes, and Blu-ray. Long planning horizons add uncertainty to the cost-calculations, but also lead to better global outcomes.

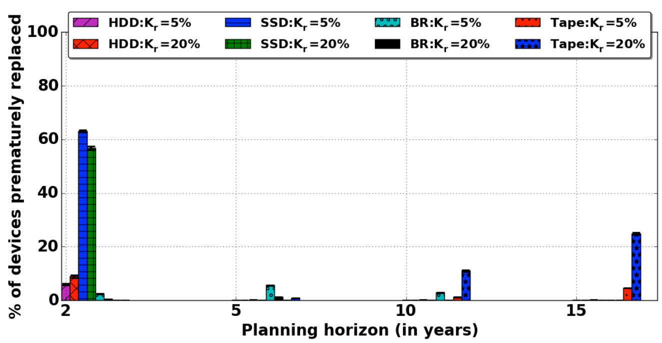


Fig. 7: Cumulative percentage of total devices replaced early in the system for the entire duration of 100 years. Number of devices replaced early decrease as the planning horizon is increased forcing devices to be used for a long time. Devices are retired, if not prematurely replaced, at the end of their service life. In the times of low media density growth rates, early replacement results in increased costs without getting much capacity benefit.

1) Cost: Historically, when devices improved quickly, new devices were so much cheaper to run that it was worth buying new devices to save on operating costs. Figure 6 shows, however, that over time this decision will not be cost-effective as the marginal improvement decreases for both hard disks and SSDs, planning for any shorter than the media service life results in higher total costs. In particular, as shown in Figure 7, archival systems designed with short planning horizons had a high number of premature replacements. For tapes, the service life is 30 years, and the purchase costs are low. Extending planning horizons increased consideration of new tapes due to lower costs over a long period of time, and hence the replacements.

In the archival system with short planning horizons, only short-term costs are considered, and the long-term benefits of keeping the device are ignored. However, when considering long-term costs, we see that using the current device for its service life and buying another device later might be more cost-effective than retiring the current device and getting the new device right away because we would get a longer archival life covered in the first case at a low cost. For SSDs, we see an almost 6% improvement in cost, for $K_r = 5\%$, as shown in Figure 6, though in the event of high K_r extending planning horizons would have resulted in 9% extra costs.

For the removable device types we modeled, media is cheap, and the major contribution to the cost came from external factors such as the readers required to maintain the data. We model one Blu-ray drive per 100 Blu-ray disks in the system, and the Blu-ray drive annual fail rate is kept at 5%. As media gets denser over time, new readers would be required for better density even if the old ones do not fail. Tape readers can typically only support three generations of media before format obsolescence causes them to need replacing. Although the device replacement is minimal for a long planning horizon for both tape and Blu-ray, as shown in Figure 7, we do not see much change in the costs, as shown in Figure 6.

2) Reliability: For longer planning horizons, many devices remain in the system for the entirety of their expected service lives, increasing the number of failures. However, as devices fail, they are replaced with higher capacity devices, reducing the total number of devices, and so offsetting much of the decrease in reliability. We also studied the effect of an extended planning horizon on the data loss in the system. Extending the planning horizon does not result in compromised reliability and increase data loss for all media.

3) Power Consumption and Bandwidth: Similar to extending service life, extending planning horizon leads to the system having a higher proportion of low capacity devices. However, over the long-term, extending the planning horizon leads to larger devices in a more stable equilibrium. Short planning horizons are better to keep the power consumption low. Power consumption is higher in archival systems where planning horizon is as long as the media service life, because the system used old devices for a long time and to meet the storage demand, the system had to buy more devices, as shown in Figure 8a, for hard disks and SSDs. For tape and Blu-ray, most power is consumed by tape libraries and tape drives. Hence, we do not see much change in the power consumption for all density growth rates and different planning horizons.

As the planning horizon increases, the system's bandwidth would increase, as shown in Figure 8b. It also depends on the time when the devices got replaced, because of increasing media densities, and also on K_r because bandwidth of the media growth is dependent on the density growth. If media gets replaced early during the preservation, it is replaced by lower density media than it would be if it was replaced late. For hard disks and Blu-ray, the bandwidth does not grow linearly with the density growth, and having a high number of them yields high bandwidth. For SSDs, we saw an approximately $5\times$ growth in bandwidth for K_r=5%.

C. Combining Planning Horizon with Service Life Extension

As Figure 9 shows, combining a long planning horizon with extended service life helps in minimizing the costs further for SSDs. It results in almost 12% fewer costs with a long planning horizon for a three year extension and almost 15% cost reduction with a five year extension. Since the costs are amortized over longer periods, a short planning horizon for media with a long life and corresponding increased residual value lowers the total cost, because there are few purchase events. As service life increases, the optimal planning horizon for the archival system rise correspondingly. For hard disks,

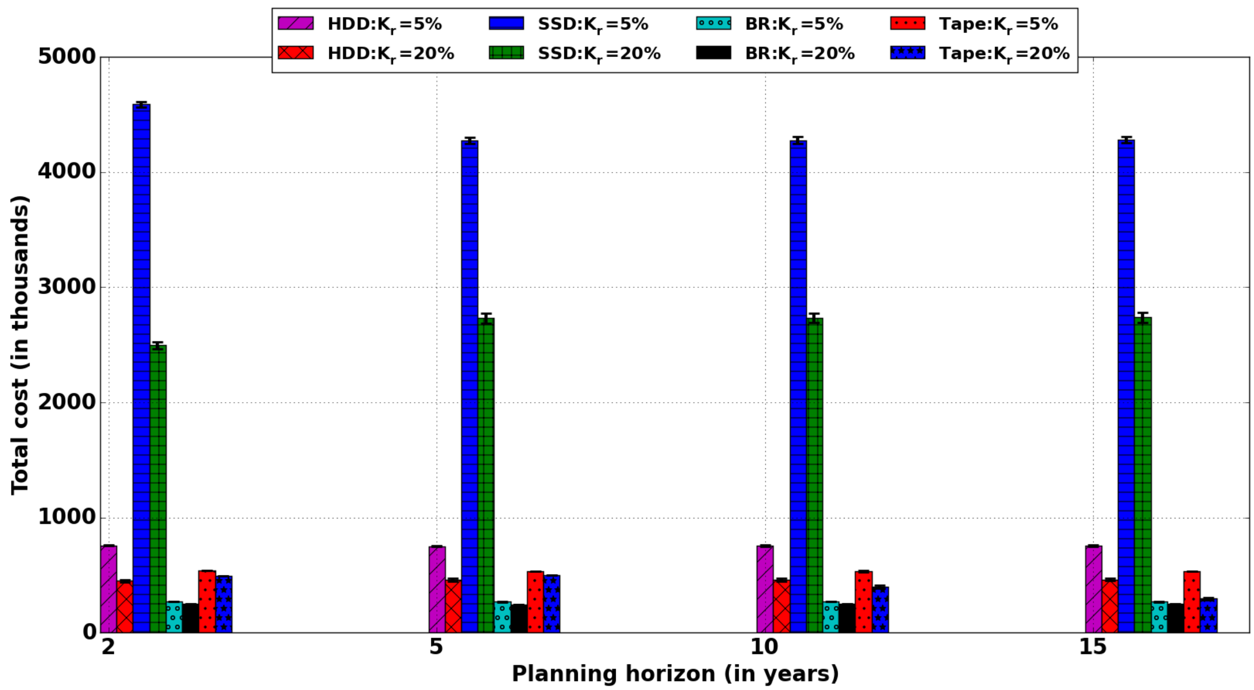
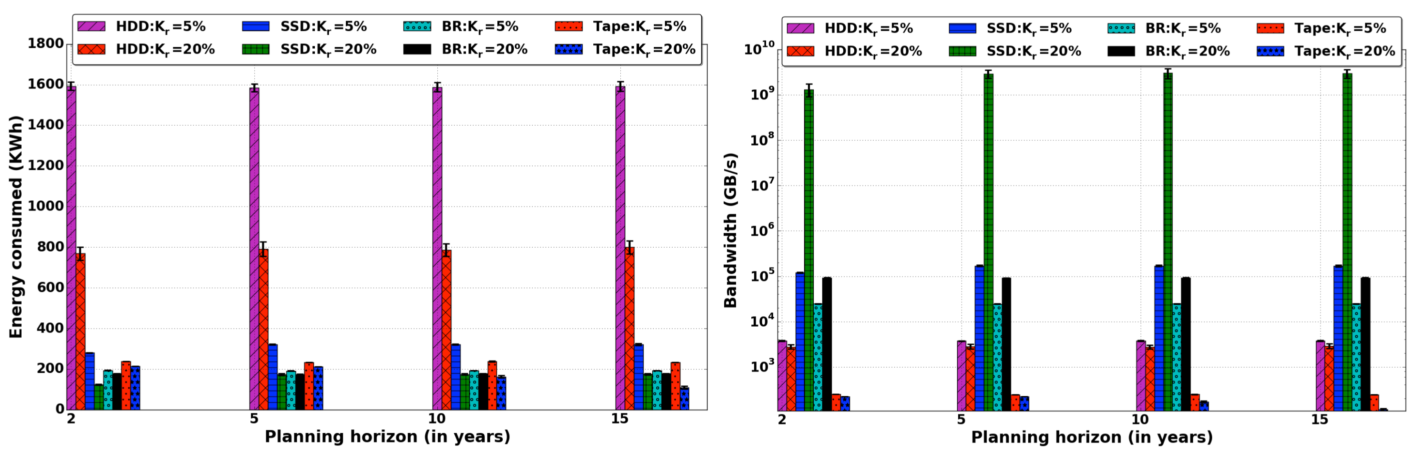


Fig. 6: For SSDs, choosing the optimal planning horizon results in cost as much as 6% lower for K_r=5%, which would have resulted in 9% higher costs for K_r=20%. For hard disks, the savings are not as significant as of SSDs, but it helps save almost 4% on the total costs. For tape and Blu-ray, the majority of the costs come from media readers, hence, we do not see the planning horizon affecting the total cost much.



(a) (b)

Fig. 8: As devices are used for a longer period of time, the power consumption of the archival system increases, since the system will have more devices of low capacity. This effect is greater for hard disks, since they consume three times more power than SSDs (Devices are retired at the end of their manufacture-recommended service life, hence the increase is most from a two to five year horizon). For tape and Blu-ray, all power is consumed by the media library and drives. Changing the planning horizon does not affect power consumption for them, since the drives and libraries are not replaced early due to a short planning horizon.

prolonged media usage does not make sense, due to their high operational costs, and high failure rates as they age. Extending the planning horizon to consider prolonged usage of hard disks results in approximately a 2% cost savings.

VI. Discussion

Using media beyond their suggested life-time was effective for SSDs without causing any data loss, in the times of low media density growth rate, but surprisingly made little difference for hard disks, tapes, and optical discs. For hard disks, the cost benefit of extending the service life was countered by the rapidly increasing failure rates compared to other media types.

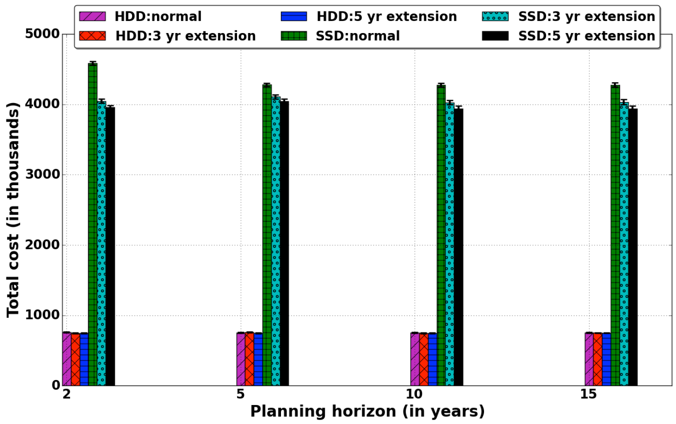


Fig. 9: Extending the planning horizon for old media, in the system, is economically effective for SSDs. It results in almost 12% fewer costs with a long planning horizon for a three year extension and almost 15% cost reduction with a five year extension. A short planning horizon, such as two years, and a long planning horizon, such as ten years, both are cost-effective for media running for a long time. Since residual value is calculated over a long-period (service life + period of extension), the short-term media has higher residual value, and does not get replaced early. This figure was plotted for $K_r = 5\%$.

For SSDs, however, we observed cost-benefit, up to 10%, if we used them for an extended period of time in the times of low K_r, such as 5%. Although, we save by using old SSDs in the times of low K_r, it would have resulted in a loss of 35% in the times of high K_r. Based on limits in hardware manufacturing, it is reasonable to expect that K_r will decrease over the system lifetime.

Extending media service life by 10 and 15 years, for SSDs, resulted in another 4% and 6% of cost-savings (beyond savings with a five year extension), respectively. Drives built to these specifications are good candidates for cost-efficient long term archiving. According to our results, running media beyond their manufacturer suggested service life yields more savings for SSDs than with an optimal planning horizon. For SSDs, we observed cost-benefits of up to 10%, and an 7% reduction in hard drive cost, by asserting that the planning horizon may be at least as long as the device service life. The savings are due to low device replacements, which is cost-effective in the times of low annual media density growth rates, but would have caused high costs, up to 9%, shown in Figure 6, in the time of high annual media density growth rates. Also, by extending media service life in the system, we see a significant increase in the system bandwidth, which would be highly appreciated in active archives.

The sharp increase in HDD failure rates over time makes running these drives beyond their service life less cost efficient than other devices. So, for systems using hard disks, where using old devices is a poor option, choosing an optimal planning horizon is important. For Blu-ray and tapes, the majority of the total costs come from expenditures made on readers and libraries, including purchases, computer room space, and power, so extending the planning horizon and the service life does not reduce the archival system's cost. We do not see a major drop in the system bandwidth or increase in power consumption due to this extension, because using a high number of low density media would provide more bandwidth than using a low number of high density media (we noticed an almost $10\times$ growth in the system bandwidth for SSDs), and that is a plus point for even enterprise storage.

Prior to 2011, due to high annual bit density growth rates, buying a high capacity new disk at a low cost was preferred over keeping a high cost, and low capacity hard disk for a long time, hence manufacturers had little or no interest in making such devices. Today, when hard disk density growth rates are growing at much slower rates than in the past, it makes sense to invest in long-lived hard disks, and to use such hard disks. Archival system designers need to update their planning horizons, and be ready to use the media for periods longer than the usual five year life-span. SSDs, media whose wear mostly is dependent upon the data read/write cycles, may perform better in archival systems than in enterprise systems.

Recent announcements of huge improvements in SSD capacities make it seem likely that we are moving towards high capacity with low operational cost long-lived devices with no external dependencies [48], [49]. On the other hand, hard disks with Heat-assisted magnetic recording (HAMR) are expected to arrive by 2016, but in 4TB capacity [50]. SSDs are advancing much faster than hard disks, and the archiving community needs to be ready for their use in data-centers to keep the data foot-print small with less costs.

In light of our results, we can say that archival systems could benefit from archival media having a high CAPEX/OPEX ratio, high reliability, and a long service life. Though these devices do not yet exist, there are multiple efforts in progress to bring a specialized archival device to the market [51]–[55]. For all these media, initial investment is high (expensive media and readers/writers), but they have high reliability, an extremely long life, and low operational costs in comparison with traditional archiving media (hard disk, tape, and Blu-ray). To make the best use of such media and keep the archiving costs low, the system designers would have to broaden their horizons, and change their policies to account for economic realities, and limited planning capabilities. As service life increases in specialized media, it will be critical to balance the certainty of the short-term system design with the benefit of a longer planning horizon.

VII. Future work

In this work, the choice of media is based only on the costs, and we noted significant bandwidth improvements without incurring heavy costs, but organizations may have other requirements, such as high capacity, bandwidth, input/output operations per second, or low probability of failure and power. It is important to explore the effect of other selection criteria on the total costs of preservation. Also, the model presented in this paper assumes a single device type over the life of the archive. We are developing a heterogeneous device model with data modeling and finer granularity to better inform large-scale archival system design.

Recently, industry and research labs are realizing the importance of reliable and long-lived media in long term preservation. Hitachi's Glass disks [51], DNA [52], [53], Rosetta disks [54], and Sapphire disks [55], are some of the efforts in this direction. As old technologies mature and we see new archival devices on the horizon, archival systems are bound to change as well. In the future, we plan to extend our work to consider using upcoming media into an archival system's cost calculations.

Changing archival access patterns will also affect the choice of the media, and eventually the costs of the archival system. Historically, the archives were considered to be cold. However, the demand for active archives is continuously increasing [56]–[58]. Researchers in many fields, such as medical science, computer science, weather science, animation, and archaeology, are realizing the importance of data retention. As this demand increases, archival systems need to be designed differently, and that will affect the costs.

VIII. Conclusion

It is important to note the change slow growing media density growth rates impose on long-term storage. In times of high media density growth rates, extending media usage would have resulted in increased costs by as much as 35%, but today to make better use of archival media while keeping the costs of archiving low, system designers need to update their media selection policies. We suggest prolonged usage of media and planning horizon extension to minimize the archival system's cost. Using SSDs beyond their service life was cost-effective, saving almost 15% on total costs.

The relationship between the system's planning horizon, media service life, and cost will be important for future archival devices, which will likely have a long service life, and high cost. The demand for archival media is increasing, and enterprises are increasing their effort towards designing archival media [16], [59], [60]. As archival media hits the market, we recommend that corporations adjust their planning horizon to match the extended service life of this more robust media. This also suggests device manufacturers should design hardware that is long lived and reliable to take advantage of archival trade-offs.

Acknowledgments

This research was based on work supported in part by the DOE under Award Number DE-FC02-10ER26017/DE-SC0005417 and NSF under award IIP-1266400 and industrial members of the Center for Research in Storage Systems. The authors would like to thank our colleagues in the Storage Systems Research Center at University of California, Santa Cruz and Emory University for their valuable feedback on this paper.

References

[1] EMC, "New Digital Universe Study Reveals Big Data Gap: Less Than 1of Worldś Data is Analyzed; Less Than 20% is Protected," 2012, http://www.emc.com/about/news/press/2012/20121211-01.htm.

[2] R. L. Moore, J. D'Aoust, R. H. McDonald, and D. Minor, "Disk and tape storage cost models," *Archiving 2007*, 2007.

[3] D. S. Rosenthal, D. C. Rosenthal, E. L. Miller, I. F. Adams, M. W. Storer, and E. Zadok, "The economics of long-term digital storage," *Memory of the World in the Digital Age, Vancouver, BC*, 2012.

[4] P. Gupta, A. Wildani, E. L. Miller, D. Rosenthal, I. F. Adams, C. Strong, and A. Hospodor, "An economic perspective of disk vs. flash media in archival storage," in *Modelling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS), 2014 IEEE 22nd International Symposium on*. IEEE, 2014, pp. 249–254.

[5] K. Finley, "IBM cloud storage partner nirvanix files for bankruptcy," 2013, http://www.wired.com/2013/10/nirvanix-bankrupt/.

[6] CBC News, "Federal libraries, archives shutting down," May 2012, http://www.cbc.ca/news/canada/ottawa/federal-libraries-archives-shutting-down-1.1139085.

[7] C. Allen, "WHY ARE SOME DOCTOR WHO EPISODES 'MISSING'?" 2013, http://www.doctorwho.tv/whats-new/article/why-are-some-doctor-who-episodes-missing/.

[8] Wuala, "Wuala: Termination of free storage and storage promotion," October 2014, http://cdn.wuala.com/files/termination_free_storage.pdf?winst=1414602574471&of=0.

[9] O. W. Eshbach, *Eshbach's handbook of engineering fundamentals*. John Wiley & Sons, 1990.

[10] D. Anderson, "Archive Drive Study," 2009, http://www.digitalpreservation.gov/meetings/documents/othermeetings/5-4_Anderson-seagate-v3_archive_study.pdf.

[11] B. Schroeder and G. A. Gibson, "Disk failures in the real world: What does an MTTF of 1, 000, 000 hours mean to you?" in *FAST*, vol. 7, 2007, pp. 1–16.

[12] Backblaze, "Hard Drive Reliability Update," 2014, https://www.backblaze.com/blog/hard-drive-reliability-update-september-2014/.

[13] E. Pinheiro, W.-D. Weber, and L. A. Barroso, "Failure trends in a large disk drive population." in *FAST*, vol. 7, 2007, pp. 17–23.

[14] D. Anderson, "Hard drive directions," *Designing Storage Architectures for Preservation Collections*, pp. 22–23, 2009.

[15] Samsung, "SAMSUNG SSD Limited Warranty For All Samsung SSD," 2014, http://www.samsung.com/global/business/semiconductor/minisite/SSD/global/html/support/warranty.html.

[16] T. Coughlin, "Keeping Data For A Long Time," 2014, http://www.forbes.com/sites/tomcoughlin/2014/06/29/keeping-data-for-a-long-time/.

[17] G. Gasior, "Behind the scenes with Intel's SSD division," 2014, http://techreport.com/review/26269/behind-the-scenes-with-intel-ssd-division.

[18] A. Ku, "Investigation: Is Your SSD More Reliable Than A Hard Drive?" 2011, http://www.tomshardware.com/reviews/ssd-reliability-failure-rate,2923-5.html.

[19] L. Mearian, "SSDs do die, as Linus Torvalds just discovered," 2013, http://www.computerworld.com/article/2484998/solid-state-drives/ssds-do-die--as-linus-torvalds-just-discovered.html?page=2.

[20] M. Bamburic, "Modern SSDs can last a lifetime," 2014, http://betanews.com/2014/12/05/modern-ssds-can-last-a-lifetime/.

[21] G. Gasior, "The SSD Endurance Experiment: Two freaking petabytes," 2014, http://techreport.com/review/27436/the-ssd-endurance-experiment-two-freaking-petabytes.

[22] J. Meza, Q. Wu, S. Kumar, and O. Mutlu, "A large-scale study of flash memory failures in the field," 2015.

[23] B. Schroeder, R. Lagisetty, and A. Merchant, "Flash reliability in production: The expected and the unexpected," in *14th USENIX Conference on File and Storage Technologies (FAST 16)*. Santa Clara, CA: USENIX Association, Feb. 2016, pp. 67–80. [Online]. Available: http://usenix.org/conference/fast16/technical-sessions/presentation/schroeder

[24] Seagate, "Seagate Archive HDD," 2014, http://www.seagate.com/www-content/product-content/hdd-fam/seagate-archive-hdd/en-us/docs/archive-hdd-dS1834-3-1411us.pdf.

[25] Seagate, "Kinetic HDD," 2015, http://www.seagate.com/www-content/product-content/hdd-fam/kinetic-hdd/en-us/docs/kinetic-ds1835-1-1110us.pdf.

[26] Seagate, "OEM Warranty Statements," 2015, http://www.seagate.com/in/en/support/warranty-and-replacements/oem-warranty/.

[27] WD, "Datacenter Hard Drives," 2015, http://www.wdc.com/en/products/internal/enterprise/.

[28] P. Williams, D. S. Rosenthal, M. Roussopoulos, and S. Georgis, "Predicting archival life of removable hard disk drives," in *Archiving Conference*, vol. 2008, no. 1. Society for Imaging Science and Technology, 2008, pp. 188–192.

[29] Z. Li, A. Mukker, and E. Zadok, "On the Importance of Evaluating Storage Systems' $Costs," in *6th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 14)*. Philadelphia, PA: USENIX Association, Jun. 2014.

[30] U. B. Kejser, A. B. Nielsen, and A. Thirifays, "Cost model for digital preservation: Cost of digital migration," *International Journal of Digital Curation*, vol. 6, no. 1, pp. 255–267, 2011.

[31] S. Chapman, "Counting the costs of digital preservation: is repository storage affordable?" *Journal of digital information*, vol. 4, no. 2, 2006.

[32] P. Wheatley, P. Ayris, R. Davies, R. McLeod, and H. Shenton, "LIFE: costing the digital preservation lifecycle," 2007.

[33] P. Wheatley and B. Hole, "LIFE3: Predicting Long Term Digital Preservation Costs," 2009.

[34] M. Addis and M. Jacyno, "Tools for modelling and simulating migration based preservation," Project Report, December 2010. [Online]. Available: http://eprints.soton.ac.uk/339147/

[35] N. Beagrie, M. Duke, M. Patel, L. Lyon, C. Hardman, D. Kalra, B. Lavoie, and M. Woollard, "The krds benefit analysis toolkit: Development and application," *International Journal of Digital Curation*, vol. 7, no. 2, pp. 64–67, 2012.

[36] S. Abrams, P. Cruse, J. Kunze, and M. Mundrane, "Total Cost of Preservation (TCP): Cost Modeling for Sustainable Services," 2012.

[37] P. Xue, E. Shehab, P. Baguley, and M. Badawy, "Cost modelling for long-term digital preservation: Challenges and issues," in *Proc. 9th Intl. Conf. on Manufacturing Research ICMR*, 2011, pp. 187–192.

[38] D. Floyer, "The Emergence of a New Architecture for Long-term Data Retention," 2014, http://wikibon.org/wiki/v/The_Emergence_of_a_New_Architecture_for_Long-term_Data_Retention.

[39] Y. Sverdlik, "OPEN COMPUTE: FACEBOOK TO MOVE ALL ARCHIVAL DATA ONTO ALL-FLASH STORAGE," 2013, http://www.datacenterdynamics.com/focus/archive/2013/01/open-compute-facebook-move-all-archival-data-all-flash-storage.

[40] D. Olds, "How not to evaluate hard disk reliability: Backblaze vs world+dog," 2014, http://www.theregister.co.uk/2014/02/17/backblaze_how_not_to_evaluate_disk_reliability/.

[41] R. Fraimow, "American (Archives) Horror Story: LTO Failure and Data Loss," 2015, http://code4lib.org/conference/2015/fraimow.

[42] A. Lafrance, "The Library of Congress Wants to Destroy Your Old CDs (For Science)," 2014, http://www.theatlantic.com/technology/archive/2014/05/the-library-of-congress-wants-to-destroy-your-old-cds-for-science/370804/.

[43] A. Cunningham, "Next-gen "Archival Disc" will squeeze 1TB of data onto optical discs," 2014, http://arstechnica.com/gadgets/2014/03/next-gen-archival-disc-will-squeeze-1tb-of-data-onto-optical-discs/.

[44] tom's HARDWARE, "All SSD Charts 2014," 2014, http://www.tomshardware.com/charts/ssd-charts-2014/benchmarks,129.html.

[45] tom's HARDWARE, "All HDD Charts 2013," 2014, http://www.tomshardware.com/charts/hdd-charts-2013/benchmarks,134.html.

[46] McAdam, Dianne, "Tape and Disk Costs–What It Really Costs to Power the Devices," 2006, https://www.spectralogic.com/index.cfm?fuseaction=home.displayFile\&DocID=1944.

[47] J. Brodkin, "Why Facebook thinks Blu-ray discs are perfect for the data center," Jan 2014, http://arstechnica.com/information-technology/2014/01/why-facebook-thinks-blu-ray-discs-are-perfect-for-the-data-center/.

[48] C. Mellor, "Watch out, Tokyo! Samsung readies a 15 TERABYTE SSD," September 2015, http://www.theregister.co.uk/2015/08/13/godzilla_roars_15tb_ssd_from_samsung/.

[49] M. Beasley, "Intel and Micron announce new '3D XPoint' memory solution 1,000 times faster than current NAND flash storage," July 2015, http://9to5mac.com/2015/07/28/intel-micron-3d-xpoint/.

[50] C. Mellor, "Muted HAMR blow from Seagate: damp squib drive coming in 2016," August 2015, http://www.theregister.co.uk/2015/08/31/muted_hamr_blows_coming/.

[51] Phys.org, "Data that lives forever is possible: Japan's Hitachi," 2012, http://phys.org/news/2012-09-japan-hitachi.html.

[52] N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E. M. LeProust, B. Sipos, and E. Birney, "Towards practical, high-capacity, low-maintenance information storage in synthesized dna," *Nature*, 2013.

[53] J. Bohannon, "DNA: The Ultimate Hard Drive," 2012, http://news.sciencemag.org/math/2012/08/dna-ultimate-hard-drive?ref=hp.

[54] Norsam, "HD-Rosetta^{TM} and Related Processes," 2008, http://www.norsam.com/rosetta.html.

[55] S. Anthony, "The 10-million-year sapphire hard disk," 2013, http://www.extremetech.com/extreme/132793-the-10-million-year-sapphire-hard-disk.

[56] I. F. Adams, M. W. Storer, and E. L. Miller, "Analysis of workload behavior in scientific and historical long-term data repositories," *ACM Transactions on Storage (TOS)*, vol. 8, no. 2, p. 6, 2012.

[57] Allen, Erin, "Update on the Twitter Archive At the Library of Congress," 2013, http://www.loc.gov/today/pr/2013/files/twitter_report_2013jan.pdf.

[58] T. Coughlin, "Adventures in Archiving," 2013, http://www.forbes.com/sites/tomcoughlin/2013/02/27/adventures-in-archiving.

[59] Seagate, "Affordable Active Archive Hard Drives for Cloud Storage," 2015, http://www.seagate.com/products/enterprise-servers-storage/nearline-storage/archive-hdd/.

[60] Western Digital, "Datacenter Archive HDD," 2014, http://www.wdc.com/en/products/products.aspx?id=1340.