Flash Reliability in Production:
The Expected and the Unexpected

Bianca Schroeder, *University of Toronto;* **Raghav Lagisetty and Arif Merchant,** *Google Inc.*

**This paper is included in the Proceedings of the
14th USENIX Conference on
File and Storage Technologies (FAST '16).**

February 22–25, 2016 • Santa Clara, CA, USA

**Open access to the Proceedings of the
14th USENIX Conference on
File and Storage Technologies
is sponsored by USENIX**

Flash Reliability in Production: The Expected and the Unexpected

Bianca Schroeder
University of Toronto
Toronto, Canada

Raghav Lagisetty
Google Inc.
Mountain View, CA

Arif Merchant
Google Inc.
Mountain View, CA

Abstract

As solid state drives based on flash technology are becoming a staple for persistent data storage in data centers, it is important to understand their reliability characteristics. While there is a large body of work based on experiments with individual flash chips in a controlled lab environment under synthetic workloads, there is a dearth of information on their behavior in the field. This paper provides a large-scale field study covering many millions of drive days, ten different drive models, different flash technologies (MLC, eMLC, SLC) over 6 years of production use in Google's data centers. We study a wide range of reliability characteristics and come to a number of unexpected conclusions. For example, raw bit error rates (RBER) grow at a much slower rate with wear-out than the exponential rate commonly assumed and, more importantly, they are not predictive of uncorrectable errors or other error modes. The widely used metric UBER (uncorrectable bit error rate) is not a meaningful metric, since we see no correlation between the number of reads and the number of uncorrectable errors. We see no evidence that higher-end SLC drives are more reliable than MLC drives within typical drive lifetimes. Comparing with traditional hard disk drives, flash drives have a significantly lower replacement rate in the field, however, they have a higher rate of uncorrectable errors.

1 Introduction

The use of solid state drives based on NAND flash technology in data center servers is continuously growing. As more data lives on flash, data durability and availability critically depend on flash reliability. While it is widely understood that flash drives offer substantial performance improvements relative to hard disk drives, their failure characteristics are not well understood. The datasheets that manufacturers provide only contain vague guarantees, such as the number of times a flash chip can be erased before wearing out. Our current understanding is based on work that studies flash reliability in controlled lab experiments (such as accelerated life tests), using a small population of raw flash chips under synthetic workloads. There is a dearth of studies that report on the reliability of flash drives and their failure characteristics in large-scale production use in the field.

This paper provides a detailed field study of flash reliability based on data collected over 6 years of production use in Google's data centers. The data spans many millions of drive days [1], ten different drive models, different flash technologies (MLC, eMLC and SLC) and feature sizes (ranging from 24nm to 50nm). We use this data to provide a better understanding of flash reliability in production. In particular, our contributions include a detailed analysis of the following aspects of flash reliability in the field:

1. The different types of errors experienced by flash drives and their frequency in the field (Section 3).
2. Raw bit error rates (RBER), how they are affected by factors such as wear-out, age and workload, and their relationship with other types of errors (Section 4).
3. Uncorrectable errors, their frequency and how they are affected by various factors (Section 5).
4. The field characteristics of different types of hardware failure, including block failures, chip failures and the rates of repair and replacement of drives (Section 6).
5. A comparison of the reliability of different flash technologies (MLC, eMLC, SLC drives) in Sections 7, and between flash drives and hard disk drives in Section 8.

As we will see, our analysis uncovers a number of aspects of flash reliability in the field that are different from common assumptions and reports in prior work, and will hopefully motivate further work in this area.

---

[1] The size of their fleet and the number of devices in it is considered confidential at Google, so we can not provide precise numbers. We are making sure throughout this work that the reported numbers are statistically significant.

Model name	MLC-A	MLC-B	MLC-C	MLC-D	SLC-A	SLC-B	SLC-C	SLC-D	eMLC-A	eMLC-B
Generation	1	1	1	1	1	1	1	1	2	2
Vendor	I	II	I	I	I	I	III	I	I	IV
Flash type	MLC	MLC	MLC	MLC	SLC	SLC	SLC	SLC	eMLC	eMLC
Lithography (nm)	50	43	50	50	34	50	50	34	25	32
Capacity	480GB	480GB	480GB	480GB	480GB	480GB	480GB	960GB	2TB	2TB
PE cycle limit	3,000	3,000	3,000	3,000	100,000	100,000	100,000	100,000	10,000	10,000
Avg. PE cycles	730	949	529	544	860	504	457	185	607	377

Table 1: *Overview of drive models*

2 Background on data and systems

2.1 The flash drives

The drives in our study are custom designed high performance solid state drives, which are based on commodity flash chips, but use a custom PCIe interface, firmware and driver. We focus on two generations of drives, where all drives of the same generation use the same device driver and firmware. That means that they also use the same error correcting codes (ECC) to detect and correct corrupted bits and the same algorithms for wear-levelling. The main difference between different drive models of the same generation is the type of flash chips they comprise.

Our study focuses on the 10 drive models, whose key features are summarized in Table 1. Those models were chosen as they each span millions of drive days, comprise chips from four different flash vendors, and cover the three most common types of flash (MLC, SLC, eMLC).

2.2 The data

The data was collected over a 6-year period and contains for each drive aggregated monitoring data for each day the drive was in the field. Besides daily counts for a variety of different types of errors, the data also includes daily workload statistics, including the number of read, write, and erase operations, and the number of bad blocks developed during that day. The number of read, write, and erase operations includes user-issued operations, as well as internal operations due to garbage collection. Another log records when a chip was declared failed and when a drive was being swapped to be repaired.

3 Prevalence of different error types

We begin with some baseline statistics on the frequency of different types of errors in the field. We distinguish transparent errors, which the drive can mask from the user, and non-transparent errors, which will lead to a failed user operation. The device driver of the flash drives reports the following transparent types of errors:

Correctable error: During a read operation an error is detected and corrected by the drive internal ECC (error correcting code).

Read error: A read operation experiences a (non-ECC) error, but after retrying it succeeds.
Write error: A write operation experiences an error, but after retrying the operation succeeds.
Erase error: An erase operation on a block fails.

The devices report the following types of non-transparent errors:

Uncorrectable error: A read operation encounters more corrupted bits than the ECC can correct.
Final read error: A read operation experiences an error, and even after retries the error persists.
Final write error: A write operation experiences an error that persists even after retries.
Meta error: An error accessing drive-internal metadata.
Timeout error: An operation timed out after 3 seconds.

Uncorrectable errors include errors that were detected either during user-initiated operations or internal operations due to garbage collection, while final read errors include only errors encountered during user operations.

Note that errors vary in the severity of their possible impact. Besides the distinction between transparent and non-transparent errors, the severity of non-transparent errors varies. In particular, some of these errors (final read error, uncorrectable error, meta error) lead to data loss, unless there is redundancy at higher levels in the system, as the drive is not able to deliver data that it had previously stored.

We consider only drives that were put into production at least 4 years ago (for eMLC drives 3 years ago, as they are more recent drives), and include any errors that they experienced during their first 4 years in the field. Table 2 reports for each error type the fraction of drives for each model that experienced at least one error of that type (top half of table) and the fraction of drives days that had an error of that type (bottom half of table).

3.1 Non-transparent errors

We find that the most common non-transparent errors are final read errors, i.e. read errors that cannot be resolved even after retrying the operation. Depending on the model, between 20-63% of drives experience at least one such error and between 2-6 out of 1,000 drive days are affected. We find that the count of final read errors

Model name	MLC-A	MLC-B	MLC-C	MLC-D	SLC-A	SLC-B	SLC-C	SLC-D	eMLC-A	eMLC-B
	Fraction of drives affected by different types of errors									
final read error	2.63e-01	5.64e-01	3.25e-01	3.17e-01	5.08e-01	2.66e-01	1.91e-01	6.27e-01	1.09e-01	1.27e-01
uncorrectable error	2.66e-01	5.75e-01	3.24e-01	3.24e-01	5.03e-01	2.84e-01	2.03e-01	6.34e-01	8.63e-01	9.05e-01
final write error	1.73e-02	2.11e-02	1.28e-02	1.85e-02	2.39e-02	2.33e-02	9.69e-03	5.67e-03	5.20e-02	3.16e-02
meta error	9.83e-03	7.97e-03	9.89e-03	1.93e-02	1.33e-02	3.68e-02	2.06e-02	7.04e-03	0.00e+00	0.00e+00
timeout error	5.68e-03	9.17e-03	5.70e-03	8.21e-03	1.64e-02	1.15e-02	8.47e-03	5.08e-03	0.00e+00	0.00e+00
response error	7.95e-04	3.90e-03	1.29e-03	1.88e-03	4.97e-03	2.08e-03	0.00e+00	9.78e-04	1.97e-03	8.76e-04
correctable error	9.89e-01	9.98e-01	9.96e-01	9.91e-01	9.99e-01	9.61e-01	9.72e-01	9.97e-01	9.97e-01	9.94e-01
read error	8.64e-03	1.46e-02	9.67e-03	1.12e-02	1.29e-02	1.77e-02	6.05e-03	1.02e-02	2.61e-01	2.23e-01
write error	6.37e-02	5.61e-01	6.11e-02	6.40e-02	1.30e-01	1.11e-01	4.21e-01	9.83e-02	5.46e-02	2.65e-01
erase error	1.30e-01	3.91e-01	9.70e-02	1.26e-01	6.27e-02	3.91e-01	6.84e-01	4.81e-02	1.41e-01	9.38e-02
	Fraction of drive days affected by different types of errors									
final read error	1.02e-03	1.54e-03	1.78e-03	1.39e-03	1.06e-03	9.90e-04	7.99e-04	4.44e-03	1.67e-04	2.93e-04
uncorrectable error	2.14e-03	1.99e-03	2.51e-03	2.28e-03	1.35e-03	2.06e-03	2.96e-03	6.07e-03	8.35e-03	7.82e-03
final write error	2.67e-05	2.13e-05	1.70e-05	3.23e-05	2.63e-05	4.21e-05	1.21e-05	9.42e-06	1.06e-04	6.40e-05
meta error	1.32e-05	1.18e-05	1.16e-05	3.44e-05	1.28e-05	5.05e-05	3.62e-05	1.02e-05	0.00e+00	0.00e+00
timeout error	7.52e-06	9.45e-06	7.38e-06	1.31e-05	1.73e-05	1.56e-05	1.06e-05	8.88e-06	0.00e+00	0.00e+00
response error	7.43e-07	3.45e-06	2.77e-06	2.08e-06	4.45e-06	3.61e-06	0.00e+00	2.69e-06	2.05e-06	1.11e-06
correctable error	8.27e-01	7.53e-01	8.49e-01	7.33e-01	7.75e-01	6.13e-01	6.48e-01	9.00e-01	9.38e-01	9.24e-01
read error	7.94e-05	2.75e-05	3.83e-05	7.19e-05	3.07e-05	5.85e-05	1.36e-05	2.91e-05	2.81e-03	5.10e-03
write error	1.12e-04	1.40e-03	1.28e-04	1.52e-04	2.40e-04	2.93e-04	1.21e-03	4.80e-04	2.07e-04	4.78e-04
erase error	2.63e-04	5.34e-04	1.67e-04	3.79e-04	1.12e-04	1.30e-03	4.16e-03	1.88e-04	3.53e-04	4.36e-04

Table 2: *The prevalence of different types of errors. The top half of the table shows the fraction of drives affected by each type of error, and the bottom half the fraction of drive* days *affected.*

and that of uncorrectable errors is strongly correlated and conclude that these final read errors are almost exclusively due to bit corruptions beyond what the ECC can correct. For all drive models, final read errors are around two orders of magnitude more frequent (in terms of the number of drive days they affect) than any of the other non-transparent types of errors.

In contrast to read errors, write errors rarely turn into non-transparent errors. Depending on the model, 1.5-2.5% of drives and 1-4 out of 10,000 drive days experience a final write error, i.e. a failed write operation that did not succeed even after retries. The difference in the frequency of final read and final write errors is likely due to the fact that a failed write will be retried at other drive locations. So while a failed read might be caused by only a few unreliable cells on the page to be read, a final write error indicates a larger scale hardware problem.

Meta errors happen at a frequency comparable to write errors, but again at a much lower frequency than final read errors. This might not be surprising given that a drive contains much less meta-data than real data, which lowers the chance of encountering an error accessing meta data. Other non-transparent errors (timeout and response errors) are rare, typically affecting less than 1% of drives and less than 1 in 100,000 drive days.

3.2 Transparent errors

Maybe not surprisingly, we find that correctable errors are the most common type of transparent error. Virtually all drives have at least some correctable errors, and the majority of drive days (61-90%) experience correctable errors. We discuss correctable errors, including a study of raw bit error rates (RBER), in more detail in Section 4.

The next most common transparent types of error are write errors and erase errors. They typically affect 6-10% of drives, but for some models as many as 40-68% of drives. Generally less than 5 in 10,000 days experience those errors. The drives in our study view write and erase errors as an indication of a block failure, a failure type that we will study more closely in Section 6.

Errors encountered during a read operations are rarely transparent, likely because they are due to bit corruption beyond what ECC can correct, a problem that is not fixable through retries. Non-final read errors, i.e. read errors that can be recovered by retries, affect less than 2% of drives and less than 2-8 in 100,000 drive days.

In summary, besides correctable errors, which affect the majority of drive days, transparent errors are rare in comparison to all types of non-transparent errors. The most common type of non-transparent errors are uncorrectable errors, which affect 2–6 out of 1,000 drive days.

4 Raw bit error rates (RBER)

The standard metric to evaluate flash reliability is the raw bit error rate (RBER) of a drive, defined as the number of corrupted bits per number of total bits read (including correctable as well as uncorrectable corruption events). The second generation of drives (i.e. models eMLC-A and eMLC-B) produce precise counts of the number of corrupted bits and the number of bits read, allowing us to accurately determine RBER. The first generation of

Model name	MLC-A	MLC-B	MLC-C	MLC-D	SLC-A	SLC-B	SLC-C	SLC-D	eMLC-A	eMLC-B
Median RBER	2.1e-08	3.2e-08	2.2e-08	2.4e-08	5.4e-09	6.0 e-10	5.8 e-10	8.5 -09	1.0 e-05	2.9 e-06
95%ile RBER	2.2e-06	4.6e-07	1.1e-07	1.9e-06	2.8e-07	1.3e-08	3.4e-08	3.3e-08	5.1e-05	2.6e-05
99%ile RBER	5.8e-06	9.1e-07	2.3e-07	2.7e-05	6.2e-06	2.2e-08	3.5e-08	5.3e-08	1.2e-04	4.1e-05

Table 3: *Summary of raw bit error rates (RBER) for different models*

drives report accurate counts for the number of bits read, but for each page, consisting of 16 data chunks, only report the number of corrupted bits in the data chunk that had the largest number of corrupted bits. As a result, in the (unlikely) absolute worst case, where all chunks have errors and they all have the same number of errors as the worst chunk, the RBER rates could be 16X higher than the drives record. While irrelevant when comparing drives within the same generation, this subtlety must be kept in mind when comparing across generations.

4.1 A high-level view of RBER

Table 3 shows for each drive model the median RBER across all drives for that model, as well as the 95th and 99th percentile. We decided to work with medians and percentiles since we find averages to be heavily biased by a few outliers, making it hard to identify any trends.

We observe large differences in the RBER across different drive models, ranging from as little as 5.8e-10 to more than 3e-08 for drives of the first generation. The differences are even larger when considering the 95th or 99th percentile RBER, rather than the median. For example, the 99th percentiles of RBER ranges from 2.2e-08 for model SLC-B to 2.7e-05 for MLC-D. Even within drives of the same model, there are large differences: the RBER of a drive in the 99th percentile tends to be at least an order of magnitude higher than the RBER of the median drive of the same model.

The difference in RBER between models can be partially explained by differences in the underlying flash technology. RBER rates for the MLC models are orders of magnitudes higher than for the SLC models, so the higher price point for the SLC models pays off with respect to RBER. We will see in Section 5 whether these differences will translate to differences in user-visible, non-transparent errors.

The eMLC models report RBER that are several orders of magnitude larger than for the other drives. Even taking into account that the RBER for the first generation drives are a lower bound and might in the worst case be 16X higher, there is still more than an order of magnitude difference. We speculate that feature size might be a factor, as the two eMLC models have the chips with the smallest lithography of all models.

Finally, there is not one vendor that consistently outperforms the others. Within the group of SLC and eMLC drives, respectively, the same vendor is responsible for
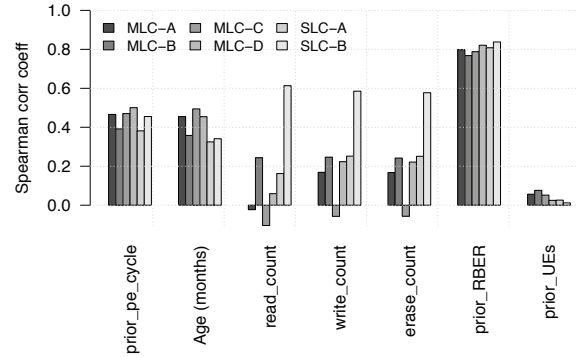

Figure 1: *The Spearman rank correlation coefficient between the RBER observed in a drive month and other factors.*

one of the worst and the best models in the group.

In summary, RBER varies greatly across drive models and also across drives within the same model. This motivates us to further study what factors affect RBER.

4.2 What factors impact RBER

In this section, we consider the effect of a number of factors on RBER: wear-out from program erase (PE) cycles; physical age, i.e. the number of months a device has been in the field, independently of PE cycles; workload, measured by the number of read, write, and erase operations, as an operation to a page can potentially disturb surrounding cells; and the presence of other errors.

We study the effect of each factor on RBER in two different ways. We use visual inspection by plotting the factor against RBER and we quantify the relationship using correlation coefficients. We use the Spearman rank correlation coefficient as it can also capture non-linear relationships, as long as they are monotonic (in contrast, for example, to the Pearson correlation coefficient).

Before analyzing individual factors in detail, we present a summary plot in Figure 1. The plot shows the Spearman rank correlation coefficient between the RBER observed in a given drive month, and other factors that were present, including the device age in months, the number of previous PE cycles, the number of read, write or erase operations in that month, the RBER observed in the previous month and the number of uncorrectable errors (UEs) in the previous month. Values for the Spearman correlation coefficient can range from -1
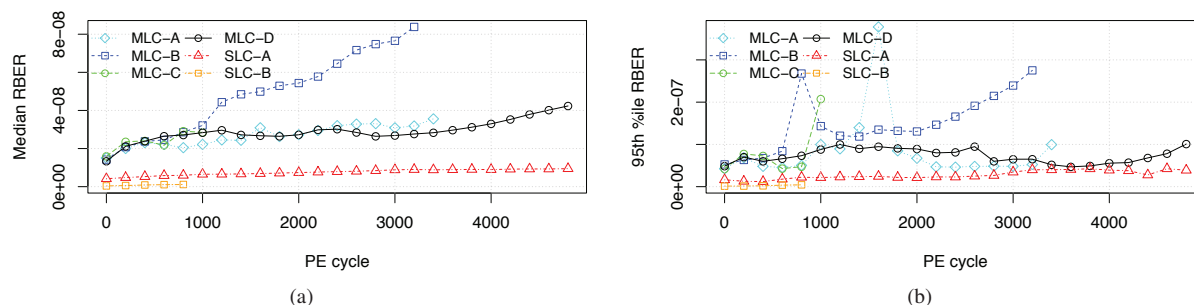
Figure 2: *The figures show the median and the 95th percentile RBER as a function of the program erase (PE) cycles.*

(strong negative correlation) to +1 (strong positive correlation). Each group of bars shows the correlation coefficients between RBER and one particular factor (see label on X-axis) and the different bars in each group correspond to the different drive models. All correlation coefficients are significant at more than 95% confidence.

We observe that all of the factors, except the prior occurrence of uncorrectable errors, show a clear correlation with RBER for at least some of the models. We also note that some of these correlations might be spurious, as some factors might be correlated with each other. We will therefore investigate each factor in more detail in the following subsections.

4.2.1 RBER and wear-out

As the endurance of flash cells is limited, RBER rates are expected to grow with the number of program erase (PE) cycles, with rates that have previously been reported as exponential [5, 8, 18, 22]. The high correlation coefficients between RBER and PE cycles in Figure 1 confirm that there is a correlation.

To study the effect of PE cycles on RBER in more detail, the two graphs in Figure 2 plot the median and the 95th percentile RBER against the number of PE cycles. We obtain these graphs by dividing all drive days in our data into different bins, based on their PE cycle count, and then determine the median and 95th percentile RBER across all days in a bin.

We observe that, as expected, RBER grows with the number of PE cycles, both in terms of median and 95th percentile RBER. However, the growth rate is slower than the commonly assumed exponential growth, and more closely resembles a linear increase. We verified this observation through curve fitting: we fit a linear model and an exponential model to the data and find that the linear model has a better fit than the exponential model.

The second interesting observation is that the RBER rates under wear-out vary greatly across drive models, even for models that have very similar RBER rates for low PE cycles. For example, the four MLC models start out with nearly identical RBER at very low PE cycles,

but by the time they reach their PE cycle limit (3,000 for all MLC models) there is a 4X difference between the model with the highest and the lowest RBER.

Finally, we find that the increase in RBER is surprisingly smooth, even when a drive goes past its expected end of life (see for example model MLC-D with a PE cycle limit of 3,000). We note that accelerated life tests for the devices showed a rapid increase in RBER at around 3X the vendor's PE cycle limit, so vendors PE cycle limits seem to be chosen very conservatively.

4.2.2 RBER and age (beyond PE cycles)

Figure 1 shows a significant correlation between age, measured by the number of months a drive has been in the field, and RBER. However, this might be a spurious correlation, since older drives are more likely to have higher PE cycles and RBER is correlated with PE cycles.

To isolate the effect of age from that of PE cycle wear-out we group all drive months into bins using deciles of the PE cycle distribution as the cut-off between bins, e.g. the first bin contains all drive months up to the first decile of the PE cycle distribution, and so on. We verify that within each bin the correlation between PE cycles and RBER is negligible (as each bin only spans a small PE cycle range). We then compute the correlation coefficient between RBER and age separately for each bin. We perform this analysis separately for each model, so that any observed correlations are not due to differences between younger and older drive models, but purely due to younger versus older drives within the same model.

We observe that even after controlling for the effect of PE cycles in the way described above, there is still a significant correlation between the number of months a device has been in the field and its RBER (correlation coefficients between 0.2 and 0.4) for all drive models.

We also visualize the effect of drive age, by separating out drive days that were observed at a young drive age (less than one year) and drive days that were observed when a drive was older (4 years or more) and then plotting each group's RBER as a function of PE cycles. The results for one drive model (MLC-D) are shown in Fig-
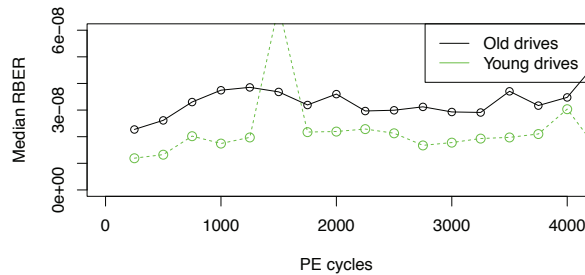
Figure 3: *RBER rates as a function of P/E cycles for young and old drives, showing that age has an effect on RBER, independently of P/E cycle induced wear-out.*

ure 3. We see a marked difference in the RBER rates between the two groups, across all PE cycles.

We conclude that age, as measured by days in the field, has a significant effect on RBER, independently of cell wear-out due to PE cycles. That means there must be other aging mechanisms at play, such as silicon aging.

4.2.3 RBER and workload

Bit errors are thought to be caused by one of four different mechanisms: retention errors, where a cell loses charge over time; read disturb errors, where a read operation disturbs the charge in a nearby cell; write disturb errors, where a write disturbs the charge in a nearby cell; or an incomplete erase errors, where an erase operation did not fully reset the cells in an erase block.

Errors that are of the latter three types (read disturb, write disturb, incomplete erase) will be correlated with workload, so understanding the correlation between RBER and workload helps us understand the prevalence of different error mechanisms. A recent field study [16] concludes that errors in the field are dominated by retention errors, while read disturb errors are negligible.

Figure 1 shows a significant correlation between the RBER in a given drive month and the number of read, write, and erase operations in the same month for some models (e.g. a correlation coefficient above 0.2 for model MLC-B and above 0.6 for model SLC-B). However, this might be a spurious correlation, as the per-month workload might be correlated with the total number of PE cycles seen so far. We use the same technique as described in Section 4.2.2 to isolate the effects of workload from that of PE cycles, by binning the drive months based on the prior PE cycles, and then determining correlation coefficients separately for each bin.

We find that the correlation between the number of read operations in a given drive month and the RBER in the same month does persist for models MLC-B and SLC-B, even when controlling for the PE cycles. We also repeat a similar analysis, where we isolate the effect of read operations from the count of concurrent write and erase operations, and find that for model SLC-B the correlation between RBER and read counts persists.

Figure 1 also showed a correlation between RBER and write and erase operations. We therefore repeat the same analysis we performed for read operations, for write and erase operations. We find that the correlation between RBER and write and erase operations is not significant, when controlling for PE cycles and read operations.

We conclude that there are drive models, where the effect of read disturb is significant enough to affect RBER. On the other hand there is no evidence for a significant impact of write disturb and incomplete erase operations on RBER.

4.2.4 RBER and lithography

Differences in feature size might partially explain the differences in RBER across models using the same technology, i.e. MLC or SLC. (Recall Table 1 for an overview of the lithography of different models in our study.) For example, the two SLC models with a 34nm lithography (models SLC-A and SLC-D) have RBER that are an order of magnitude higher than the two 50nm models (models SLC-B and SLC-C). For the MLC models, the only 43nm model (MLC-B) has a median RBER that is 50% higher than that of the other three models, which are all 50nm. Moreover, this difference in RBER increases to 4X with wear-out, as shown in Figure 2. Finally, their smaller lithography might explain the higher RBER for the eMLC drives compared to the MLC drives.

In summary, there is clear evidence that lithography affects RBER.

4.2.5 Presence of other errors

We investigate the relationship between RBER and other errors (such as uncorrectable errors, timeout errors, etc.), in particular whether RBER is higher in a month that also experiences other types of errors.

Figure 1 shows that while RBER experienced in the previous month is very predictive of future RBER (correlation coefficient above 0.8), there is no significant correlation between uncorrectable errors and RBER (see the right-most group of bars in Figure 1). Correlation coefficients are even lower for other error types (not shown in plot). We will further investigate the relationship between RBER and uncorrectable errors in Section 5.2.

4.2.6 Effect of other factors

We find evidence that there are factors with significant impact on RBER that our data does not directly account for. In particular, we observe that the RBER for a particular drive model varies depending on the cluster where the drive is deployed. One illustrative example is Figure 4, which shows RBER against PE cycles for drives of
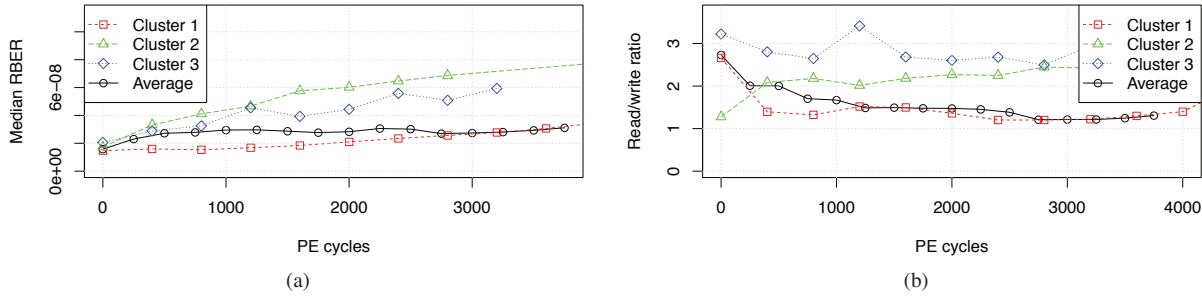
Figure 4: *Figure (a) shows the median RBER rates as a function of PE cycles for model MLC-D for three different clusters. Figure (b) shows for the same model and clusters the read/write ratio of the workload.*

model MLC-D in three different clusters (dashed lines) and compares it to the RBER for this model across its entire population (solid line). We find that these differences persist even when we control for other factors, such as age or read count.

One possible explanation are differences in the type of workload in different clusters, as we observe that those clusters, whose workload has the highest read/write ratios, tend to be among the ones with the highest RBER. For example, Figure 4(b) shows the read/write ratio of model MLC-D. However, the read/write ratio does not explain differences across clusters for all models, so there might be other factors the data does not account for, such as environmental factors or other workload parameters.

4.3 RBER in accelerated life tests

Much academic work and also tests during the procurement phase in industry rely on accelerated life tests to derive projections for device reliability in the field. We are interested in how well predictions from such tests reflect field experience.

Analyzing results from tests performed during the procurement phase at Google, following common methods for test acceleration [17], we find that field RBER rates are significantly higher than the projected rates. For example, for model eMLC-A the median RBER for drives in the field (which on average reached 600 PE cycles at the end of data collection) is 1e-05, while under test the RBER rates for this PE cycle range were almost an order of magnitude lower and didn't reach comparable rates until more than 4,000 PE cycles. This indicates that it might be very difficult to accurately predict RBER in the field based on RBER estimates from lab tests.

We also observe that some types of error, seem to be difficult to produce in accelerated tests. For example, for model MLC-B, nearly 60% of drives develop uncorrectable errors in the field and nearly 80% develop bad blocks. Yet in accelerated tests none of the six devices under test developed any uncorrectable errors or

bad blocks until the drives reached more than 3X of their PE cycle limit. For the eMLC models, more than 80% develop uncorrectable errors in the field, while in accelerated tests no device developed uncorrectable errors before 15,000 PE cycles.

We also looked at RBER reported in previous work, which relied on experiments in controlled environments. We find that previously reported numbers span a very large range. For example, Grupp et al. [10, 11] report RBER rates for drives that are close to reaching their PE cycle limit. For SLC and MLC devices with feature sizes similar to the ones in our work (25-50nm) the RBER in [11] ranges from 1e-08 to 1e-03, with most drive models experiencing RBER close to 1e-06. The three drive models in our study that reach their PE cycle limit experienced RBER between 3e-08 to 8e-08. Even taking into account that our numbers are lower bounds and in the absolute worst case could be 16X higher, or looking at the 95th percentile of RBER, our rates are significantly lower.

In summary, while the field RBER rates are higher than in-house projections based on accelerated life tests, they are lower than most RBER reported in other work for comparable devices based on lab tests. This suggests that predicting field RBER in accelerated life tests is not straight-forward.

5   Uncorrectable errors

Given the high prevalence of uncorrectable errors (UEs) we observed in Section 3, we study their characteristics in more detail in this section, starting with a discussion of what metric to use to measure UEs, their relationship with RBER and then moving to the impact of various factors on UEs.

5.1   Why UBER is meaningless

The standard metric used to report uncorrectable errors is UBER, i.e. the number of uncorrectable bit errors per
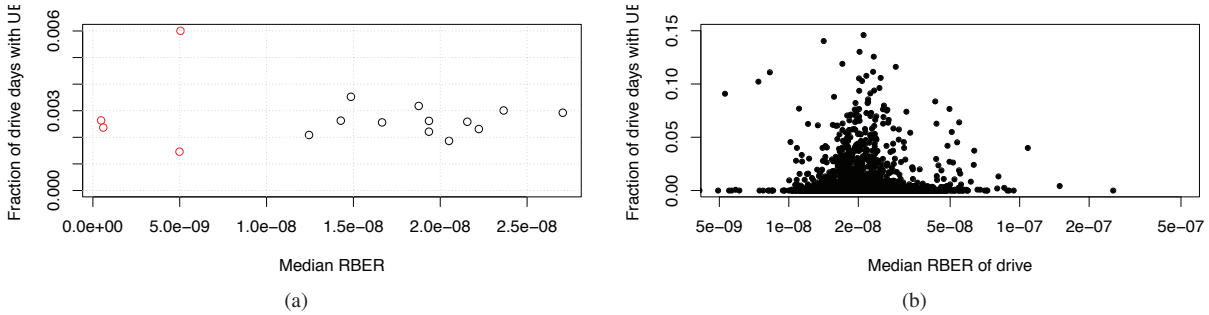
Figure 5: *The two figures show the relationship between RBER and uncorrectable errors for different drive models (left) and for individual drives within the same model (right).*

total number of bits read. This metric makes the implicit assumption that the number of uncorrectable errors is in some way tied to the number of bits read, and hence should be normalized by this number.

This assumption makes sense for correctable errors, where we find that the number of errors observed in a given month is strongly correlated with the number of reads in the same time period (Spearman correlation coefficient larger than 0.9). The reason for this strong correlation is that one corrupted bit, as long as it is correctable by ECC, will continue to increase the error count with every read that accesses it, since the value of the cell holding the corrupted bit is not immediately corrected upon detection of the error (drives only periodically rewrite pages with corrupted bits).

The same assumption does not hold for uncorrectable errors. An uncorrectable error will remove the affected block from further usage, so once encountered it will not continue to contribute to error counts in the future. To formally validate this intuition, we used a variety of metrics to measure the relationship between the number of reads in a given drive month and the number of uncorrectable errors in the same time period, including different correlation coefficients (Pearson, Spearman, Kendall) as well as visual inspection. In addition to the number of uncorrectable errors, we also looked at the incidence of uncorrectable errors (e.g. the probability that a drive will have at least one within a certain time period) and their correlation with read operations.

We find no evidence for a correlation between the number of reads and the number of uncorrectable errors. The correlation coefficients are below 0.02 for all drive models, and graphical inspection shows no higher UE counts when there are more read operations.

As we will see in Section 5.4, also write and erase operations are uncorrelated with uncorrectable errors, so an alternative definition of UBER, which would normalize by write or erase operations instead of read operations, would not be any more meaningful either.

We therefore conclude that UBER is not a meaningful

metric, except maybe in controlled environments where the number of read operations is set by the experimenter. If used as a metric in the field, UBER will artificially decrease the error rates for drives with high read count and artificially inflate the rates for drives with low read counts, as UEs occur independently of the number of reads.

5.2 Uncorrectable errors and RBER

RBER is relevant because it serves as a measure for general drive reliability, and in particular for the likelihood of experiencing UEs. Mielke et al. [18] first suggested to determine the expected rate of uncorrectable errors as a function of RBER. Since then many system designers, e.g. [2,8,15,23,24], have used similar methods to, for example, estimate the expected frequency of uncorrectable errors depending on RBER and the type of error correcting code being used.

The goal of this section is to characterize how well RBER predicts UEs. We begin with Figure 5(a), which plots for a number of first generation drive models [2] their median RBER against the fraction of their drive days with UEs. Recall that all models within the same generation use the same ECC, so differences between models are not due to differences in ECC. We see no correlation between RBER and UE incidence. We created the same plot for 95th percentile of RBER against UE probability and again see no correlation.

Next we repeat the analysis at the granularity of individual drives, i.e. we ask whether drives with higher RBER have a higher incidence of UEs. As an example, Figure 5(b) plots for each drive of model MLC-C its median RBER against the fraction of its drive days with UEs. (Results are similar for 95th percentile of RBER.) Again we see no correlation between RBER and UEs.

Finally, we perform an analysis at a finer time granularity, and study whether drive months with higher RBER are more likely to be months that experience a UE. Fig-

---

[2]Some of the 16 models in the figure were not included in Table 1, as they do not have enough data for some other analyses in the paper.
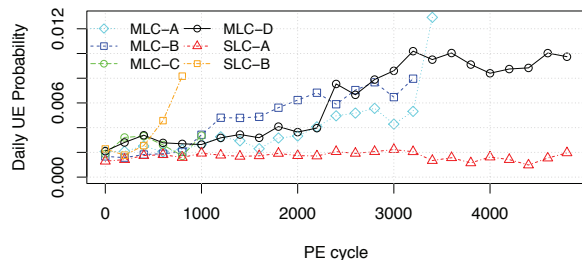
Figure 6: *The figure shows the daily probability of a drive experiencing an uncorrectable error as a function of the PE cycles the drive has experienced.*

Figure 7: *The monthly probability of a UE as a function of whether there were previous errors of various types.*

ure 1 already indicated that the correlation coefficient between UEs and RBER is very low. We also experimented with different ways of plotting the probability of UEs as a function of RBER for visual inspection, and did not find any indication of a correlation.

In summary, we conclude that RBER is a poor predictor of UEs. This might imply that the failure mechanisms leading to RBER are different from those leading to UEs (e.g. retention errors in individual cells versus larger scale issues with the device).

5.3 Uncorrectable errors and wear-out

As wear-out is one of the main concerns with flash drives, Figure 6 shows the daily probability of developing an uncorrectable error as a function of the drive's P/E cycles. We observe that the UE probability continuously increases with age. However, as was the case for RBER, the increase is slower than commonly assumed: both visual inspection and curve fitting indicate that the UEs grow linearly with PE cycles rather than exponentially.

Also two other observations we made for RBER apply to UEs as well: First, there is no sharp increase in error probabilities after the PE cycle limit is reached, e.g. consider model MLC-D in Figure 6, whose PE cycle limit is 3,000. Second, error incidence varies across models, even within the same class. However, the differences are not as large as they were for RBER.

Finally, further supporting the observations we make in Section 5.2 we find that within a class of models (MLC versus SLC) the models with the lowest RBER rates for a given PE cycle count are not necessarily the ones with the lowest probabilities of UEs. For example, for 3,000 PE cycles model MLC-D had RBER rates 4X lower than that of MLC-B, yet its UE probability at the same PE cycles is slightly higher than that of MLC-B.

5.4 Uncorrectable errors and workload

For the same reasons that workload can affect RBER (recall Section 4.2.3) one might expect an effect on UEs. For example, since we observed read disturb errors af-

fecting RBER, read operations might also increase the chance of uncorrectable errors.

We performed a detailed study of the effect of workload on UEs. However, as noted in Section 5.1, we find no correlation between UEs and the number of read operations. We repeated the same analysis for write and erase operations and again see no correlation.

Note that at first glance one might view the above observation as a contradiction to our earlier observation that uncorrectable errors are correlated with PE cycles (which one would expect to be correlated with the number of write and erase operations). However, in our analysis of the effect of PE cycles we were correlating the number of uncorrectable errors in a given month with the total number of PE cycles the drive has experienced in its life at that point (in order to measure the effect of wear-out). When studying the effect of workload, we look at whether drive months that had a higher read/write/erase count in *that particular month* also had a higher chance of uncorrectable errors in that particular month, i.e. we do not consider the cumulative count of read/write/erase operations.

We conclude that read disturb errors, write disturb errors or incomplete erase operations are not a major factor in the development of UEs.

5.5 Uncorrectable errors and lithography

Interestingly, the effect of lithography on uncorrectable errors is less clear than for RBER, where smaller lithography translated to higher RBER, as expected. Figure 6 shows, for example, that model SLC-B has a higher rate of developing uncorrectable errors than SLC-A, although SLC-B has the larger lithography (50nm compared to 34nm for model SLC-A). Also, the MLC model with the smallest feature size (model MLC-B), does not generally have higher rates of uncorrectable errors than the other models. In fact, during the first third of its life ($0-1,000$ PE cycles) and the last third ($> 2,200$ PE cycles) it has lower rates than, for example, model MLC-D. Recall, that all MLC and SLC drives use the same ECC, so these effects cannot be attributed to differences in the ECC.

Model name	MLC-A	MLC-B	MLC-C	MLC-D	SLC-A	SLC-B	SLC-C	SLC-D	eMLC-A	eMLC-B
Drives w/ bad blocks (%)	31.1	79.3	30.7	32.4	39.0	64.6	91.5	64.0	53.8	61.2
Median # bad block	2	3	2	3	2	2	4	3	2	2
Mean # bad block	772	578	555	312	584	570	451	197	1960	557
Drives w/ fact. bad blocks (%)	99.8	99.9	99.8	99.7	100	97.0	97.9	99.8	99.9	100
Median # fact. bad block	1.01e+03	7.84e+02	9.19e+02	9.77e+02	5.00e+01	3.54e+03	2.49e+03	8.20e+01	5.42e+02	1.71e+03
Mean # fact. bad block	1.02e+03	8.05e+02	9.55e+02	9.94e+02	3.74e+02	3.53e+03	2.55e+03	9.75e+01	5.66e+02	1.76e+03

Table 4: *Overview of prevalence of factory bad blocks and new bad blocks developing in the field*

Overall, we find that lithography has a smaller effect on uncorrectable errors than expected and a smaller effect than what we observed for RBER.

5.6   Other types of errors versus UEs

Next we look at whether the presence of other errors increases the likelihood of developing uncorrectable errors. Figure 7 shows the probability of seeing an uncorrectable error in a given drive month depending on whether the drive saw different types of errors at some previous point in its life (yellow) or in the previous month (green bars) and compares it to the probability of seeing an uncorrectable error in an average month (red bar).

We see that all types of errors increase the chance of uncorrectable errors. The increase is strongest when the previous error was seen recently (i.e. in the previous month, green bar, versus just at any prior time, yellow bar) and if the previous error was also an uncorrectable error. For example, the chance of experiencing an uncorrectable error in a month following another uncorrectable error is nearly 30%, compared to only a 2% chance of seeing an uncorrectable error in a random month. But also final write errors, meta errors and erase errors increase the UE probability by more than 5X.

In summary, prior errors, in particular prior uncorrectable errors, increase the chances of later uncorrectable errors by more than an order of magnitude.

6   Hardware failures

6.1   Bad blocks

Blocks are the unit at which erase operations are performed. In our study we distinguish blocks that fail in the field, versus factory bad blocks that the drive was shipped with. The drives in our study declare a block bad after a final read error, a write error, or an erase error, and consequently remap it (i.e. it is removed from future usage and any data that might still be on it and can be recovered is remapped to a different block).

The top half of Table 4 provides for each model the fraction of drives that developed bad blocks in the field, the median number of bad blocks for those drives that had bad blocks, and the average number of bad blocks among drives with bad blocks. We only include drives
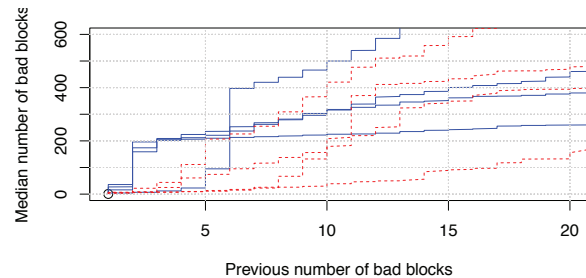

Figure 8: *The graph shows the median number of bad blocks a drive will develop, as a function of how many bad blocks it has already developed.*

that were put into production at least four years ago, and consider only bad blocks that developed during the first four years in the field. The bottom half of the table provides statistics for factory bad blocks.

6.1.1   Bad blocks developed in the field

We find that bad blocks are a frequent occurrence: Depending on the model, 30-80% of drives develop bad blocks in the field. A study of the cumulative distribution function (CDF) for the number of bad blocks per drive shows that most drives with bad blocks experience only a small number of them: the median number of bad blocks for drives with bad blocks is 2-4, depending on the model. However, if drives develop more than that they typically develop many more. Figure 8 illustrates this point. The figure shows the median number of bad blocks drives develop, as a function of how many bad blocks a drive has already experienced. The blue solid lines correspond to MLC models, while the red dashed lines correspond to the SLC models. We observe, in particular for MLC drives, a sharp increase after the second bad block is detected, when the median number of total bad blocks jumps to close to 200, i.e. 50% of those drives that develop two bad blocks will develop close to 200 or more bad blocks in total.

While we don't have access to chip-level error counts, bad block counts on the order of hundreds are likely due to chip failure, so Figure 8 indicates that after experiencing only a handful of bad blocks there is a large chance of developing a chip failure. This might imply potential for predicting chip failures, based on previous counts of

Model name	MLC-A	MLC-B	MLC-C	MLC-D	SLC-A	SLC-B	SLC-C	SLC-D	eMLC-A	eMLC-B
Drives w/ bad chips (%)	5.6	6.5	6.6	4.2	3.8	2.3	1.2	2.5	1.4	1.6
Drives w/ repair (%)	8.8	17.1	8.5	14.6	9.95	30.8	25.7	8.35	10.9	6.2
MTBRepair (days)	13,262	6,134	12,970	5,464	11,402	2,364	2,659	8,547	8,547	14,492
Drives replaced (%)	4.16	9.82	4.14	6.21	5.02	10.31	5.08	5.55	4.37	3.78

Table 5: *The fraction of drives for each model that developed bad chips, entered repairs and were replaced during the first four years in the field.*

bad blocks, and by potentially taking other factors (such as age, workload, PE cycles) into account.

Besides the frequency of bad blocks, we are also interested in how bad blocks are typically detected – in a write or erase operation, where the block failure is transparent to the user, or in a final read error, which is visible to the user and creates the potential for data loss. While we don't have records for individual block failures and how they were detected, we can turn to the observed frequencies of the different types of errors that indicate a block failure. Going back to Table 2, we observe that for all models, the incidence of erase errors and write errors is lower than that of final read errors, indicating that most bad blocks are discovered in a non-transparent way, in a read operation.

6.1.2 Factory bad blocks

While the discussion above focused on bad blocks that develop in the field, we note that nearly all drives ($>99\%$ for most models) are shipped with factory bad blocks and that the number of factory bad blocks can vary greatly between models, ranging from a median number of less than 100 for two of the SLC models, to more typical values in the range of 800 or more for the others. The distribution of factory bad blocks looks close to a normal distribution, with mean and median being close in value.

Interestingly, we find that the number of factory bad blocks is to some degree predictive of other issues the drive might develop in the field: For example, we observe that for all but one drive model the drives that have above the 95%ile of factory bad blocks have a higher fraction of developing new bad blocks in the field and final write errors, compared to an average drive of the same model. They also have a higher fraction that develops some type of read error (either final or non-final). The drives in the bottom 5%ile have a lower fraction of timeout errors than average.

We summarize our observations regarding bad blocks as follows: Bad blocks are common: 30-80% of drives develop at least one in the field. The degree of correlation between bad blocks in a drive is surprisingly strong: after only 2-4 bad blocks on a drive, there is a 50% chance that hundreds of bad blocks will follow. Nearly all drives come with factory bad blocks, and the number of factory bad blocks shows a correlation with the number of bad

blocks the drive will develop in the field, as well as a few other errors that occur in the field.

6.2 Bad chips

The drives in our study consider a chip failed if more than 5% of its blocks have failed, or after the number of errors it has experienced within a recent time window exceed a certain threshold. Some commodity flash drives contain spare chips, so that the drive can tolerate a bad chip by remapping it to a spare chip. The drives in our study support a similar feature. Instead of working with spare chips, a bad chip is removed from further usage and the drive continues to operate with reduced capacity. The first row in Table 5 reports the prevalence of bad chips.

We observe that around 2-7% of drives develop bad chips during the first four years of their life. These are drives that, without mechanisms for mapping out bad chips, would require repairs or be returned to the vendor.

We also looked at the symptoms that led to the chip being marked as failed: across all models, around two thirds of bad chips are declared bad after reaching the 5% threshold on bad blocks, the other third after exceeding the threshold on the number of days with errors. We note that the vendors of all flash chips in these drives guarantee that no more than 2% of blocks on a chip will go bad while the drive is within its PE cycle limit. Therefore, the two thirds of bad chips that saw more than 5% of their blocks fail are chips that violate vendor specs.

6.3 Drive repair and replacement

A drive is being swapped and enters repairs if it develops issues that require manual intervention by a technician. The second row in Table 5 shows the fraction of drives for each model that enter repairs at some point during the first four years of their lives.

We observe significant differences in the repair rates between different models. While for most drive models 6-9% of their population at some point required repairs, there are some drive models, e.g. SLC-B and SLC-C, that enter repairs at significantly higher rates of 30% and 26%, respectively. Looking at the time between repairs (i.e. dividing the total number of drive days by the total number of repair events, see row 3 in Table 5) we see a range of a couple of thousand days between repairs for the worst models to nearly 15,000 days between repairs

for the best models. We also looked at how often in their life drives entered repairs: The vast majority (96%) of drives that go to repairs, go there only once in their life.

We also check whether a drive returns to the fleet after visiting repairs or not, the latter indicating that it was permanently replaced. The fourth row in Table 5 shows that most models see around 5% of their drives permanently removed from the field within 4 years after being deployed, while the worst models (MLC-B and SLC-B) see around 10% of their drives replaced. For most models less than half as many drives are being replaced as being sent to repairs, implying that at least half of all repairs are successful.

7 Comparison of MLC, eMLC, and SLC drives

eMLC and SLC drives target the enterprise market and command a higher price point. Besides offering a higher write endurance, there is also the perception that the enterprise drives are higher-end drives, which are overall more reliable and robust. This section evaluates the accuracy of this perception.

Revisiting Table 3, we see that this perception is correct when it comes to SLC drives and their RBER, as they are orders of magnitude lower than for MLC and eMLC drives. However, Tables 2 and 5 show that SLC drives do not perform better for those measures of reliability that matter most in practice: SLC drives don't have lower repair or replacement rates, and don't typically have lower rates of non-transparent errors.

The eMLC drives exhibit higher RBER than the MLC drives, even when taking into account that the RBER for MLC drives are lower bounds and could be up to 16X higher in the worst case. However, these differences might be due to their smaller lithography, rather than other differences in technology.

Based on our observations above, we conclude that SLC drives are not generally more reliable than MLC drives.

8 Comparison with hard disk drives

An obvious question is how flash reliability compares to that of hard disk drives (HDDs), their main competitor. We find that when it comes to replacement rates, flash drives win. The *annual* replacement rates of hard disk drives have previously been reported to be 2-9% [19, 20], which is high compared to the 4-10% of flash drives we see being replaced in a *4 year* period. However, flash drives are less attractive when it comes to their error rates. More than 20% of flash drives develop uncorrectable errors in a four year period, 30-80% develop bad

blocks and 2-7% of them develop bad chips. In comparison, previous work [1] on HDDs reports that only 3.5% of disks in a large population developed bad sectors in a 32 months period – a low number when taking into account that the number of sectors on a hard disk is orders of magnitudes larger than the number of either blocks or chips on a solid state drive, and that sectors are smaller than blocks, so a failure is less severe.

In summary, we find that the flash drives in our study experience significantly lower replacement rates (within their rated lifetime) than hard disk drives. On the downside, they experience significantly higher rates of uncorrectable errors than hard disk drives.

9 Related work

There is a large body of work on flash chip reliability based on controlled lab experiments with a small number of chips, focused on identifying error patterns and sources. For example, some early work [3, 4, 9, 12–14, 17, 21] investigates the effects of retention, program and read disturbance in flash chips, some newer work [5–8] studies error patterns for more recent MLC chips. We are interested in behaviour of flash drives in the field, and note that our observations sometimes differ from those previously published studies. For example, we find that RBER is not a good indicator for the likelihood of uncorrectable errors and that RBER grows linearly rather than exponentially with PE cycles.

There is only one, very recently published study on flash errors in the field, based on data collected at Facebook [16]. Our study and [16] complement each other well, as they have very little overlap. The data in the Facebook study consists of a single snapshot in time for a fleet consisting of very young (in terms of the usage they have seen in comparison to their PE cycle limit) MLC drives and has information on uncorrectable errors only, while our study is based on per-drive time series data spanning drives' entire lifecycle and includes detailed information on different types of errors, including correctable errors, and different types of hardware failures, as well as drives from different technologies (MLC, eMLC, SLC). As a result our study spans a broader range of error and failure modes, including wear-out effects across a drive's entire life. On the other hand, the Facebook study includes the role of some factors (temperature, bus power consumption, DRAM buffer usage) that our data does not account for.

Our studies overlap in only two smaller points and in both of them we reach slightly different conclusions: (1) The Facebook paper presents rates of uncorrectable errors and studies them as a function of usage. They observe significant infant mortality (which they refer to as early detection and early failure), while we don't. Be-

sides differences in burn-in testing at the two companies, which might affect infant mortality, the differences might also be due to the fact that the Facebook study presents more of a close-up view of a drive's early life (with no datapoints past a couple of hundred PE cycles for drives whose PE cycle limits are in the tens of thousands) while our view is more macroscopic spanning the entire lifecycle of a drive. (2) The Facebook study concludes that read disturb errors are not a significant factor in the field. Our view of read disturb errors is more differentiated, showing that while read disturb does not create uncorrectable errors, read disturb errors happen at a rate that is significant enough to affect RBER in the field.

10   Summary

This paper provides a number of interesting insights into flash reliability in the field. Some of these support common assumptions and expectations, while many were unexpected. The summary below focuses on the more surprising results and implications from our work:

• Between 20–63% of drives experience at least one uncorrectable error during their first four years in the field, making uncorrectable errors the most common non-transparent error in these drives. Between 2–6 out of 1,000 drive days are affected by them.

• The majority of drive days experience at least one correctable error, however other types of transparent errors, i.e. errors which the drive can mask from the user, are rare compared to non-transparent errors.

• We find that RBER (raw bit error rate), the standard metric for drive reliability, is not a good predictor of those failure modes that are the major concern in practice. In particular, higher RBER does not translate to a higher incidence of uncorrectable errors.

• We find that UBER (uncorrectable bit error rate), the standard metric to measure uncorrectable errors, is not very meaningful. We see no correlation between UEs and number of reads, so normalizing uncorrectable errors by the number of bits read will artificially inflate the reported error rate for drives with low read count.

• Both RBER and the number of uncorrectable errors grow with PE cycles, however the rate of growth is slower than commonly expected, following a linear rather than exponential rate, and there are no sudden spikes once a drive exceeds the vendor's PE cycle limit, within the PE cycle ranges we observe in the field.

• While wear-out from usage is often the focus of attention, we note that independently of usage the age of a drive, i.e. the time spent in the field, affects reliability.

• SLC drives, which are targeted at the enterprise market and considered to be higher end, are not more reliable than the lower end MLC drives.

• We observe that chips with smaller feature size tend

to experience higher RBER, but are not necessarily the ones with the highest incidence of non-transparent errors, such as uncorrectable errors.

• While flash drives offer lower field replacement rates than hard disk drives, they have a significantly higher rate of problems that can impact the user, such as uncorrectable errors.

• Previous errors of various types are predictive of later uncorrectable errors. (In fact, we have work in progress showing that standard machine learning techniques can predict uncorrectable errors based on age and prior errors with an interesting accuracy.)

• Bad blocks and bad chips occur at a signicant rate: depending on the model, 30-80% of drives develop at least one bad block and and 2-7% develop at least one bad chip during the first four years in the field. The latter emphasizes the importance of mechanisms for mapping out bad chips, as otherwise drives with a bad chips will require repairs or be returned to the vendor.

• Drives tend to either have less than a handful of bad blocks, or a large number of them, suggesting that impending chip failure could be predicted based on prior number of bad blocks (and maybe other factors). Also, a drive with a large number of factory bad blocks has a higher chance of developing more bad blocks in the field, as well as certain types of errors.

11   Acknowledgements

References

[1] BAIRAVASUNDARAM, L. N., GOODSON, G. R., PASUPATHY, S., AND SCHINDLER, J. An analysis of latent sector errors in disk drives. In *Proceedings of the 2007 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems* (New York, NY, USA, 2007), SIGMETRICS '07, ACM, pp. 289–300.

[2] BALAKRISHNAN, M., KADAV, A., PRABHAKARAN, V., AND MALKHI, D. Differential RAID: Rethinking RAID for SSD reliability. *Trans. Storage 6*, 2 (July 2010), 4:1–4:22.

[3] BELGAL, H. P., RIGHOS, N., KALASTIRSKY, I., PETERSON, J. J., SHINER, R., AND MIELKE, N. A new reliability model for post-cycling charge retention of flash memories. In *Reliability Physics Symposium Proceedings, 2002. 40th Annual* (2002), IEEE, pp. 7–20.

[4] BRAND, A., WU, K., PAN, S., AND CHIN, D. Novel read disturb failure mechanism induced by flash cycling. In *Reliability*

Physics Symposium, 1993. 31st Annual Proceedings., International (1993), IEEE, pp. 127–132.

[5] CAI, Y., HARATSCH, E. F., MUTLU, O., AND MAI, K. Error patterns in MLC NAND flash memory: Measurement, characterization, and analysis. In *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2012* (2012), IEEE, pp. 521–526.

[6] CAI, Y., HARATSCH, E. F., MUTLU, O., AND MAI, K. Threshold voltage distribution in MLC NAND flash memory: Characterization, analysis, and modeling. In *Proceedings of the Conference on Design, Automation and Test in Europe* (2013), EDA Consortium, pp. 1285–1290.

[7] CAI, Y., MUTLU, O., HARATSCH, E. F., AND MAI, K. Program interference in MLC NAND flash memory: Characterization, modeling, and mitigation. In *Computer Design (ICCD), 2013 IEEE 31st International Conference on* (2013), IEEE, pp. 123–130.

[8] CAI, Y., YALCIN, G., MUTLU, O., HARATSCH, E. F., CRISTAL, A., UNSAL, O. S., AND MAI, K. Flash correct-and-refresh: Retention-aware error management for increased flash memory lifetime. In *IEEE 30th International Conference on Computer Design (ICCD 2012)* (2012), IEEE, pp. 94–101.

[9] DEGRAEVE, R., SCHULER, F., KACZER, B., LORENZINI, M., WELLEKENS, D., HENDRICKX, P., VAN DUUREN, M., DORMANS, G., VAN HOUDT, J., HASPESLAGH, L., ET AL. Analytical percolation model for predicting anomalous charge loss in flash memories. *Electron Devices, IEEE Transactions on 51*, 9 (2004), 1392–1400.

[10] GRUPP, L. M., CAULFIELD, A. M., COBURN, J., SWANSON, S., YAAKOBI, E., SIEGEL, P. H., AND WOLF, J. K. Characterizing flash memory: Anomalies, observations, and applications. In *Proceedings of the 42Nd Annual IEEE/ACM International Symposium on Microarchitecture* (New York, NY, USA, 2009), MICRO 42, ACM, pp. 24–33.

[11] GRUPP, L. M., DAVIS, J. D., AND SWANSON, S. The bleak future of NAND flash memory. In *Proceedings of the 10th USENIX Conference on File and Storage Technologies* (Berkeley, CA, USA, 2012), FAST'12, USENIX Association, pp. 2–2.

[12] HUR, S., LEE, J., PARK, M., CHOI, J., PARK, K., KIM, K., AND KIM, K. Effective program inhibition beyond 90nm NAND flash memories. *Proc. NVSM* (2004), 44–45.

[13] JOO, S. J., YANG, H. J., NOH, K. H., LEE, H. G., WOO, W. S., LEE, J. Y., LEE, M. K., CHOI, W. Y., HWANG, K. P., KIM, H. S., ET AL. Abnormal disturbance mechanism of sub-100 nm NAND flash memory. *Japanese journal of applied physics 45*, 8R (2006), 6210.

[14] LEE, J.-D., LEE, C.-K., LEE, M.-W., KIM, H.-S., PARK, K.-C., AND LEE, W.-S. A new programming disturbance phenomenon in NAND flash memory by source/drain hot-electrons generated by GIDL current. In *Non-Volatile Semiconductor Memory Workshop, 2006. IEEE NVSMW 2006. 21st* (2006), IEEE, pp. 31–33.

[15] LIU, R., YANG, C., AND WU, W. Optimizing NAND flash-based SSDs via retention relaxation. In *Proceedings of the 10th USENIX conference on File and Storage Technologies, FAST 2012, San Jose, CA, USA, February 14-17, 2012* (2012), p. 11.

[16] MEZA, J., WU, Q., KUMAR, S., AND MUTLU, O. A large-scale study of flash memory failures in the field. In *Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems* (New York, NY, USA, 2015), SIGMETRICS '15, ACM, pp. 177–190.

[17] MIELKE, N., BELGAL, H. P., FAZIO, A., MENG, Q., AND RIGHOS, N. Recovery effects in the distributed cycling of flash memories. In *Reliability Physics Symposium Proceedings, 2006. 44th Annual., IEEE International* (2006), IEEE, pp. 29–35.

[18] MIELKE, N., MARQUART, T., WU, N., KESSENICH, J., BELGAL, H., SCHARES, E., TRIVEDI, F., GOODNESS, E., AND NEVILL, L. Bit error rate in NAND flash memories. In *2008 IEEE International Reliability Physics Symposium* (2008).

[19] PINHEIRO, E., WEBER, W.-D., AND BARROSO, L. A. Failure trends in a large disk drive population. In *FAST* (2007), vol. 7, pp. 17–23.

[20] SCHROEDER, B., AND GIBSON, G. A. Disk failures in the real world: What does an MTTF of 1, 000, 000 hours mean to you? In *FAST* (2007), vol. 7, pp. 1–16.

[21] SUH, K.-D., SUH, B.-H., LIM, Y.-H., KIM, J.-K., CHOI, Y.-J., KOH, Y.-N., LEE, S.-S., SUK-CHON, S.-C., CHOI, B.-S., YUM, J.-S., ET AL. A 3.3 v 32 mb NAND flash memory with incremental step pulse programming scheme. *Solid-State Circuits, IEEE Journal of 30*, 11 (1995), 1149–1156.

[22] SUN, H., GRAYSON, P., AND WOOD, B. Quantifying reliability of solid-state storage from multiple aspects. In *SNAPI* (2011).

[23] WU, G., HE, X., XIE, N., AND ZHANG, T. Exploiting workload dynamics to improve ssd read latency via differentiated error correction codes. *ACM Trans. Des. Autom. Electron. Syst. 18*, 4 (Oct. 2013), 55:1–55:22.

[24] ZAMBELLI, C., INDACO, M., FABIANO, M., DI CARLO, S., PRINETTO, P., OLIVO, P., AND BERTOZZI, D. A cross-layer approach for new reliability-performance trade-offs in MLC NAND flash memories. In *Proceedings of the Conference on Design, Automation and Test in Europe* (San Jose, CA, USA, 2012), DATE '12, EDA Consortium, pp. 881–886.