

大规模生产环境下的闪存可靠性研究

院（系）名 称： 计算机学院

班 级： 2017 级硕士 5 班

学 生 姓 名： 赵知非

二〇一七年十二月

摘 要

以磁盘为主的外部存储经常是整个计算机系统的性能瓶颈，难于满足现代应用的存储性能需求，但 Flash 固态盘的出现为存储系统提供了新的选择。由于其性能优异、功耗较低、体积较小的优点，固态硬盘正在成为数据center里数据持久化的主要部分。由于固态硬盘的重要性日益凸显，我们必须对其可靠性有深入的了解。虽然目前已有大量关于单个 Flash 芯片的负载测试研究，但是大部分是在可控的实验室环境下进行的，缺乏大规模部署环境下的固态硬盘行为研究。

本文在认真研读 FAST 2016 论文《Flash Reliability in Production: The Expected and the Unexpected》及相关论文的基础上撰写。原文中进行了大规模的实地研究，研究数据来自于使用 6 年以上的谷歌数据中心，涵盖 10 种不同的驱动模式，3 种不同的闪存技术（MLC, eMLC, SLC）。并在广泛研究了其可靠性特点的基础上，得出了一些意想不到的结论。例如，原始误码率（RBER）的增长速度远远低于通常假设的指数速率，并且难以预测到不可纠正错误或者其他错误模式的产生。除此之外，研究还发现以往使用的经典度量 UBER（不可纠正的误码率）并没有体现出其意义，因为我们难以证明读取次数和不可纠正错误数量之间的相关性。同时，我们也没有找到任何证据表明，在存储器的使用寿命内，高端 SLC 驱动器比 MLC 驱动器更可靠。与传统的机械硬盘相比，固态硬盘的替换率明显较低，但不可纠正错误出现的概率较高。

目 录

1	绪论	1
1.1	研究背景	1
1.2	主要研究工作	1
2	不同类型错误的出现频率研究	3
2.1	非透明错误	4
2.2	透明错误	5
3	原始误码率 (RBER) 研究	6
3.1	RBER 总览	6
3.2	影响 RBER 的因素	7
3.3	RBER 加速寿命测试	8
4	不可纠正错误 (UE) 研究	10
4.1	为什么说 UBER 是毫无意义的	10
4.2	和 RBER 的关系	11
4.3	和闪存磨损的关系	12
4.4	和工作量的关系	13
4.5	和光刻尺寸的关系	13
5	MLC, eMLC 和 SLC 驱动器的比较	15
6	与硬盘驱动器比较	16
7	总结	17
	参考文献	18

1 绪论

1.1 研究背景

基于 NAND 闪存技术的固态硬盘的在生产过程中使用越来越广泛，随着越来越多的数据存储于 Flash 存储器上，数据的持久存储能力和可用性将取决于闪存自身的可靠性。虽然人们普遍认为闪存驱动器相对于传统硬盘驱动器有着性能改进，但对其故障的特点、产生原理还不是很清楚。制造商提供的数据中仅包含一些十分模糊的故障参数，如闪存芯片在磨损前的可擦除的次数等。

针对固态硬盘可靠性来说，目前的大部分研究工作是在实验室环境下完成的（如加速寿命测试），在模拟正常工作负载情况下仅使用了少量的闪存芯片，缺乏关于闪存驱动器可靠性及其在大规模生产环境中故障特征的研究报告。本文将根据 Google 数据中心 6 年的生产使用中收集到的数据，对闪存可靠性进行详细的实际使用研究。数据涵盖十种不同的闪存型号，不同的闪存技术（MLC，eMLC 和 SLC）以及多种颗粒尺寸（范围从 24nm 到 50nm）。本文使用这些数据来更好地了解闪存存在实际生产中的可靠性。

1.2 主要研究工作

本文主要从以下几个方面来详细分析闪存可靠性：

1. 闪存驱动器在生产中遇到过的不同类型的错误及其出现的频率（第 2 章）。
2. 原始误码率（RBER）与固态硬盘磨损程度、使用年龄和工作量等因素的联系（第 3 章）。
3. 不可纠正错误出现的频率及其如何受各种因素影响（第 4 章）。
4. 比较不同闪存技术（MLC，eMLC，SLC 驱动器）的可靠性以及闪存驱动器和硬盘驱动器之间的可靠性（第 5、6 章）。

本文用于研究的闪存驱动器主要有 10 种型号，其关键参数如表 1.1 所示。

表 1.1 不同闪存型号的关键参数

模型	闪存类型	光刻工艺	容量	最大擦除次数	平均擦除次数
MLC-A	MLC	50nm	480GB	3,000	730
MLC-B	MLC	43nm	480GB	3,000	949
MLC-C	MLC	50nm	480GB	3,000	529
MLC-D	MLC	50nm	480GB	3,000	544
SLC-A	SLC	34nm	480GB	100,000	860
SLC-B	SLC	50nm	480GB	100,000	504
SLC-C	SLC	50nm	480GB	100,000	457
SLC-D	SLC	34nm	960GB	100,000	185
eMLC-A	eMLC	25nm	960GB	10,000	607
eMLC-B	eMLC	32nm	2TB	10,000	377

最后，本文将分析并揭示闪存可靠性方面不同于以往工作一般假设的部分，并有望推动这一领域的进步。

2 不同类型错误的出现频率研究

本文从一些基准统计数据中得到了不同类型的错误发生频率。我们可以将错误分为两类，即用户可忽略的透明错误，与有可能导致用户操作失败的非透明错误。

闪存的设备驱动程序通常会报告的透明错误有以下几种类型：

- 可纠正错误：在读取操作过程中，闪存内部 ECC（错误纠正码）检测到错误并进行纠正。
- 读取错误：读取操作遇到（非 ECC）错误，但重试后成功。
- 写入错误：写入操作遇到错误，但重试后操作成功。
- 擦除错误：块上的擦除操作失败产生的错误。

非透明错误有以下几种类型：

- 不可纠正错误：读取操作遇到的错误位数多于 ECC 可以纠正的位数。
- 最终读取错误：读取操作遇到错误，重试后错误仍然存在。
- 最终写入错误：写入操作遇到错误，重试后错误仍然存在。
- 元错误：访问驱动器内部元数据时发生错误。
- 超时错误：3 秒后无响应产生超时错误。

不可纠正错误包括由于垃圾收集而在用户启动的操作或内部操作期间检测到的错误，而最终读取错误仅包括在用户操作期间遇到的错误。

值得注意的是，错误的严重性会根据其可能造成的影响有所不同。除去透明和非透明错误的区别外，非透明错误内部的严重程度也会有所不同。因为这些错误中的一部分（例如最终读取错误，不可纠正错误，元错误）会导致数据丢失，除非系统中存在较高级别的冗余，

因为驱动器无法传送先前存储的数据。本文只考虑至少投入生产 4 年的闪存（或者使用超过 3 年的 eMLC 驱动器，因为它们出现较晚），以及他们在实际运行中所产生的任何错误。表 2.1 展示了每种错误在不同闪存型号的出现频率（表格的上半部分）和每种错误在不同闪存型号的每日出现频率（表格的下半部分）。

表 2.1 各种错误在不同闪存型号上的产生频率

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

2.1 非透明错误

从数据中可以发现，最常见的非透明错误是最终读取错误，即重试操作后无法解决的读取错误。

根据型号的不同，20-63%的闪存驱动器至少会遇到一次这样的错误，而在1000个工作日内至少有2-6个会受到影响。此外，我们还发现最终读取错误的数量和不可纠正错误的数量是密切相关的，并且认为这些最终读取错误几乎完全是由于比ECC的可纠正位损坏所导致。对于所有的闪存型号，最终读取错误比其他不透明类型错误出现更频繁两个数量级。

而与读取错误相比，写入错误很少会变成非透明错误。不同闪存型号中，大约只有1.5-2.5%的闪存驱动器和10,000工作日中的1-4次会出现最终写入错误。最终读取和最终写入错误的频率差异，很可能是写入失败后在其他位置上的重试所导致。所以读取错误可能是由于页面上少数不可靠的单元被读取而引起的，但是最终写入错误往往意味着发生了较大规模的硬件问题。

元错误的发生频率与写入错误相当，但又是以比最终读取错误频率低得多。考虑到驱动器比真实数据包含的元数据少得多，这降低了访问元数据时遇到错误的机会，这可能并不令人惊讶。其他非透明错误（超时错误和响应错误）很少见，通常只有小于1%的驱动器会发生，或者说100,000工作日中会发生小于1个其他非透明错误。

2.2 透明错误

从数据中可以发现可纠正错误是最常见的透明错误类型。这并不奇怪，几乎所有闪存驱动器都至少有一些可纠正错误，大部分（61-90%）工作日都会遇到可纠正错误。本文将在下一章中更详细地讨论可纠正错误，包括原始误码率（RBER）的研究。

除此之外，最常见的透明错误类型是写入错误和擦除错误。它们通常影响 6-10% 的驱动器，但对于一些型号多达 40-68%。一般在 1 万个工作日内不到 5 天就会遇到这些错误。我们把写入和擦除错误，作为闪存块出错的一个标志，这是我们在第 5 章将要深入研究的一种错误类型。

而读取操作期间产生的错误很少是透明的，可能是由于超出 ECC 可以纠正的损坏所致，这是一个通过重试无法解决的问题。透明的读取错误，即可通过重试恢复的读取错误出现频率不到 2%，在 10 万个工作日内只影响不到 2-8 个。

总而言之，除了可纠正错误会大量产生外，与非透明错误相比，透明错误是比较罕见的。最常见的非透明错误类型是不可纠正错误，这会影响 1000 个工作日中的 2-6 天。

3 原始误码率（RBER）研究

评估闪存可靠性的度量标准是原始误码率（RBER），它定义为单次读取中（包括可纠正以及不可纠正损坏位）的损坏位数。第二代闪存驱动器（即 eMLC-A 和 eMLC-B）可以精确地计算损坏的位数和读取的位数，从而可以准确地确定 RBER。第一代驱动器仅能给出读取位数的精确计数，以及由 16 个数据块组成的每个页面中的最大损坏位数。因此，我们假设一个在绝对最差的情况下，即所有块都有错误，并且它们都与最差块有相同的错误数量，所以 RBER 有可能会比真实情况高 16 倍。

3.1 RBER 总览

表 3.1Z 展示了不同型号中所有闪存驱动器的 RBER 中位数，以及第 95 和第 99 百分位数。

表 3.1 不同闪存型号的 RBER

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

我们观察到不同闪存型号的 RBER 差异很大，从 5.8e-10 到 3e-08 不等。当考虑第 95 或第 99 百分位的 RBER 时，差异甚至更大。例如，从 SLC-B 型号的闪存的 2.2e-08 到 MLC-D 的 2.7e-05。即使在相同型号中，也存在很大差异：第 99 百分位的闪存 RBER 比相同型号中位数 RBER 要高至少一个数量级。

底层闪存技术的差异可以部分解释不同型号之间的 RBER 差异，MLC 的 RBER 明显要比 SLC 高很多。

而 eMLC 闪存的 RBER 比其他型号闪存明显要大几个数量级。即使考虑到第一代闪存的 RBER 只是一个下限（最坏的情况下可能比真实高出 16 倍），但仍然有一个数量级以上的差别。我们推测颗粒尺寸可能是一个因素，因为两种 eMLC 芯片具有所有型号中最小的光刻尺寸。

除此之外，没有一家供应商的表现是一直超越其他供应商的。在 SLC 和 eMLC 闪存组中，表现最差和最好的闪存可能出自于同一个供应商。

总之，不同的闪存型号之间、同一型号里不同芯片之间的 RBER 差别都很大。本文将进一步研究影响 RBER 的因素。

3.2 影响 RBER 的因素

本节探索了许多因素对 RBER 的影响，其中包括：

- 每个擦除周期（PE）造成的磨损；
- 闪存的物理年龄，即除去 PE 周期后的设备工作月数；
- 工作量，即闪存读取，写入和擦除操作的数量。因为对页面的操作可能会干扰周围的单元；
- 其他错误。

本节用了两种不同的方法来研究每个因素对 RBER 的影响。先是绘制了不同因素对 RBER 的影响曲线图，并对相关系数加以量化。然后，我们还使用 Spearman 相关系数来衡量不同因素对 RBER 的影响，因为它也可以捕获所有的单调非线性关系。

在详细分析各个因素之后，我们给出了一个总结图（图 3.1）。该图显示了在一个给定的运行时间段内观察到的 RBER 与其他因素之间的 Spearman 等级相关系数，其中包括闪存物理年龄，PE 次数的数量，读取、写入或擦除操作的数量，前一时间段的 RBER 和不可纠正错误（UE）的数量。Spearman 相关系数的值范围为 -1（强负相关）到 +1（强正相关）。每组条形图显示了 RBER 和一个特定因素（见 X 轴上的标签）之间的相关系数，每组中的不同条形对应于不同的闪存型号。所有相关系数都在 95% 以上的置信度下。我们观察到除了先前发生的不可纠正错误（UE）因素外，几乎所有因素都与至少一种型号的闪存 RBER 有明确的相关性。

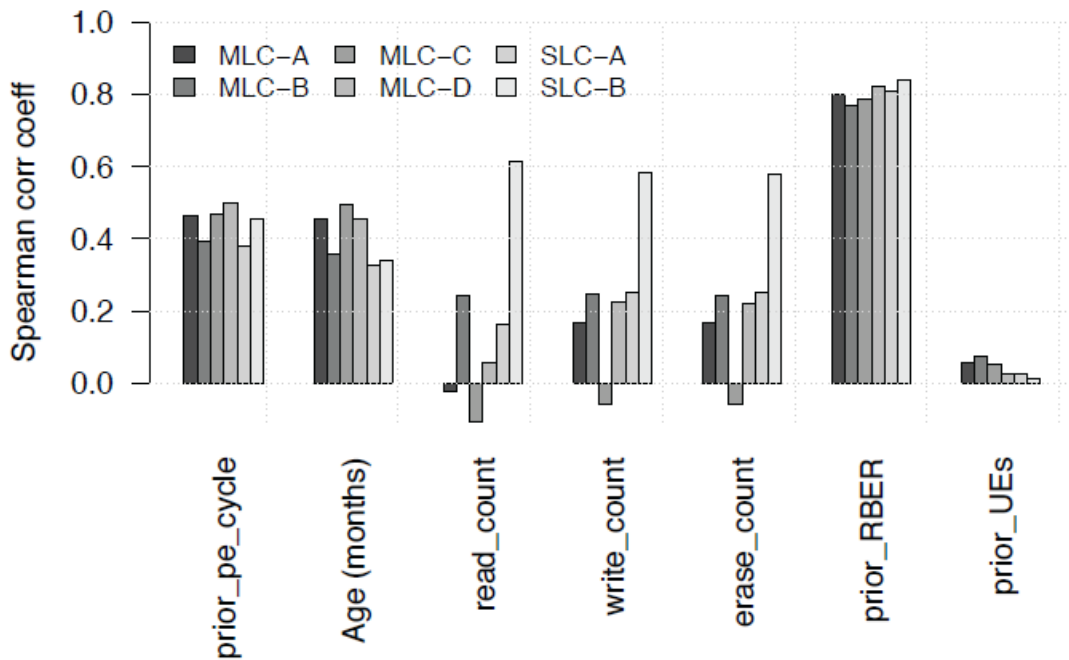


图 3.1 不同闪存型号的 RBER 与多个影响因素之间的 Spearman 系数

3.3 RBER 加速寿命测试

在学术上和工业采购阶段进行的测试一般使用加速寿命测试来预测设备的可靠性。本节将加速寿命测试情况和实际使用的情况做了一个对比。

按照常见的加速方法来测试,我们发现实际使用中的 RBER 明显要高于预期。例如,对 eMLC-A 闪存(在数据收集结束时平均达到 600 个 PE 周期)来说,实际使用的平均 RBER 为 $1e-05$,而在测试中,该 PE 次数范围的 RBER 速率几乎呈数量级递减,直到超过 4,000 个 PE 周期才达到可比较的数量级。这表明,仅根据实验室测试的 RBER 估计来准确预测实际使用中 RBER 可能是非常困难的。

我们也发现某些类型的错误在加速测试中似乎很难产生。例如,对于 MLC-B 型号,近 60%的驱动器在实际使用中会出现不可纠正错误,近 80%的驱动器会出现坏块。然而在加速测试中,六个型号的被测闪存都没有发现任何不可纠正错误或坏块,直到超过其 PE 次数限制的 3 倍以上时才会出现类似错误。而在 eMLC 型号的测试中,超过 80%在实际使用中发生了不可纠正错误,而在加速测试中,没有任何器件在 15,000 个 PE 周期之前就产生不可纠正错误。

我们还研究了以前记录过的 RBER,这些实验是在可控的内部环境下进行的。对于特征参数(25-50nm)类似的接近 PE 循环极限的 SLC 和 MLC 器件,RBER 范

围从 $1\text{e-}08$ 到 $1\text{e-}03$ 不等，大多数闪存具有接近 $1\text{e-}06$ 的 RBER。而在我们的研究中，达到 PE 循环极限的三个闪存芯片只具有 $3\text{e-}08$ 到 $8\text{e-}08$ 的 RBER。即使考虑到我们的数字是下限，在绝对最坏的情况下可能高出真实值 16 倍，我们的数据也明显较低。

总而言之，虽然实际使用中 RBER 明显高于加速寿命测试所预测，但它们还是低于其他文献中类似设备的 RBER（实验室环境下测试）。这表明通过加速寿命试验得到的真实 RBER 预测值是不具备可信度的。

4 不可纠正错误（UE）研究

鉴于我们在第 2 章中发现了不可纠正错误（UE）的高频次出现，我们将在本章中更详细地研究其特性，主要讨论适合于 UE 的度量单位，它们与 RBER 的关系，以及各种因素对 UE 的影响。

4.1 为什么说 UBER 是毫无意义的

过去用于衡量不可纠正错误的标准度量是 UBER，即读取总位数中不可纠正位错误的数量。该度量标准隐性地假设了不可纠正错误的数量在某种程度上与读取的位数有关，因此应该用该数字来标准化。

这个假设对于可纠正错误是有意义的，我们在一个给定时间段中观察到的错误数量与同一时间段中的读取数量有很强的相关性（Spearman 相关系数大于 0.9）。这种强关联的原因在于，一个被损坏的位，只要它能被 ECC 纠正，就会继续增加每次读取访问它的错误计数，因为含有损坏位的单元不会立即被修正（闪存驱动只会周期性地重写带有损坏位的页面）。

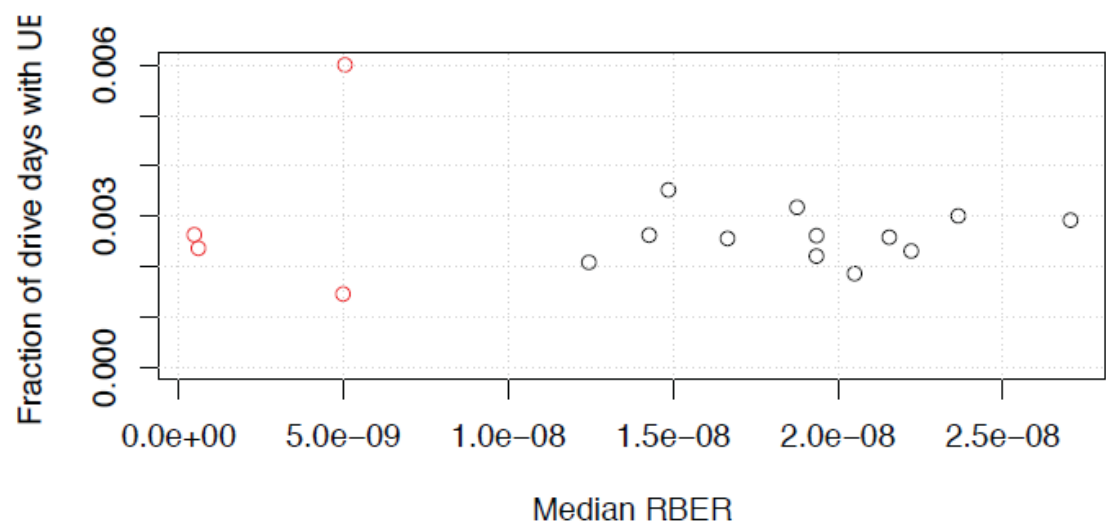
但相同的假设并不适用于不可纠正错误。一个不可纠正错误产生后，闪存驱动将直接删除受影响的块，所以它不会继续增加错误计数。为了验证这个假设，我们使用了各种度量标准来衡量一个给定时间段中的读取次数与同一时间段内不可纠正错误数量之间的关系，包括不同的相关系数（Pearson, Spearman, Kendall）以及肉眼检测。除了不可纠正错误数量之外，我们还研究了不可纠正错误的发生率以及它们与读取操作的相关性。

我们发现没有任何证据可以显示读取次数与不可纠正错误数量之间的相关性。所有闪存型号的相关系数都低于 0.02，当读取操作更多时，并没有更多的 UE 产生。

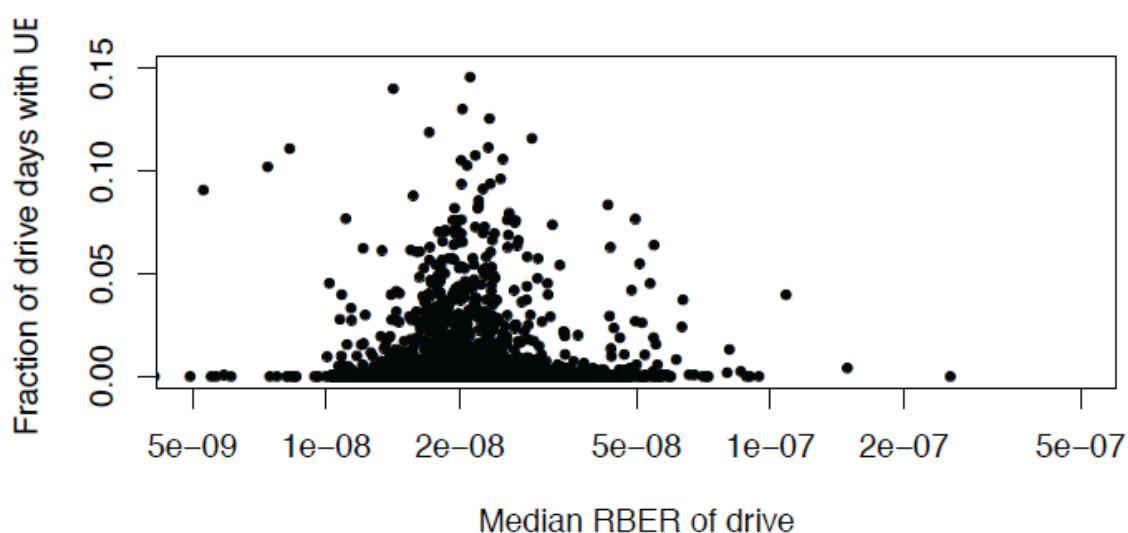
正如我们将在本章第 4 节中看到的那样，写入和擦除操作与不可纠正错误也是不相关的，所以 UBER 的另一个定义（写入或擦除操作的总位数中不可纠正位错误的数量）也不具有意义。

因此，我们可以得出这样的结论：UBER 不是一个有意义的指标，除非是预设读取操作次数的可控环境中。如果在实际使用中拿 UBER 作度量标准，将人为地降低较多读取次数的闪存的真实错误率，同时增加读取次数较少的闪存的真实错误率，因为 UE 发生的次数与读取次数无关。

4.2 和 RBER 的关系



(a)



(b)

图 4.1 RBER 和不可纠正错误 (UE) 之间的关系

RBER 一般用来作为衡量闪存可靠性的标准,尤其是用在预测不可纠正错误产生的可能性上。Mielke 等人曾率先提出使用一个关于 RBER 的函数来计算不可纠正错误的产生概率。此后,许多系统设计者使用了类似的方法,例如根据 RBER 和正在使用的纠错码类型来估计不可纠正错误的产生频率。

本节的目标是描述如何使用 RBER 来预测 UE。从图 4.1 (a) 开始,该图针对多个不同型号的第一代闪存,绘制了它们的 RBER 中位数与一个 UE 产生所需运行时间的关系。同一代中,所有型号闪存都使用相同的 ECC,因此不同型号之间

的差异并不是由于 ECC 的差异造成的，我们无法看到 RBER 和 UE 产生频率之间的相关性。即使换成了第 95 百分位的 RBER，同样看不到相关性。

于是我们从单个闪存芯片的粒度来重复分析，即测试具有较高 RBER 的闪存是否具有更高的 UE 发生率。为此，我们在图 4.1（b）中绘制了每个 MLC-C 闪存的 RBER 中位数与一个 UE 产生所需运行时间的关系（结果与第 95 百分位的 RBER 类似）。我们可以看到 RBER 和 UE 之间依然没有相关性。

最后，我们在更细粒度的时间内进行分析，研究具有更高 RBER 的时间段是否同样是产生更多 UE 的时间段。图 3.1 已经表明，UE 和 RBER 之间的相关系数非常低。我们还尝试了用不同方式来绘制 UE 的产生概率与 RBER 的函数联系，并没有发现的任何相关迹象。

总之，我们得出的结论是：RBER 并不是一个好的 UE 预测指标。这可能意味着导致 RBER 产生的机制并不同于导致 UE 产生的机制。

4.3 和闪存磨损的关系

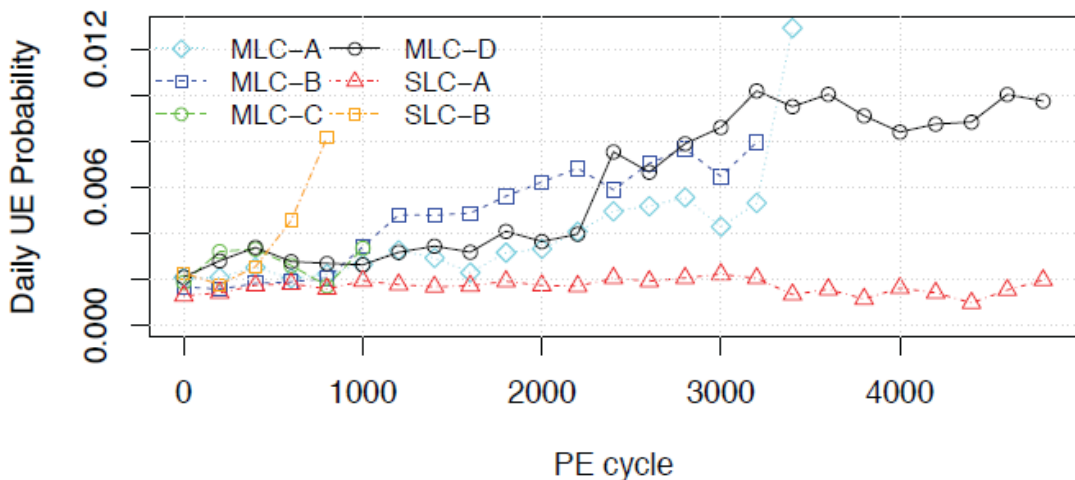


图 4.2 闪存磨损和不可纠正错误（UE）之间的关系

由于磨损是闪存驱动器的主要问题之一，因此，图 4.2 显示了随闪存的 PE 次数而变化的单日产生 UE 的可能性。我们观察到 UE 的产生概率随着 PE 次数的增长而不断增加。然而，与 RBER 的情况一样，增加比通常假设的要慢：肉眼观察和曲线拟合都表明 UE 产生率随着 PE 次数的增加线性增长，而不是按指数规律增长。

另外两个对 RBER 观察结果同样也适用于 UE：首先，在达到 PE 次数极限之后，UE 产生概率没有急剧增加。其次，错误发生率在不同型号的闪存中是不同的，即使它们使用了同一类颗粒。不过差别并不像 RBER 中那么大。

最后我们发现，无论是 MLC 还是 SLC 闪存，对于给定的 PE 次数，具有最低 RBER 的型号不一定同时具有最低的 UE 产生率。例如，对于进行了 3000 次 PE 闪存芯片，MLC-D 的 RBER 比 MLC-B 低 4 倍，但是它的 UE 产生概率略高于 MLC-B。这个结论进一步支持了我们在第 4.2 节中提出的观点。

4.4 和工作量的关系

出于同样的原因，工作负载会影响 RBER，也可能会对 UE 产生影响。例如，由于读取错误确实会附带影响 RBER（3.2 节），读取操作也可能会导致不可纠正错误的增加。

我们详细研究了工作负载对 UE 的影响。但是，如第 4.1 节所述，我们发现 UE 和读取操作的数量之间没有相关性。我们对写入和擦除操作重复相同的分析，并再次看到没有关联。

乍一看，人们可能会认为上述观察结果与之前的观察相矛盾，即不可纠正错误与 PE 次数（预期与写入和擦除操作的次数相关）有关。然而，在我们对 PE 次数的影响的分析中，我们将给定时间段中不可纠正错误的数量与该时间进行的 PE 操作总次数相关联（为了测试闪存磨损的影响）。在研究工作量的影响时，我们考察一下这个特定月份中具有较高读/写/擦除次数的驱动器月份在该特定月份中是否具有不可纠正错误的可能性较高，即，我们不考虑读取的累计计数/写/擦除操作。

我们得出这样的结论：读取干扰错误，写入干扰错误或不完整的擦除操作不是 UE 开发中的主要因素。

4.5 和光刻尺寸的关系

从 3.2 节可知，较小的光刻尺寸往往意味着为更高的 RBER，然而它对产生不可纠正错误的影响并没有 RBER 那样清晰。如图 4.2 所示，虽然 SLC-B 具有较大的光刻尺寸（50 纳米，而 SLC-A 型号为 34 纳米），但 SLC-B 却拥有比 SLC-A 更高的不可纠正错误产生率。小尺寸的 MLC 型号（MLC-B）通常不具有比其他型号更高的不可纠正错误产生率。事实上，在其寿命的前三分之一（0-1,000 个 PE 周期）和最后三分之一（> 2200 个 PE 周期）中，其 UE 产生率比 MLC-D 型低。回想一下，所有 MLC 和 SLC 驱动器都使用相同的 ECC，因此这些影响不能归因于 ECC 的差异。

总体而言，我们发现光刻尺寸对不可纠正错误的影响比预期的小，并且相比对比 RBER 的影响也要小。

5 MLC, eMLC 和 SLC 驱动器的比较

eMLC 和 SLC 驱动器是瞄准企业市场的产品，价格往往较高。除了提供有更高的极限写入量之外，人们往往有一种观点，即企业级固态硬盘是更高端的固态硬盘，一定更可靠和更强大。本章将评估这种看法的准确性。

如果只看表 3.1，该观点是正确的，SLC 驱动器的 RBER 远低于 MLC 和 eMLC 驱动器。然而，表 2.1 显示，SLC 在实际运用中的可靠性表现并没有那么出色，通常它的不透明度错误产生率并不低。

即使在考虑到 MLC 驱动器的 RBER 是下限的情况下，eMLC 驱动器也会比 MLC 驱动器具有更高的 RBER，并且在最坏的情况下可能会高出 16 倍。然而，这些差异可能是由于它们的光刻尺寸较小，而不是技术上的其他差异。

于是我们得出以下结论，即 SLC 驱动器在实际运行中并不比 MLC 驱动器更可靠。

6 与硬盘驱动器比较

闪存的可靠性与其主要竞争对手硬盘（HDD）相比有一个明显的问题。我们发现，在更换率方面，闪存驱动器更胜一筹。硬盘驱动器每年的替换率为 2-9%，这比固态盘的 4 年替换率 4-10% 要高。但是，闪存驱动器在错误率方面的吸引力不大。超过 20% 的闪存在四年内会产生不可纠正错误，30-80% 的会有坏块产生，其中 2-7% 的会发生芯片损坏。相比之下，有关硬盘驱动器的研究报告指出，在 32 个月的时间内，只有 3.5% 的磁盘出现坏道，考虑到硬盘上的扇区数目是数量级大于固态驱动器上的块或芯片的数量，并且这些扇区比块小，所以故障不那么严重。

总之，在本文的研究中，闪存驱动器比硬盘驱动器具有更低的替代率（在其额定寿命内）。不利的一面是，他们拥有比硬盘驱动器更高的不可纠正错误产生率。

7 总结

本文提供了一些关于闪存可靠性的有趣见解。其中一些符合常见的假设和期望，但还有许多是意想不到的。下面将集中总结这些结论：

1. 20-63%的闪存在实际使用的前四年内至少会遇到一次不可纠正错误，并且不可纠正错误也是闪存中最常见的非透明错误。闪存驱动器 1,000 个工作日中有 2-6 天会受其影响。
2. 大多数运行时间内至少会遇到一个可纠正错误，但是其他类型的透明错误与非透明错误相比较是极为罕见的。
3. 我们发现作为闪存可靠性度量标准的 RBER（原始误码率）并不能很好地预测那些实际中主要关心的故障模式。特别是，较高的 RBER 并不意味着不可纠正错误的发生率较高。
4. 我们发现 UBER（不可纠正的误码率）对于衡量不可纠正错误而言，并不是很有意义。我们看不到 UE 与读取次数之间的相关性，因此通过读取的位数来计算不可纠正错误位数，将会人为地夸大读取错误率。
5. RBER 和不可纠正错误的数量随着 PE 次数而增长，但增长率比通常预期的要慢，遵循线性而非指数的速度。即使使用超过了供应商的 PE 次数限制，也不会出现突然的增长高峰。
6. 面向企业市场并被认为是高端产品的 SLC 驱动器不如低端 MLC 驱动器更可靠。
7. 具有较小光刻尺寸的芯片往往有更高的 RBER，但不一定是非透明错误发生率最高的错误，例如不可纠正错误。
8. 虽然闪存驱动器的硬盘驱动器的真实更换率低于硬盘驱动器，但出现错误的概率明显高于磁盘驱动器。

参考文献

- [1] Schroeder, Bianca, Raghav Lagisetty, and Arif Merchant. "Flash Reliability in Production: The Expected and the Unexpected." FAST. 2016.
- [2] 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.
- [3] 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.
- [4] 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
- [5] 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.