|  |  |
| --- | --- |
| **分 数：** |  |
| **评卷人：** |  |

****

**研究生（数据中心技术）课程论文（报告）**

**题 目：SSD可靠性**

**学 号 M202073173**

**姓 名 任哲旋**

**专 业 电子信息**

**课程指导教师 施展 童薇**

**院（系、所） 武汉光电国家研究中心**

**2020年 12月15日**

**目录**

[SSD可靠性综述 I](#_Toc59561861)

[A Survey on SSD Reliability I](#_Toc59561862)

[1. 研究背景 1](#_Toc59561863)

[1.1 SSD介绍 1](#_Toc59561864)

[1.2 SSD结构 1](#_Toc59561865)

[2. SSD可靠性 2](#_Toc59561866)

[2.1 闪存特性 2](#_Toc59561867)

[2.2 错误来源 3](#_Toc59561868)

[2.2.1 擦写次数过多（磨损） 3](#_Toc59561869)

[2.2.2 Data Retention(数据保留损失) 3](#_Toc59561870)

[2.2.3 读干扰 3](#_Toc59561871)

[2.2.4 存储单元之间干扰 3](#_Toc59561872)

[2.2.5 写错误 4](#_Toc59561873)

[2.3 主要解决方式 4](#_Toc59561874)

[2.3.1 ECC（Error correction code）纠错 4](#_Toc59561875)

[2.3.2 数据冗余（Redundancy /RAID） 4](#_Toc59561876)

[2.3.3 Re-read（重读） 4](#_Toc59561877)

[2.3.4 Data Scrub/ Relocation （重写） 4](#_Toc59561878)

[2.3.5 数据随机化 4](#_Toc59561879)

[3. 相关研究 4](#_Toc59561880)

[3.1 大规模企业存储部署中SSD可靠性研究 5](#_Toc59561881)

[3.1.1 介绍 5](#_Toc59561882)

[3.1.2 影响ARR的因素 5](#_Toc59561883)

[3.1.3 硬盘之间的联系 6](#_Toc59561884)

[3.1.4 总结 7](#_Toc59561885)

[3.2 SSD可靠性的设计权衡 7](#_Toc59561886)

[3.2.1 介绍 7](#_Toc59561887)

[3.2.2 观察 7](#_Toc59561888)

[3.2.3 整体可靠性管理方法（HRM） 8](#_Toc59561889)

[3.2.4 实验与评估 8](#_Toc59561890)

[3.3 通过利用自我恢复和温度意识来提高3D NAND闪存设备的可靠性 9](#_Toc59561891)

[3.3.1 介绍 9](#_Toc59561892)

[3.3.2研究背景 9](#_Toc59561893)

[3.3.3 观察结论 10](#_Toc59561894)

[3.3.4 自恢复模型 11](#_Toc59561895)

[3.3.5 实验与评估 11](#_Toc59561896)

[4. 总结 12](#_Toc59561897)

[5. 参考文献 12](#_Toc59561898)

# 

# SSD可靠性综述

任哲旋1)

1)(华中科技大学武汉光电国家研究中心 湖北 武汉 430074)

摘 要 随着互联网的发展，企业和数据中心产生的数据量呈爆炸式增长。因此为了提高存储设备性能、提高用户体验，许多企业和数据中心开始广泛应用了固态硬盘（SSD）来替换之前使用的机械硬盘（HDD），这主要是因为SSD具有多方面的优势。但是由于NAND闪存具有一定的特性，并且随着制作工艺不断提高，晶体管密度不断增大，导致近年来误码率不断提高，因此解决SSD可靠性问题变得尤为重要。之后本文针对近年来SSD可靠性的研究列出了三篇相关文献。首先介绍了针对企业app应用中SSD的可靠性相关研究，分析了特定场景下可靠性问题来源以及多种影响因素。目前对SSD可靠性问题的来源仍然不太清晰，需要深入研究，并且希望通过非算法的配置手段来降低SSD更换率。之后是对多种不同可靠性保证算法进行权衡，针对冷热数据来对应选择适合的算法，从而结合形成一个整体可靠性管理方法。与之前的做法相比可保证SSD处于原有的可靠性强度，同时优化了处理过程、提高了性能。最后是通过利用NAND闪存的自恢复物理现象，根据当前的工作温度来调整重读的电压，实现了一个HeatWatch机制。这样不仅可以降低误码率，提高可靠性，同时SSD的整体寿命也有所提高。

**关键词** SSD；固态硬盘；NAND；可靠性；自恢复；

# A Survey on SSD Reliability

Ren ZheXuan1)

1)( Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, Hubei 430074)

**Abstract** With the development of the Internet, the amount of data generated by enterprises and data centers has exploded. Therefore, in order to improve storage device performance and user experience, many companies and data centers have begun to widely use solid state drives (SSD) to replace previously used HDD, mainly because SSD has many advantages. However, because NAND flash memory has certain characteristics, and as the manufacturing process continues to improve, the transistor density continues to increase, resulting in the continuous increase of bit error rate in recent years, so it is particularly important to solve the reliability problem of SSD. After that, this article lists three related documents for the research on SSD reliability in recent years. First, it introduces the related research on the reliability of SSD in enterprise app applications, and analyzes the source of reliability problems and various influencing factors in specific scenarios. At present, the source of the SSD reliability problem is still not clear and requires in-depth research, and it is hoped that non-algorithmic configuration methods can be used to reduce the SSD replacement rate. After that, a variety of different reliability assurance algorithms are weighed, and suitable algorithms are selected corresponding to the cold and hot data, thereby forming an overall reliability management method. Compared with the previous method, it can ensure that the SSD is at the original reliability strength, while optimizing the processing process and improving the performance. Finally, by using the self-recovery physical phenomenon of NAND flash memory to adjust the reread voltage according to the current operating temperature, a HeatWatch mechanism is implemented. This not only reduces the bit error rate and improves reliability, but also improves the overall lifespan of the SSD.

**Key words** SSD; Solid State Drive; NAND; Reliability; Self-recovery;

# 1. 研究背景

## 1.1 SSD介绍

近年来，随着互联网的不断发展，企业和数据中心每日产生的数据量呈爆炸式的增长，但是单位时间内服务器处理数据的数量是有限的，而其中存储设备的IO响应成为了阻碍其响应速度的一个很大的问题，因此为了提高用户体验，也就迫切需要提高存储设备的性能。因此许多企业和数据中心都开始广泛应用了固态硬盘（SSD），来替换之前使用的机械硬盘（HDD）。

固态硬盘以其高速的读写速度而被人们广泛关注，与机械硬盘相比，其性能可提高几倍到数百倍，其中随机读写性能（速度和时延）差异最为明显。当然除了性能以外，SSD还有许多其他的优势。（1）首先是SSD的功耗较低，由于SSD极高的性能，相对于HDD而言，相当于单位功耗产生出了百倍的性能。此外当处于空闲或省电模式下，SSD会启动内部休眠，尽可能多地关掉不工作的硬件模块，只需使用极低的功耗来保持运行状态。所以SSD被称为高性能、低功耗的节能产品，符合数据中心（Data Center）的使用定位。（2）抗震防摔。SSD内部不存在任何机械部件，相比HDD更加抗震。HDD是机械式结构，磁头和磁片之间发生跌落时接触碰撞会产生物理损坏，无法复原。SSD是电子和PCB结构，PCB加半导体芯片，跌落时不存在机械损伤问题，因此更加抗震和可靠。无噪声，

由于结构上没有马达的高速运转，SSD是静音的。（3）体积小多尺寸。HDD一般只有3.5寸和2.5寸两种形式，SSD除了这两种，还有更小的可以贴放在主板上的M.2形式，甚至可以小到芯片级，例如BGA SSD的大小只有16mm×30mm，甚至可做得更小。

当然由于目前技术原因，SSD的制造成本比HDD更昂贵，但是近年来由于工艺不断发展，SSD的价格逐渐下降，再考虑其具有多方面的优势，与HDD相比性价比逐渐凸显，也正因此SSD才会越来越被人们欢迎。

## 1.2 SSD结构

如图1.1所示，SSD由主控制器，存储单元，缓存(可选)，接口(SATA，SAS, PCIe等)这几个主要部件构成。对于存储单元，主要分为DRAM，Flash，3D X Point三种，而其中闪存（Flash）类型又细分为NAND和NOR两种,NOR的特点是程序可在晶片内执行，而且传输效率很高，但是写入和擦除速度较低。NAND结构能提供极高的单元密度，并且写入擦除的速度很快。而企业和数据中心使用的主要是NAND-based SSDs，因此本文也主要是去研究此种SSD的可靠性，（见图1.2）。再根据存储单元内的信息位数对NAND类型的SSD细分，又可分为SLC，MLC和TLC等，不同的类型各方面的差异较大，如表1.4。

表1.1HDD与SSD主要区别



表1.2HDD与SSD读写性能对比

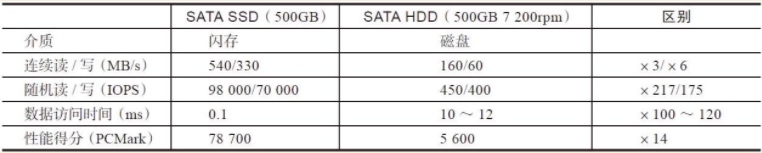


表1.3 HDD与SSD功耗对比



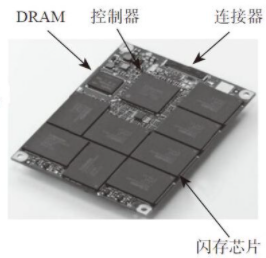


图1.1 SSD内部结构

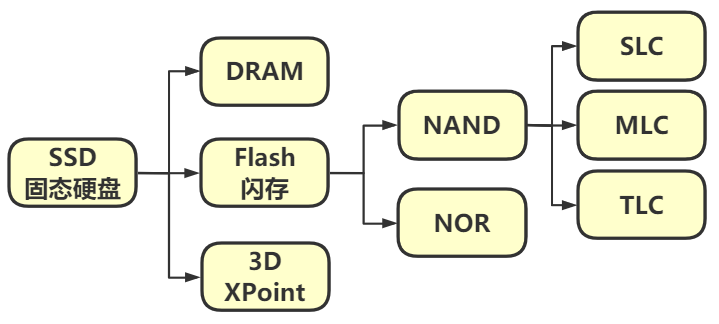


图1.2 SSD主要分类

表1.4 SSD主要分类



其中可将主控制器看作包括多种算法的软件层FTL，负责对SSD数据进行管理。这是因为NAND闪存不能原地更新，而且数据必须先擦除才能重写。数据按页存储，但是必须以块（页的父级单位）为单位擦除。所以为了高效的更新数据，会先将有用的信息挑选出合并写到其他地方，然后对原块全部擦除。所以内部包含了垃圾回收（GC）算法，负责更新数据，并且擦除掉无用数据，如图1.3。而NAND闪存的每个存储单元的擦写次数是有一定寿命的，如果擦写次数过多会导致存储单元磨损甚至损坏，显著降低可靠性，因此需要添加磨损平衡算法，尽可能的分散存储数据，使得各个存储单元的磨损程度尽可能一致，保证数据最大写入量，而不总是对某一部分存储单元频繁擦写，导致提前损坏。而为了实现更好的兼容性，SSD内部维护了一张表，负责将逻辑数据映射并存储到闪存空间上，这样可以像HDD一样来操作SSD。坏块管理，主要是为了屏蔽掉由于频繁擦写或者其他原因导致的坏块，保证数据的可靠性。当然还有其他的多种可靠性保证算法，本文会在之后给出更具体的描述。

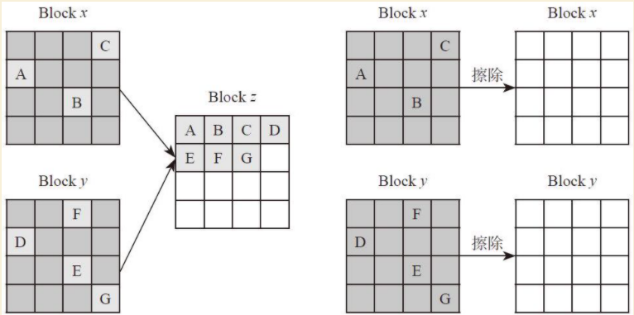


图1.3 垃圾回收算法示例

# 2. SSD可靠性

## 2.1 闪存特性

NAND闪存的主要结构是浮栅晶体管，如图2.1。在源极（Source）和漏极（Drain）之间电流单向传导的半导体上形成存储电子的浮栅，浮栅上下被绝缘层包围，存储在里面的电子不会因为掉电而消失，所以闪存是非易失性存储器。对于写操作是在控制极加正电压，使电子通过绝缘层进入浮栅极，这样在浮栅层内部会具有一定的电压，当之后读取时可根据电压的值来表示具体存储的数据。而擦除操作正好相反，是在衬底加正电压，把电子从浮栅极中吸出来，如图2.2。

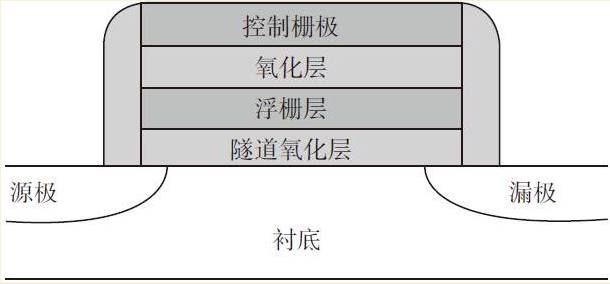


图2.1 NAND存储单元物理结构

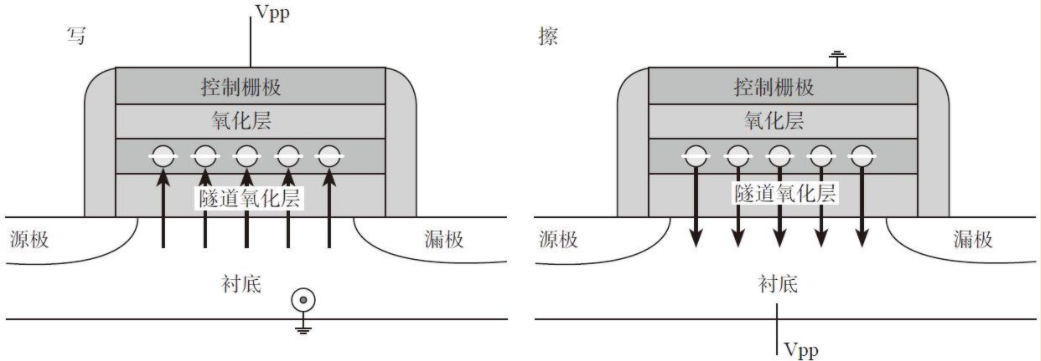


图2.2 NAND存储单元擦（右）写（左）示例

NAND闪存分为SLC，MLC，TLC多种，主要区别是内部表示数据的位数，分别为1位，2位和3位二进制数据。如下图2.3为MLC阈值电压与数据表示的示例，表示表示所有两位数值的四种不同情况（即00、01、10、11）。因为写入数据时，会拉入不同数量的电子，则每个存储单元内部相应的会具有不同的电压。而单元内所有可能的阈值电压的范围分为四个电压窗口或状态，每个不同状态之间具有一定间隔，很容易分辨出。因此可根据这些间隔制定多个参考电压，当写入后读数据（即读取电压的值）时，通过多次比较当前电压和参考电压的大小来确定存储的具体数据时什么。

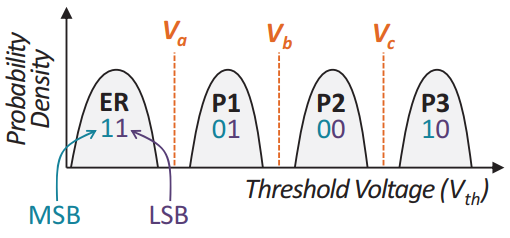


图2.3 MLC阈值电压示例

而图2.3中可以看到，LSB（Least significant bit）表示可存储数据的最低有效位，MSB（Most significant bit）对应表示数据的最高有效位。区分高低为的主要原因是对于MLC或者TLC 2-pass等类型的闪存，会分为Lower Page和 Upper Page。Lower Page的操作时间较短，称为“快页”，Upper Page操作时间较长，称为“快页”。SSD厂家为了使得存储速度更加高效，提高擦写寿命，解决掉电安全隐患，会把MSB和LSB一个存储单元上的两个bit分别映射到两个不同的page（页）上，那么，LSB对应的page就是Lower page，MSB对应的page就是Upper page，而具体如何映射，每个厂商各有不同。其中“慢页”需要根据自身的阈值电压和Lower Page阈值电压分布情况来得到最终电压。

但是随着闪存的使用，存储单元由于多次擦写数据后会磨碎氧化层，这会导致写数据之后，进入浮栅的电子数量会减少，电压分布的状态间距会减少，并且产生重叠，导致无法区分内部存储的数据。而如果继续进行擦写可能会出现坏块，造成该物理区域的永久性损坏。而如果数据存储时间变长，存储在闪存里面的电子也会出现逃逸，被氧化层错误捕获，从而发生比特翻转。

而随着单个存储单元内表示的数据位数不断提高，以及SSD制作工艺不断提高，晶体管密度不断增大，导致近年来误码率不断提高，因此解决SSD可靠性问题变得尤为重要，使用闪存作为存储介质的固态硬盘需要采用一些数据完整性的技术（可靠性保证算法）来确保用户数据更可靠且不易丢失。

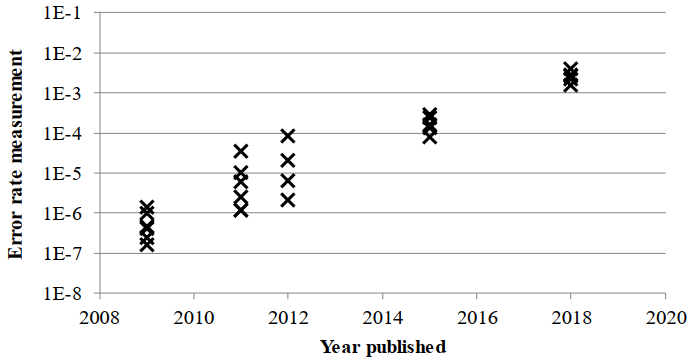


图2.4 误码率逐年提高

## 2.2 错误来源

### 2.2.1 擦写次数过多（磨损）

闪存块擦写次数增多，氧化层逐渐老化，电子进出存储单元越来越容易，因此存储在存储单元的电荷容易发生异常，发生bit翻转，导致数据读错误。

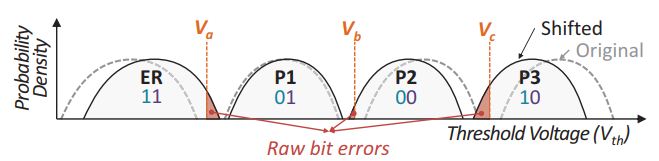


图2.5 阈值电压偏移导致读错误

### 2.2.2 Data Retention(数据保留损失)

随着存储单元存储数据的时间不断增加，会导致在其空闲时发生电荷泄漏而引起的错误。电荷既可以通过隧道氧化物层泄漏到衬底，对于3d-NAND也可以通过电荷捕获层泄漏到相邻的单元，因为隧道氧化物和电荷捕获层都是不完善的绝缘体。由于该泄漏，闪存单元的阈值电压随时间降低，从而将高压状态的阈值电压分布移动到较低的电压。在较长的时间段内，每个分布的左尾部的增加部分在读取的参考电压上移动，从而导致更多的原始位错误。

### 2.2.3 读干扰

读一个Wordline数据时，需要施加Vpass 电压在其他Wordline上，导致其他闪存页发生轻微写。如果读的次数过多，轻微写累积起来就会使阈值电压分布发生右移，导致读数据时候发生误判，即读数据错误。

### 2.2.4 存储单元之间干扰

由于存储电子的浮栅极是导体，两个导体之间构成电容，一个存储单元电荷的变化会导致其他存储单元电荷变化（影响最大的是相邻的存储单元）。周围的单元不同状态时，中心单元的阈值电压不一样。

### 2.2.5 写错误

上一节中已提到对于MLC或者TLC 2-pass的存储单元的bit 会映射到不同page（LSB: Lower Page，MSB: Upper Page）。当写数据时会先写LP（Lower Page），再写UP（Upper Page）写。写UP是基于之前LP的状态，然后再写每个存储单元到目标状态。若写UP时，LP数据已出错（写UP时，LP不会经过ECC纠错，写过程发生在闪存内部），就会导致存储单元写到一个不期望的状态，即发生写错误。且如果一开始擦除状态不对，还是会发生写错误。

而且在NAND闪存中，同一页中的单元被一起写入，在每个写入的脉冲期间都会注入随机量的电荷，电流震荡等许多因素都会增加写入的随机性，从而导致写错误。

## 2.3 主要解决方式

### 2.3.1 ECC（Error correction code）纠错

ECC（纠错码），是常用的一个解决方式，根据需求选择对应的纠错强度，并通过存储额外的校验信息可纠正n位的错误。SSD控制器上面都有ECC纠错模块，有些闪存内部也集成了ECC纠错模块。而因为ECC是通过写入额外的校验数据作为后续纠错的依据的，因此分配越多的预留空间就能提供越强的ECC纠错能力。

目前大多数SSD使用的是BCH算法，但是也有的使用的是LDPC算法，与比BCH相比，LDPC的纠错能力提升了近3倍。这主要是因为LDPC的算法更加复杂，计算量更大，因此随着错误率的提升，延迟也会不断提高。

ECC分为硬解码和软解码两类，硬解码是只需要计算一次来直接解码，可配合重读来提高纠错能力。软解码的纠错能力更强，但延迟也会更高，LDPC软解码可根据多次重读的数据结合成的对数似然比进行判决。

### 2.3.2 数据冗余（Redundancy /RAID）

ECC是将数据和校验信息存在同一位置的，但是如果出现整块硬件区域损坏，数据是无法纠正甚至恢复的。而数据内部冗余的方式是跨多个闪存芯片存储额外信息，可避免硬件芯片、Word Line等错误。数据冗余可认为是RAID技术，目前SSD普遍使用的是RAID5。但是由于SSD只能异地更新，并且涉及到GC等特性，会使得此方法的实现变得很困难。而且冗余数据也伴随着更大的读写流量，进一步增加延迟。

### 2.3.3 Re-read（重读）

对于电压分布平移的问题仍可恢复，因为表示数据的不同阈值电压之间还是有明显间隔的，只是读取的电压不处于间隔内。因此可不断改变读取的参考电压来尝试找到可读出数据的电压点。直到正确读出数据，理论上，只要状态电压分布没有发生重叠，就可以通过重读恢复数据。还有更复杂的重读，叫作Advanced Read Retry。先读附近的单元确定状态，再用不同参考电压读两次要读的单元，根据附近单元数据决定选择哪一个。



图2.6 重读/多次调整读取的阈值电压

### 2.3.4 Data Scrub/ Relocation （重写）

因为当闪存存储数据的时间较长后会丢失电子，所以可在未出现严重错误之前提前将该块数据擦除，并写到其他地方，防止之后出现不可纠正错误。如已经丢失电子，但仍可通过其他机制恢复例如可在ECC或调整电压进行重读成功纠错后重写。甚至是未丢失电子而只是为了防范。重写后可恢复到正常电压，相当于定时刷新。但是过多的重写也会增加额外的擦写次数，加重磨损。而且需要在适当的实际去重写，以防增加用户请求的延迟。

### 2.3.5 数据随机化

闪存是通过控制栅施加电压来存储数据的，对某些写入的数据样式很敏感，不断地输入全0或者全1，很容易导致闪存内部电量不均衡，从而造成信号抗干扰性下降，导致这些数据在闪存中可靠性变差。而且当0，1分布不均匀时，表示数据的对应电压也会相差的很大，而且直接相邻的存储单元之间很容易互相影响，从而造成电压变化。因此可考虑在写入数据之前添加一个转换，让0，1的分布均匀，充分隔离，各存储状态电压分布明显不产生交集并且降低相邻单元之间的耦合电压产生的影响。

SSD控制器或者闪存内部都有数据随机化模块，它对用户写入的数据加入扰码，使最终写入闪存数据0和1基本保持均衡， 一般会推荐使用AES加密算法实现数据随机化。

# 3. 相关研究

本节主要是对三篇关于NAND-based SSDs可靠性的相关文献[1][2][3]进行简单讲述与介绍，并列出文献中作者观测得到的结论以及实验后的研究成果。

## 3.1 大规模企业存储部署中SSD可靠性研究

### 3.1.1 介绍

本论文出自FAST 20[1]，主要是对大规模企业存储系统中 NAND-based SSDs 的现场研究。之前有许多论文分析了互联网或者公有云厂商中SSD盘的可靠性相关的问题，但本文重点研究企业级存储领域中的SSD可靠性，并且从多个角度分析了造成SSD损坏更换的原因。

对于SSD来说，会出现多种错误，出现bit翻转是可通过相关手段恢复的，对于不可纠正错误概率较低，而且即使发生，SSD仍然可用，只需下次避免即可，但是整盘更换的这种错误则是较为严重的，被更换的SSD大概率不能再使用，而这会增加企业的成本，而本论文主要是去考虑造成SSD更换的错误原因。其中衡量更换率的指标为ARR（平均年替换率）。

本文研究的数据来自NetApp企业的140万个SSD的运行数据，覆盖了3个不同制造商，18个不同型号，12种不同容量，4种闪存技术（SLC，cMLC，eMLC，3D-TLC），多个固件版本等众多市面上不同类型的SSD，如表3.1。首先根据前期对数据的统计与观察，作者得出了几个比较明显的结论：

（1）平均年替换率ARR为0.22%（范围：0.07%-1.2%）.，这比google数据中心ARR低（1.0%-2.5%）并且远比HDD故障率低（2%-9%）。

（2）即使对于技术上非常相似的驱动器型号规格，ARR有很大差异。例如0.53% for II-G 15TB drives。1.13% for II-C 15.3TB drives。

（3）为坏块而预留的备用空间利用率很低，平均使用率低于15%，99.9%的备用空间利用率为17%。

（4）SSD远未达到其PE极限，所以因达到最大擦写次数而失效可能性为零。统计发现只适用2-3年的SSD实际擦写利用率仅为2%。本次研究的99.9%的SSD使用寿命为15% - 33%。

之后作者又将更换原因类型分为4类进行观测，如表3.2。其中A类错误最严重，且占比最大约为 ，而程度最轻的D类错误主要是系统检测到某些指标超过警戒线，然后随机给出更换提示，是一种预测型的更换原因，硬盘仍可使用只是为了在出现更严重错误之前及时替换下来，其占比也约为 。

### 3.1.2 影响ARR的因素

之后作者开始从多个方面来分析造成SSD更换的因素。这部分统计都基于他们生产环境中的eMLC和3D-TLC SSD两类SSD得出，且大部分eMLC和3D-TLC大约有10k次的擦写次数上限，小部分eMLC的擦写次数上限为30K。

（1）**使用率和使用时间**

擦写过多无疑会造成更严重影响。但是作者发现，反而擦写使用率越少替换率却更高，如图3.1。经过最终统计可知在刚开始使用的1年内ARR是一年后的2-3倍，ARR并没有随使用时间增加，即擦写次数远未达到上限。

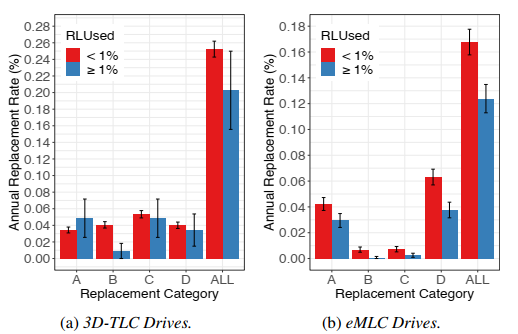
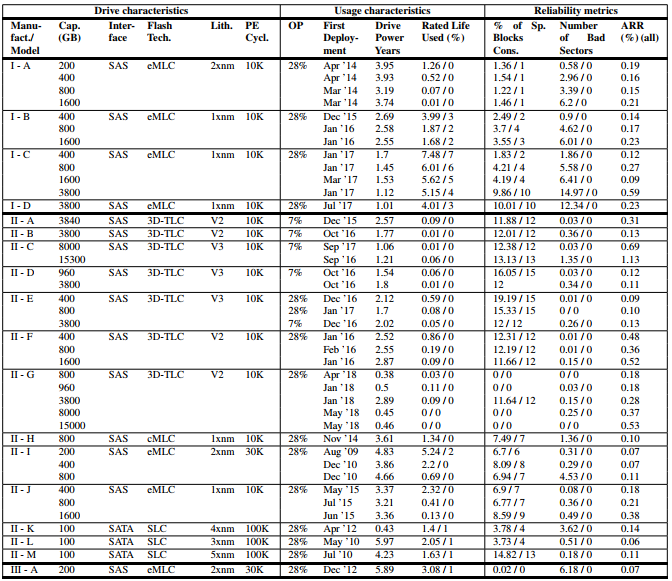
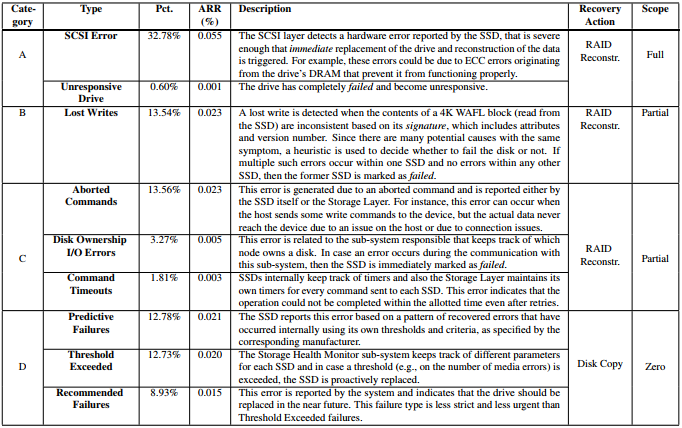


图3.1 使用率与ARR的关系

表3.1 企业SSD数据统计摘要 表3.2 企业SSD更换原因类型

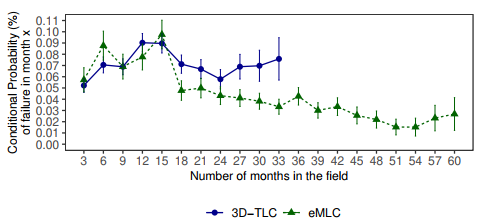


图3.2 使用时间与故障率的关系

（2）**Flash类型**

SLC，MLC，TLC这三者内部存储的数据位数是不断增加的，且擦写次数不断减少，因此理论上TLC会更易出错，有更高ARR，但是由图3.1和图3.2可以看出，闪存类型与使用率和使用时间相比，影响较小可直接忽略。

（3）**容量**

由图3.3可已看出，更高容量具有更高ARR，而且会出现更多的A类严重错误，以及更少的D类预测性错误。

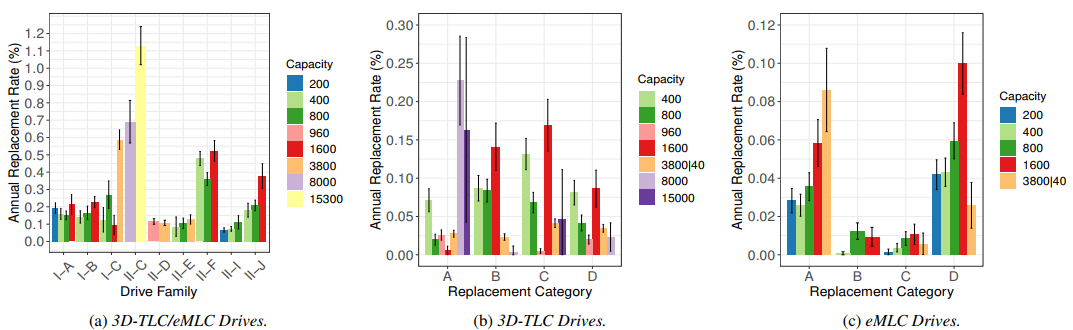


图3.3 容量与ARR的关系

（4）**工艺/密度**

之前的研究已发现NAND存储单元密度与误码率（RBER）有很强的关系。对数据中心SSD盘的研究表明，密度更高的SSD误码率更高。但是对于eMLC来说密度提高，ARR也提高，如图3.4(b),而对于3D-TLC结果却相反，如图3.4(a)，因此可知密度的提高不一定有更高更换率ARR。

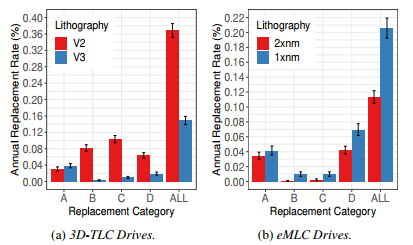


图3.4 密度与ARR的关系

（5）**固件的版本**

同参数下，老版本的固件ARR更大，作者认为可能是因为之后的版本修复了bug，但是新版本的不稳定也会增加ARR。

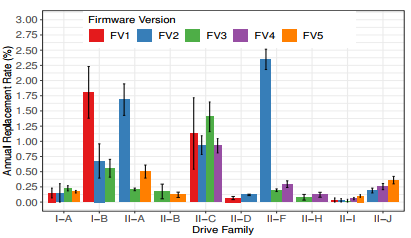


图3.5 固件版本与ARR的关系

（6）**用途**

SSD作为缓存或存储对ARR无影响。

（7）**用作RAID中的data或校验块**

用作RAID中的data或校验块对ARR无影响。

（8）**额外分配空间（**Over-Provisioning）

额外分配空间的大小对ARR无影响，但额外分配空间的利用率越高，ARR更大。

（9）**坏块个数**

出现过坏块的SSD会有更高ARR，如图3.6。

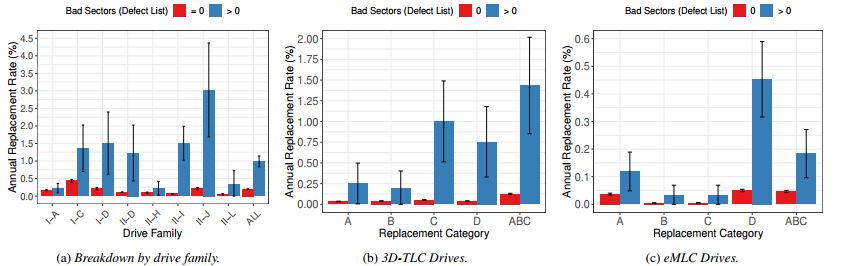


图3.6 固件版本与ARR的关系

### 3.1.3 硬盘之间的联系

上一节均为硬盘自身的问题，而在实际存储中，硬盘之间也有一定联系，如常用的RAID技术。

（1）同一个RAID组下的SSD盘的ARR

一个RAID组中某一个盘被替换，很有可能短时间内另一个盘也会被替换，如图3.7，中位数时间非常小。这可能是因为RAID重建会给其他盘带来额外的负担并暴露潜在的错误，而且同一个RAID组中的盘软硬件环境相同，导致因相同问题出错，如电力波动，过热等。

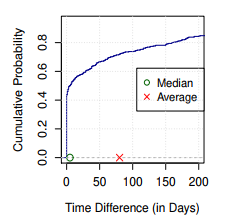


图3.7 同一RAID组中连续两次驱动器更换之间的时间

（2）不同RAID组大小对ARR的影响

更大的RAID组个数，对应的ARR更大，但如图3.8，当增长到一定程度后ARR会出现饱和（并未再增长）。

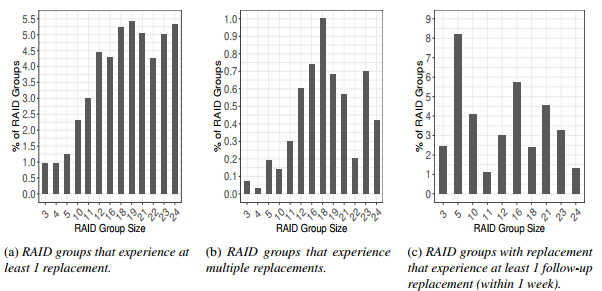


图3.8 RAID组大小

### 3.1.4 总结

（1）比起类型（eMLC,3D-TLC），更应考虑密度和容量的影响因素。

（2）虽然擦写越多故障率越高，但SSD使用初期更容易坏，因此需要更加关注其稳定性。

（3）企业中的SSD擦写次数远未达到上线，这使得未来的QLC广泛应用成为可能。

（4）生产厂商应关注固件稳定性，要着重减少bug。

（5）无直接证据表明更大的RAID组大小有更高ARR。

## 3.2 SSD可靠性的设计权衡

### 3.2.1 介绍

本论文出自FAST 19[2]，研究方向是SSD的整体可靠性管理方式。论文主要是根据几个严重问题，尝试对几种现有的解决方式分析，最后做一定的权衡和取舍，制定一个适合于多种应用场景的整体的可靠和高性能的解决方案，即同时兼顾性能，可靠性和使用寿命等。

作者认为每种方式有一定优势，但是也存在代价，这会影响性能，所以需要根据特定的问题，去针对性的使用更合适的方法。

作者主要考虑三个可靠性错误来源（问题）：Wear（擦写磨损），Retention（数据保留损失），Disturbance（干扰），更具体的细节见“2.2 错误来源”。而作者主要针对三种解决方案进行权衡：ECC（纠错码），Re-read（重读），SSD Redundancy （SSD内部数据冗余），Background Relocation（后台重放置，即重写），更具体的细节见“2.3 主要解决方式”。

本篇文章更多的是关注SSD的综合性能，其中误码率（RBER）是一个可衡量可靠性和响应性能的综合指标。所以作者先根据错误来源对误码率进行建模。如图3.9，作者对擦写次数（Cycle），数据持续存储时间（Time），Reads（读取次数），构建了一个误码率相关公式，其中，，，是系数，而k，m，n，p，q是闪存特有的指数，而每一行都对应到这三种错误原因。之后可以此公式作为不同方法优劣的衡量标准。

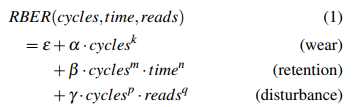
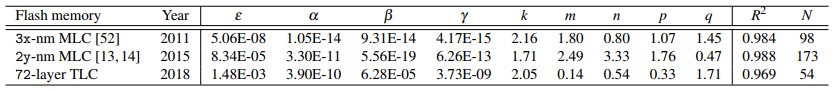


图3.9 RBER建模公式

### 3.2.2 观察

首先作者尝试使用RBER模型分析三种不同类类型闪存的错误原因（3x-nm MLC，2y-nm MLC，72-layer TLC），对应的参数见表3.3。然后分别在不同颗粒类型下进行曲线拟合，表明不同状态下的主要错误原因，如图3.6。可以看到总体错误率随着擦写次数的增加而增加，前期主要是磨损的影响。对于TLC，当擦写次数到一定程度后扰乱错误的影响反而更为严重。

表3.3 RBER模型参数



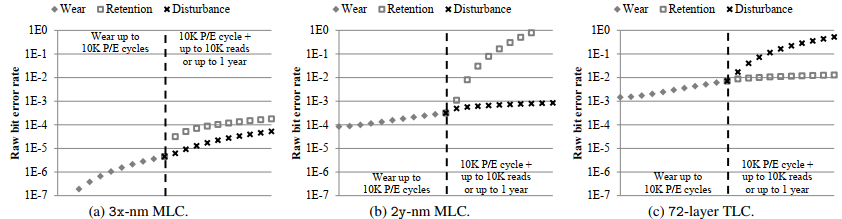
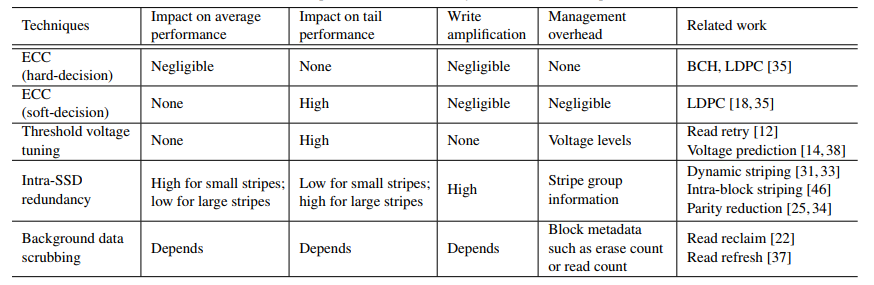


图3.6 3x-nm MLC, 2y-nm MLC, 72-layer TLC

然后作者开始权衡几种可靠性保证算法，列出了这些方法的代价（写放大，尾延迟等）以及主要做法。

表3.4 可靠性解决方案的权衡



之后作者使用DiskSim工具构造了三个SSD，每个固态硬盘具有三种不同的初始磨损状态，每个状态具有不同的错误率，但在配置上相同。而且使用了相同的内部组织来隔离可靠性增强技术对整体性能的影响。对于测试，使用7:3的混合的4KiB读/写的综合工作负载等方式来模拟真实的SSD运行状态。之后通过对比试验来观察不同状态下对应的可靠性保证算法的性能与强度。

（1）**ECC**

ECC具有固定的运行时间，因此其性能是可预测的。但是较弱的ECC进行数据重读会引起额外的读取干扰。而磨损更严重时除了降低性能,还会降低ECC校正强度较弱。

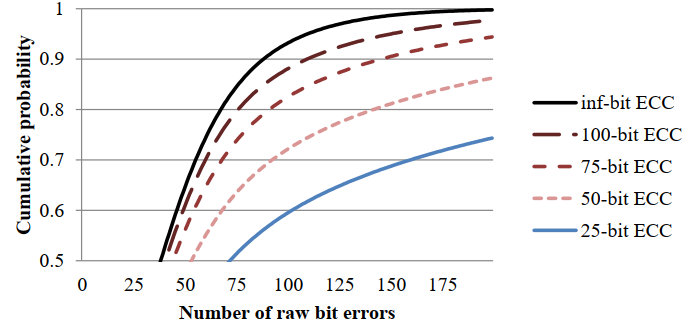


图3.7 TLC的相同占比数据对应的误码数量（弱ECC重读干扰）

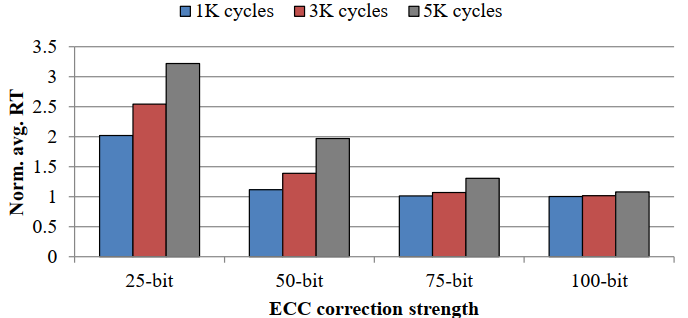


图3.8 TLC的磨损率与ECC强度关系

（2）**重读**

重读比直接ECC纠错更有效，但是可能需要多次重读才能获取到正确数据，因此会加剧延迟，并且其延迟大小是无法预期的。而且之前的研究发现过于频繁的重读反而会产生影响高磨损状态下会严重降低性能

（3）**冗余**

可防止ECC不可恢复的错误以及多种硬件错误。但会加重写放大，而且遇到错误时会加重读放大。磨损程度越高性能可靠性下降越快，同时条带越大会越加剧这一趋势。一般场景下其开销大于其好处。

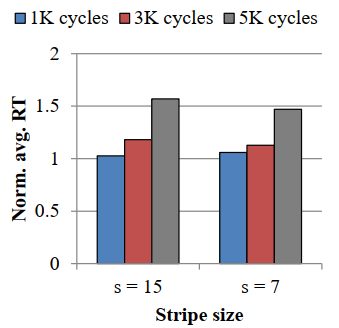


图3.8 TLC的冗余的条带大小与响应时间的关系

（4）**重写**

重写为预警方法，由于可在出错前及时避免错误，因此与其他方法相比之下更强大，但不能解决所有问题，例如本来有错的数据再次重写并不能改正，而且重写会增加延迟与内部流量。

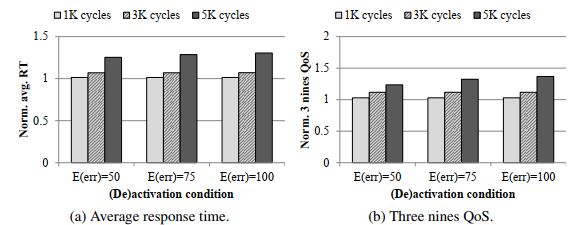


图3.9 TLC的不同磨损程度下重写与响应时间和Qos的关系

### 3.2.3 整体可靠性管理方法（HRM）

最后作者根据前期实验观察的发现与结论，考虑将数据分为冷数据，热读数据和热写数据三类，然后使用更适合于对应类型数据的算法去做可靠性保证。

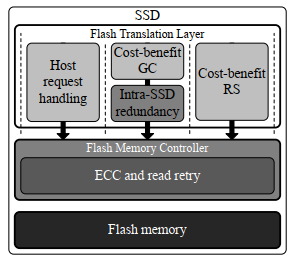


图3.10 整体可靠性管理方法HRM结构

（1）冷数据：不常用的数据，自写入后读取的次数逐渐减少，但可能之后会使用所以并未清除。因此可仅对冷叔使用强大的数据冗余方式来防止数据保留丢失以及多种严重软硬件错误，同时，不对热数据采取冗余方式可减少写放大的影响。

（2）热-读数据：经常需要读取的数据。需要防止数据的重读带来的干扰错误，可通过GC来重新放置，但应权衡GC的效用与重新放置的成本。

（3）热-写数据：写入一次但可能短时间被清楚的数据。因为其特殊性质，会通过SSD的GC被内部其他数据更新，所以无需关注。

对于冷热数据的分类依据，可用SSD现有机制来实现。

（1）热读数据：读取洗涤器（Read Scrubber）收集的数据。

（2）冷数据： GC选择的剩余数据为冷数据。

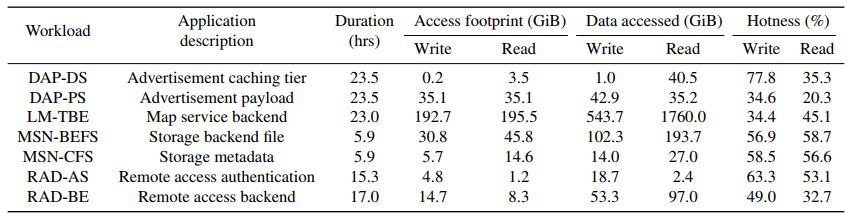
（3）热写数据：会自动在GC和RS过程中失效，被清理，或者转为冷数据或热数据。

### 3.2.4 实验与评估

（1）实验测试数据

使用了Microsoft生产服务的真实I/O工作负载数据来对HRM方法进行评估，这些工作负载分为不同类别，包含了多种实际情况，如表3.5。

表3.5 测试数据的特征



（2）对比实验设置

使用3种方法进行实验对比，并且使用无限位ECC作为对照组，即将其他3组的各种实验数据以无限位ECC的倍数来表示。

**无限位ECC**：作为对照组，无限位ECC可通过一次硬解码纠正所有错误，代表基准性能，可认为任何性能下降都是由排队延迟和垃圾回收引起等问题的。

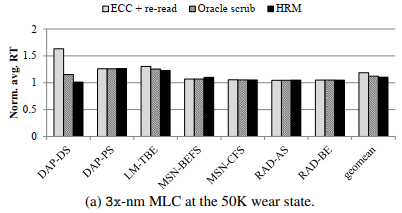
**ECC+重读**：具体配置依赖于不同的闪存内部控制器，主要是通过反复重新读取数据来尝试纠错，并且直到纠正错误为止。

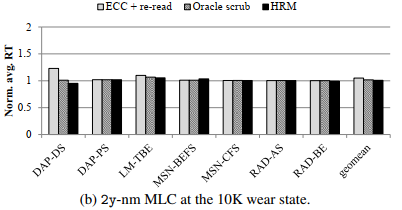
**Oracle scrub方法：**知道所有数据的预期错误数，并在错误累积之前预防性地重新定位它们。是一种理想型的可靠性最优模型。普遍用于SSD性能对比实验。

**HRM**：本文提出的整体可靠性管理方案。有选择地对由垃圾收集器收集的数据采用冗余，根据其冗余级别有条件地重新读取数据，并通过成本效益分析明智地管理数据清理。

（3）实验结果

最终由图3.11可以看出，HRM方法与其他两种方法相比，平均响应时间有一定的降低，因此整体的性能有所提高。





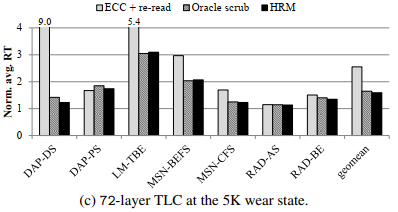


图3.11 多种方法的响应时间对比结果

而对于可靠性保证来说，这里主要是认为前期已通过RBER模型来分析并制定了HRM方法，HRM与当前使用的方法相比已经做到了对所有数据的可靠性进行覆盖，只是对可靠性保证算法进行了简单的优化，简化了冗余的可靠性算法，因此后续只需再专注于性能的提升即可。当然这也是本文的一个局限性，而且作者并未在真实的环境中进行测试，所以仍需要后续进一步的证明。

## 3.3 通过利用自我恢复和温度意识来提高3D NAND闪存设备的可靠性

### 3.3.1 介绍

本论文出自IEEE HPCA [3]，这篇论文主要考虑的是SSD的错误自恢复。论文更偏向于物理或底层层面的可靠性改进。先前已有工作研究了平面2D-NAND闪存的自恢复效应已经相关的做法，而且可通过高温来加速自恢复并提高闪存寿命。但是作者发现如果将该方式照搬到3D NAND闪存其效果并不理想，这主要是因为两种闪存的物理结构有很大差异。因此本文为3D NAND闪存提出了一个新模型：HeatWatch 。其关键思想是通过监控当前SSD温度和擦写停留时间，来调整读取数据的电压，让电压更加适应当前SSD的工作温度和擦写间隔时间，优化了读取的参考电压，从而增加了重读的准确性。最终本文证明了模型的准确性且误差仅为4.9％。使用此机制可将原始误码率降低93.5％，提高了3D NAND的可靠性，SSD整体寿命提高3.85倍。

### 3.3.2研究背景

首先作者根据3D NAND-based SSD的性质，分析了造成误码的物理层面上的原因。

（1）**保留错误**

对于3D NAND闪存来说隧道氧化物和电荷陷阱都是不完善的绝缘体，所以在单元空闲时电荷既可以通过隧道氧化物层泄漏到衬底，也可以通过电荷陷阱层泄漏到相邻的单元。

（2）**写异常错误**

同一页中的单元被一起写，每个编程脉冲期间都会注入随机量的电荷，许多因素会增加随机性，导致写的数据错误，即部分存储单元内存入的电子数与预期相比会增多或减少。

（3）**磨损**

多次重复擦写会导致氧化物层的磨损，隧道氧化物上会有电应力，大量电子被无意识的捕获在隧道氧化物内，会导致形成泄露路径，加速电子损耗。同时之后无法完全擦除所有电子，导致之后可能一直出现错误。

虽然存储单元会发生电子泄露，但是在一组连续的写入擦除时间内，闪存会通过“自恢复效应”的现象自行恢复了误码。自恢复效应已经在之前的研究中被证实，并且可针对2D-NAND闪存实现具体应用。同时，实验发现自恢复的速度和当前的温度之间是有一定的联系的。以下列出自恢复效应的两个研究结论：

（1）在Dwell Time（一组连续相邻的擦写操作之间的空闲/驻留时间）内闪存会通过一种称为自恢复效应的现象自行修复，将位于氧化层的异常电荷释放，即异常电荷会自动释放。

（2）NAND闪存的较高工作温度可提高电子迁移率，在高温下较短的保留时间与在室温下较长的保留时间具有相同的保留损失效果。

### 3.3.3 观察结论

之后本文开始尝试分析停留时间，温度，自恢复周期的关系：（1）停留时间和自我恢复效应的大小之间有什么关系。（2）工作温度如何影响保留次数和程序变化错误。（3）根据执行的恢复周期数，自我恢复的好处如何变化。希望构建一个适合于3D-NAND的自恢复模型，来降低误码率，从而提高可靠性。

首先在前期制定了一系列对比试验并得出了以下结论。

（1）先前的工作表明，在块上执行的P / E周期的最后10％，自我恢复效果的大小与停留时间相关。因此实验先通过写入伪随机数据来使存储单元达到3000个擦写周期，立即开始进行300个周期的测试实验。如图3.12可知，图的横轴为保留时间，表明某次写入数据后经过的时间，图的左半边（保留时间小于10s）为写异常的错误影响区域，表明这段时间内错误基本上都是由写异常的错误造成的误码，而对应的右侧为保留错误的影响区域，最后经整理后得到图3.13，可以看到擦除操作后更长的驻留时间（Dwell Time）可以缓解3D NAND闪存中的保留错误，而且自恢复效应以驻留时间的对数线性降低保留丢失的速度，但对写异常几乎无影响。

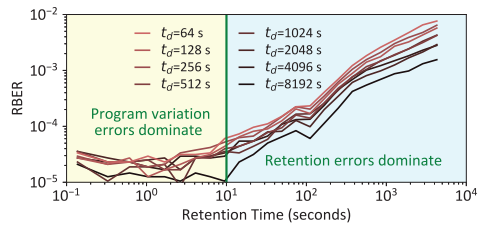


图3.12 不同Dwell Time驻留时间下R T保存时间与误码率关系

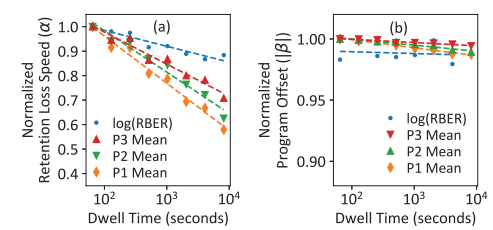


图3.13 驻留时间与保存损失和写偏移关系

同时可发现驻留时间较长时，平均阈值电压变化较小，这也表明了自恢复效应是有效的。

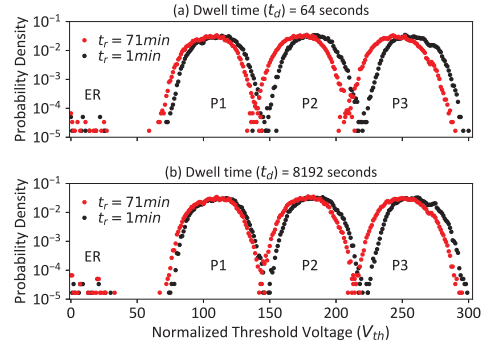


图3.14 驻留时间越长，电压偏移越小

（2）温度效应以超线性速率增加了保留损失速度，并增加了写异常偏移量。

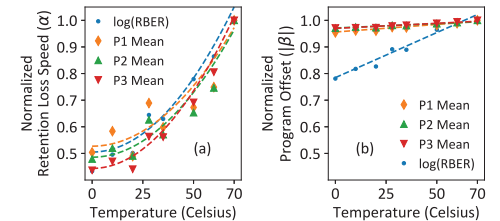


图3.15 在不同工作温度下保留损失速度(左)与写偏移(右)变化

保存时间短时，温度高会降低写异常引发的错误，但是保存时间长时，温度高反而会加速电子泄露速率。

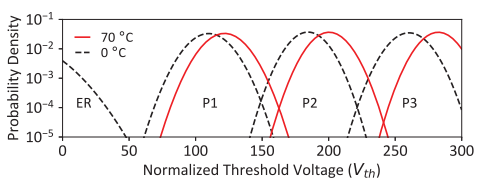


图3.16 写温度更高，电压重叠更小

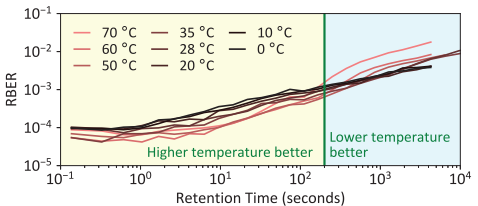


图3.17 不同工作温度下保留时间与RBER的关系

（3）在20个恢复周期后，由于自我恢复而导致的保留损失速度的降低变得微不足道。

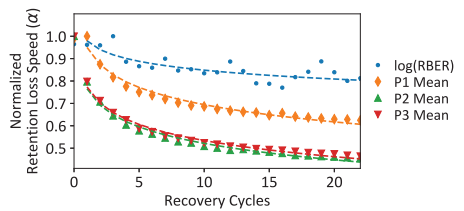


图3.18 恢复周期增加保留损失速度降低

### 3.3.4 自恢复模型

然后作者根据擦写次数，保留时间，驻留时间等值制定了公式来调节重读的电压，主公式为URT，其中的子公式的部分值为参数，是由作者反复调试训练得出的，之后发布时会作为默认初值，并且可自行调整。

（1）电压调整公式URT，即统一恢复和温度。URT为主公式，之后为多个子公式。



（2）PVM写异常组件：数据写入后立即预测阈值电压和Y0初始电压，



（3）RDTM 保留/停留组件：计算AF调节参数，同时将任何温度下保留时间缩放为室温下有效时间。





（4）SRRM自恢复和保留组件：预测阈值电压偏移在和由保留损失引起的RBER变化



总之Heat Watch由两大部分组成：

（1）三个跟踪组件：分别监视和跟踪有效记录SSD温度，停留时间，保留时间和P / E周期数。

（2）两个预测组件，它们使用此跟踪信息计算URT模型，以准确预测每次读取操作的最佳读取参考电压。

对于公式中三个主要参数的来源：

（1）SSD温度：温度传感器

（2）P/E Cycle：内部已包含

（3）停留时间与保留时间：使用SSD控制器计算与记录，同时也会考虑最近20个恢复周期的影响。

对于模型内部数据占用开销。存储每个闪存块的写温度，停留时间，编程时间。在后台每秒更新。每次读取操作会计算URT模型。

### 3.3.5 实验与评估

为了评估HeatWatch，使用4个方法进行对比实验。

**（1）固定的Vref**：始终使用默认的读取参考电压读取数据。

**（2）只考虑保留时间**：基于仅考虑P / E周期计数和保留时间的模型预测最佳读取参考电压。无论停留时间和温度如何，此模型始终假定固定的保留损耗速度。

**（3）HeatWatch**：本文提出的方法，可通过跟踪停留时间和温度并使用URT模型来准确预测最佳读取参考电压。

**（4）Oracle**：它总是最好使用测量的最佳读取基准电压，一点不产生任何性能开销。需要注意此方法仅为理想的用于测试对比的模型，实际上无法实现的。

最终在使用了28个常用的实际存储迹线的测试后，可以发现，使用了HeatWatch机制后，误码率接近了理想化的Oracle模型，这表明了HeatWatch有效的提高了可靠性。

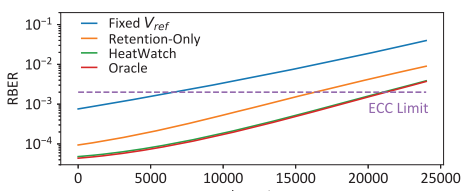


图3.19 四个模型的RBER结果对比

之后是对每个工作负载做相同的实验，来确定四种方式下每个工作负载的生存期。同样，HeatWatch非常接近Oracle的生命周期改进，表明该方法显著的提高了3D NAND闪存设备的使用寿命。

****

图3.20 4种模型不同工作负载下P/E cycle的生命周期对比

# 4. 总结

SSD可显著提高计算机的存储性能，但是随着工艺制程的发展，误码率不断提高，其可靠性问题变得尤为重要。通过不断地研究，目前已经发现了造成SSD可靠性的几种问题来源，以及几种不同强度的解决方式。但是我们对其可靠性问题还有待进一步研究。本文提出了三篇相关的研究文献。首先是对企业SSD可靠性问题来源的研究，通过对企业大数据的统计分析，得出了几点对SSD更换率的影响因素。之后是对算法的比较与权衡，将数据分为冷热数据等几种不同类型，然后有针对性的使用更合适的可靠性保证算法来提高整体性能。最后是在物理层面，利用自恢复现象，通过HeatWatch模型来监控当前工作温度动态调整读取数据的电压，从而降低了误码率，提高了可靠性也增加了SSD的使用寿命。

# 5. 参考文献

1. Maneas S, Mahdaviani K, Emami T, et al. A Study of {SSD} Reliability in Large Scale Enterprise Storage Deployments[C]. 18th {USENIX} Conference on File and Storage Technologies ({FAST} 20). 2020: 137-149.
2. Kim B S, Choi J, Min S L. Design tradeoffs for {SSD} reliability[C] .17th {USENIX} Conference on File and Storage Technologies ({FAST} 19). 2019: 281-294.
3. Luo Y, Ghose S, Cai Y, et al. HeatWatch: Improving 3D NAND flash memory device reliability by exploiting self-recovery and temperature awareness[C]. 2018 IEEE International Symposium on High Performance Computer Architecture (HPCA). IEEE, 2018: 504-517.
4. SSDFans.深入浅出SSD固态存储核心技术、原理与实战[M].机械工业出版社,2018:1-121.
5. Wang M, Hu Y. i-RAID: a novel redundant storage architecture for improving reliability, performance, and life-span of solid-state disk systems[C]. Proceedings of the 31st Annual ACM Symposium on Applied Computing. 2016: 1824-1831.