

非易失性存储器存储软件技术综述 和主存储器系统

IEEE 成员斯巴塞米塔尔和 IEEE 高级成员杰弗里维特尔

摘要—与传统存储技术(如动态随机存取存储器和磁性硬盘驱动器)相比,闪存、相变随机存取存储器、自旋转移扭矩随机存取存储器和电阻随机存取存储器等非易失性存储设备具有多种优势和挑战。在本文中,我们介绍了软件技术的概况,这些技术被提出来利用非易失性存储器的优点并减轻其缺点,当用于设计存储器系统时,特别是辅助存储器(例如,固态驱动器)和主存储器。我们按照几个维度对这些软件技术进行分类,以突出它们的异同。鉴于非易失性存储器越来越受欢迎,我们相信这项调查将推动非易失性存储器软件技术领域的进一步研究。

索引术语—回顾、分类、非易失性存储器(NVM) (NVRAM)、闪存、相变随机存取存储器(PCM) (PCM RAM)、自旋转移扭矩随机存取存储器(STT-RAM) (STT-MRAM)、电阻随机存取存储器(ReRAM) (RRAM)、存储级存储器(SCM)、固态驱动器(SSD)



1 介绍

对于从手持嵌入式系统到大型超级计算机的所有计算系统,存储系统在决定它们的功耗、可靠性以及毫无疑问的应用性能方面起着主要作用。一流应用程序不断增长的数据密集型特性要求在功耗和成本的限制下获得比传统内存技术更高的性能。例如,每分钟有超过 100 小时的视频和 250 张 K 照片分别上传到 YouTube [1] 和脸书[2]。同样,预计未来的极端规模系统将处理几十亿字节的数据;在这些系统中使用硬盘来支持检查点/重启可能会降低其 50% 以上的性能[3]。此外,这些系统至少需要 100 petabytes 的主内存,如果采用今天的 DDR3 DRAM 设计,将消耗 52 MW 功率,这远远超过整个系统规定的 20 MW 功率预算[4], [5], [6]。

这些要求促使研究人员探索新的记忆技术。非易失性存储器(NVM)设备,例如闪存、相变存储器(PCM)、自旋转移扭矩随机存取存储器(STT-RAM)和电阻随机存取存储器(ReRAM),在设计从高速缓存到主存储器到辅助存储器的存储器体系的不同组件方面,可以提供优于现有设备的许多好处

存储。这些非易失性存储器的潜在优势源于其预计的物理特性,这可能使它们既能消耗非常低的功率,又能提供比预计的传统技术高得多的密度。例如,典型的静态随机存取存储器单元的尺寸在 125-200 的范围内,而相变存储器和闪存单元的尺寸分别在 4-12F² 和 4-6F² 的范围内,其中 F 表示给定技术节点中最小的光刻尺寸[3], [7]。由于其特性,非易失性存储器越来越多地被部署在产品中。例如,英特尔 TurboMemory 使用少量闪存作为缓存来缓冲磁盘数据[8]。同样,许多大型企业也在数据中心[9]和超级计算机[10], [11]中使用闪存。事实上,最近有消息称,美国能源部名为 Summit 的下一代超级计算机,其 3400 个节点中的每一个都将拥有 800 GB 的 NVM 这种非易失性存储器可以配置为突发缓冲器或扩展存储器[12]。

另一方面,这些非挥发性物质也有一定的局限性。例如,它们的写入耐久性比传统存储器低几个数量级,并且写入操作通常具有高延迟和能量成本。很明显,除非在设备和架构层面证明了抵消这些限制的创新,否则社区将需要软件和系统层面的技术来整合未来存储系统中的非易失性存储器技术。

南米塔尔是未来技术集团,橡树岭国家实验室,橡树岭, TN 37830。电子邮件:mittals@ornl.gov.

维特尔是未来技术集团、橡树岭国家实验室和佐治亚州佐治亚理工学院的成员。电子邮件:vetter@ornl.gov.

手稿于 2014 年 11 月 13 日收到; 2015 年 5 月 27 日修订; 2015 年 6 月 2 日接受。出版日期:2015 年 6 月 8 日; 当前版本的日期为 2016 年 4 月 13 日。建议布莱特威尔接受。

有关获得本文转载的信息,请发送电子邮件至:reprints@ieee.org, 并参考下面的数字对象标识符。

数字对象标识符编号 10.1109/TPDS. 2015. 201424298086

1.1 贡献

在本文中,我们对软件和系统级技术进行了综述,这些技术被提出来利用非易失性存储器在存储器层次结构中使用时的优点并减轻其缺点,特别是辅助存储器(例如固态驱动器)和主存储器。我们按照几个维度对这些软件技术进行分类,以突出它们的异同。我们还讨论了那些比较或结合多种非易失性或传统存储技术的论文。

表 1
存储器技术的近似设备级属性[3], [7], [14]

	孔度	访问粒度	读取延迟	写入延迟	擦除延迟	忍耐	储用功率
硬盘驱动器	不适用的	512 B	5 毫秒	5 毫秒	不适用的	> 1015	1W
SLC 闪存	4 - 6F 2	4 KB	25 毫秒	500 毫秒	2 毫秒	> 1015	0
动力系统控制模块	4 - 12F 2	64 B 50 ns	500 ns	N/A	108 109	0	
STT-RAM	6 - 50F 2	64 B 10 纳秒	50 纳秒	不适用	> 1015	0	
ReRAM	4 - 10F 2	64 B 10 ns	50 ns	N/A	1011	0	

1.2 术语和范围

根据其他研究人员(例如[13]), 我们使用单片机来指代字节可寻址非易失性存储器, 如 STT-RAM、PCM 和 ReRAM, 使用非易失性存储器指代所有 SCMs 和闪存。我们讨论了为单层细胞(SLC)和多层细胞(多层细胞)理论提出的技术。除非另有说明, 否则闪存指的是 NAND 闪存, 而不是 NOR 闪存, 固态硬盘指的是基于闪存的固态硬盘, 而不是基于 PCM 的固态硬盘。此外, 为其他存储器技术(例如, 硬盘或动态随机存取存储器)提出的一些技术也可以应用于非易失性存储器, 然而, 我们只包括在非易失性存储器环境中提出的那些技术。由于不同的技术在不同的环境下被评估, 我们只展示它们的主要思想, 而不展示量化结果。

以前的一些论文集中在闪存的设备级特性[15]或者回顾 SCMs 在高速缓存和主存储器中的应用[7], [16]。通过比较, 本文考察了为所有非易失性存储器提出的系统级和软件级技术, 以解决几个重要方面, 如持久存储系统的设计、寿命增强、可靠性、成本效率、能量效率、混合存储系统的设计等。我们根据几个关键特征/特性对技术进行分类, 以突出它们的相似性和差异性。通过对现有的 NVM 管理技术前沿进行综合概述, 我们旨在为该领域的未来研究提供明确的方向。这份调查报告有望对研究人员、操作系统设计师、计算机架构师和其他人有用。

论文的其余部分组织如下。第 2 节介绍了国家风险管理的特点和局限性的背景。第三节根据几个参数对国家风险管理的研究项目进行了分类, 并对其中的几个进行了讨论。第四节讨论了研究多种存储技术集成或比较的研究项目。最后, 第 5 节提出了结论和未来的挑战。

2 存储器技术概述

在本节中, 我们简要总结了不同内存技术的特性和挑战。为了比较起见, 我们还讨论了传统的存储技术, 如硬盘驱动器和动态随机存取存储器。表 1 显示了不同内存技术的设备级属性。请注意, 这些值只能作为代表, 因为正在进行的研究可能会导致这些参数的变化。关于存储器技术的器件级属性的更多细节, 我们请读者参考以前的著作[3], [7], [17], [18], [19], [20], [21]。

在表 1 中, 访问粒度指的是每次访问中读取/写入的最小数据量, 持久性指的是一个存储块在变得不可靠之前能够承受的写入次数。所有非易失性存储器的共同特性是它们的写延迟/能量明显高于读延迟/能量。此外, 在正常情况下, 它们可以在不需要任何备用电源的情况下保留数据数年。不同非挥发性物质的具体特性将在下面讨论。

2.1 闪存

闪存有三种类型的操作, 即读取、编程(写入)和擦除。写操作只能将位从 1 更改为 0, 因此, 将页面中的位从 0 更改为 1 的唯一方法是擦除包含将该页面中的所有位设置为 1 的页面的块。由于擦除操作明显慢于写入操作, 闪存固态硬盘使用闪存转换层 (FTL) [22], 这有助于“隐藏”擦除操作, 从而只对上层进行读/写操作。FTL 在闪存上维护一个从上层到物理地址的虚拟地址映射表, 并使用它来执行损耗均衡。

闪存中的页面分为有效、无效(包含无效数据)和空闲(可用于存储新数据)。为了隐藏擦除操作的延迟, FTL 执行错位写入, 即写入空闲页面, 无效写入页面的前一个位置, 并更新映射。读写操作在页面粒度上完成, 而擦除操作在块粒度上完成。通常, 页面和块的大小分别为 4 和 256 KB。ftl 还执行垃圾收集(GC), 即通过擦除块并将其中的任何有效页重新定位到新位置(如果需要)来回收无效页。显然, 考虑到 FTL 对闪存固态硬盘性能的重要影响以及其设计中涉及的大量因素(例如, 垃圾收集策略、页面 v/s 块映射、映射表的大小和存储位置等。), FTL 需要自己的讨论, 因此, 我们请读者参考以前的作品了解更多细节[22], [23]。闪存的单元尺寸为 4-6F 2, 而静态随机存取存储器的单元尺寸为 120-200F 2 [7]。显然, 由于其高密度和延迟以及低写入耐久性, 闪存通常适合用作存储设备或动态随机存取存储器和硬盘之间的缓存层。闪存是一种成熟的非易失性存储器技术, 正在开发中并由几家商业制造商部署[9]。

与旋转介质(即硬盘), 闪存是基于半导体芯片, 这导致紧凑的尺寸, 低功耗和更好的性能

随机数据访问。此外，固态硬盘没有活动部件，没有机械磨损，并且耐热和防震。然而，固态硬盘相对于硬盘的性能优势在很大程度上取决于工作负载特性。例如，已经表明，对于许多写密集型科学工作负载，固态硬盘可能只比硬盘提供边际收益[24]。此外，由于成本较高，固态硬盘无法完全取代硬盘[25]。因此，需要在系统级解决的闪存研究挑战包括通过最小化写/擦除操作次数来提高寿命、损耗均衡、管理故障块，以及通过保留松弛和设计混合存储系统来提高性能。

2.2 动力系统控制模块、STT-RAM和ReRAM

这三种内存技术(称为SCMs)区别于闪存的最重要特征是它们是字节可寻址的。计算机系统传统上使用动态随机存取存储器作为易失性存储器，硬盘和闪存作为永久存储器。然而，它们延迟的差异导致了它们的界面的巨大差异[26]，[27]。这些SCMs提供了与闪存相似或更好的存储容量和耐用性，同时提供了与动态随机存取存储器相当的延迟。对于这些人来说，SCMs有希望被用作通用存储技术。尽管与闪存和动态随机存取存储器相比，SCMs还不太成熟，但近年来已经为开发和利用它们进行了大量的研究[7]，例如，最近已经演示了一个16 Gb的ReRAM原型[28]，其特点是八存储体并发动态随机存取存储器样的核心架构，并且

1 GB/s DDR接口。为这些存储器提出的系统级技术解决了几个关键的研究挑战，例如将它们集成到存储器/存储层次结构中以设计持久存储器系统和补充传统存储器技术，以及提高寿命、可靠性和性能等。

第3节和第4节讨论了为解决这些问题而提出的技术。

3 非挥发性物质管理技术

表2根据使用的非易失性存储器和使用非易失性存储器的存储器层级对研究项目进行了分类。许多涉及二级存储的研究需要较长的执行时间和/或操作系统操作的建模，这不是典型的用户空间模拟器所能提供的。模拟器也可能使用简单的模型，因此错过了关键的真实世界细节。另一方面，真实的硬件平台不像模拟器那样提供实验灵活性，并且考虑到SCMs的新兴特性，具有基于单片机的存储的真实硬件平台通常也是不可用的。这在国家风险管理的研究中提出了一个挑战，使得选择一个合适的评估平台至关重要。因此，表2根据技术是使用模拟还是真实硬件(如中央处理器或现场可编程门阵列)进行评估来进一步分类，以帮助读者获得洞察力。

表2还根据优化目标和基本方法对技术进行了分类。我们现在

从架构、固件、中间件(I/O、操作系统等)开始，按照软件栈中抽象级别的自下而上的顺序讨论其中一些技术。)一直到编程模型/API(应用程序编程接口)，尽管请注意，其中一些技术跨越了这些“边界”。

3.1 写入/擦除开销最小化

一般来说，非易失性存储器的写入耐久性较低，并且由于工作负载引入的写入变化，一些块可能接收到比其余块高得多的写入次数。这个问题可以通过最小化写入/擦除并将其均匀分布在所有块上(称为损耗均衡)来解决。我们现在讨论基于这些方法的技术。

黄等人[43]提出了一种减少固态硬盘写入流量的技术，这也增加了固态硬盘的寿命。他们的技术对语义块采用增量编码，对数据块采用重复数据消除。对于每个块写请求，它决定是语义块还是数据块。对于数据块写入，它计算MD5摘要来确定它是否是重复的块写入。对于重复的块，它们的技术在找到的哈希条目中返回现有的块号，从而避免了分配哈希表内存的需要。如果是语义块(包括超级块、组描述符、数据块位图等。)，他们的技术计算相对于原始内容的内容增量，然后将增量附加到增量记录区域。因为语义块比数据块被访问得更频繁，每次更新带来的变化非常小；他们的技术减少了写入次数。

Papirla和Chakrabarti[84]指出，闪存中的写入延迟和能量与数据相关，例如，在四级闪存中，写入“01”和“10”模式比写入“00”和“11”模式会产生更高的延迟和能量。他们提出了一种数据编码技术来减少写入内存的“01”和“10”模式的数量。此外，他们的技术不会损害应用程序的错误性能。此外，在四电平或非门闪存中，“11”状态被称为擦除状态，因为它不需要对存储单元进行编程，而其余状态被称为编程状态。通过减少“01”和“10”模式，他们的技术减少了编程周期的数量，从而提高了闪存的寿命。

Grupp等人[52]使用WOM(一次写入存储器)编码方案来增加闪存的寿命并节省能量。通过使用额外的位，WOM码允许写入多个逻辑值，即使物理位只能传输一次。凭借这一点，WOM代码允许在擦除数据之前将数据写入数据块两次，从而减少了所需的擦除次数。

在嵌入式系统中，闪存可以用作主要存储器，但是其大的数据访问粒度(例如高速缓存中使用的4 KB v/s 64 B)和小的耐用性带来了挑战。Shi等人[82]提出了一种技术，以减少闪存主存储器上的写入次数，从而提高其寿命，并弥合访问粒度之间的差距。他们观察到，由于数据的局部性，连续写入中访问的数据来自有限的页面。基于此，他们使用受害者缓存以及写缓冲区来

表2
技术的分类

分类参考	
使用非易失性存储器	
闪光	[8], [11], [18], [19], [20], [22], [23], [24], [25], [27], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70], [71], [72], [73], [74], [75], [76], [77], [78], [79], [80], [81], [82], [83], [84], [85], [86], [87], [88], [89], [90], [91], [92], [93], [94], [95], [96], [97], [98], [99], [100]
脉冲编码调制	[3], [13], [14], [26], [27], [31], [51], [68], [75], [85], [90], [99], [101], [102], [103], [104], [105], [106], [107], [108], [109], [110], [111], [112], [113], [114], [115], [116], [117], [118], [119], [120], [121], [122], [123], [124], [125], [126], [127], [128], [129], [130], [131], [132]
STT-RAM	[13], [26], [27], [101], [110], [112], [113], [125], [128], [133], [134]
ReRAM	[13], [26], [112], [125], [126], [127], [130], [132]
使用非易失性存储器的存储器层级	
二级存储器(及其缓存等。)	[8], [18], [19], [20], [22], [23], [24], [25], [31], [32], [33], [34], [39], [43], [44], [45], [46], [47], [48], [49], [50], [54], [55], [57], [58], [59], [60], [63], [64], [65], [66], [67], [68], [71], [72], [73], [74], [75], [76], [77], [78], [79], [80], [81], [85], [86], [87], [88], [90], [91], [92], [93], [94], [97], [98], [100], [102], [104], [106], [107], [124], [129]
主存储器	[13], [14], [62], [82], [85], [99], [101], [102], [103], [106], [110], [111], [115], [116], [117], [118], [119], [120], [121], [122], [125], [128], [131], [133], [134]
片内高速缓存	[128], [133]
评估平台	
模拟器	[3], [19], [22], [23], [24], [26], [30], [31], [33], [36], [39], [42], [45], [46], [48], [49], [56], [57], [58], [59], [60], [63], [64], [68], [70], [71], [72], [73], [74], [75], [76], [79], [80], [81], [82], [84], [85], [86], [91], [92], [93], [94], [96], [98], [99], [100], [101], [102], [103], [104], [110], [111], [114], [116], [117], [118], [120], [121], [128], [129], [130], [131], [133]
真正的硬件	[8], [13], [18], [24], [26], [27], [34], [38], [42], [43], [44], [46], [52], [53], [55], [62], [64], [65], [67], [68], [69], [74], [83], [87], [89], [90], [95], [96], [103], [105], [106], [109], [110], [112], [113], [115], [122], [125], [126], [127], [132], [134], [135]
研究/优化方法/目标	
绩效改进	[3], [8], [11], [13], [18], [19], [22], [24], [25], [27], [31], [32], [39], [40], [42], [43], [44], [52], [55], [56], [57], [58], [59], [60], [63], [64], [65], [66], [67], [68], [69], [71], [72], [73], [74], [75], [76], [77], [79], [80], [83], [84], [85], [86], [87], [88], [89], [90], [91], [92], [94], [95], [97], [98], [99], [103], [105], [106], [107], [108], [109], [110], [112], [115], [126], [131], [132], [133], [134], [135]
能源效率	[25], [31], [40], [42], [52], [84], [86], [90], [91], [99], [101], [120], [123], [131]、[133]
寿命改进	[19], [23], [31], [32], [33], [34], [36], [39], [43], [44], [46], [47], [48], [49], [50], [51], [52], [53], [57], [59], [65], [69], [72], [73], [75], [78], [81], [82], [84], [85], [88], [91], [93], [99], [100], [106], [108], [116], [117], [118], [119], [121], [122], [123], [124], [125]
损耗均衡	[19], [23], [30], [31], [45], [46], [47], [48], [49], [50], [53], [91], [93], [99], [107], [108], [111], [117], [124]
写/擦除开销最小化	[43], [44], [57], [59], [64], [65], [71], [73], [75], [82], [85], [91], [99], [100], [111], [120], [123], [129], [133]
打捞故障块并擦洗	[48]、[51]、[80]、[81]、[85]、[116]、[118]、[119]、[120]、[121]、[122]

NVM 保留弛豫[56], [72], [80], [92], [102]
持久性和一致性[13], [13], [62], [76], [96], [102], [110], [113], [114], [128], [130], [132], [133], [134], [136]
检查点、可靠性和纠错

[3], [33], [34], [35], [45], [53], [56], [60], [72], [74], [76], [78], [79], [81], [84], [85], [91], [96], [98], [103], [112], [115], [116], [117], [118], [120], [122], [128], [129]

数据值相关优化[51], [84]
成本效率[20]、[25]、[39]、[68]、[71]、[87]、[97]

执行写合并，因为对于多个末级缓存回写操作，只对闪存主存储器进行一次(或几次)写操作。

Qureshi 等人[131]注意，只有动力系统控制模块写入中的设置操作慢，而复位操作几乎与读取一样快。因此，通过在预期写入之前提前对所有位执行置位，可以减少写入时间，这也有助于节能。基于此，他们的技术会在缓存中的行变脏时立即启动 SET 操作。

3.2 磨损平衡

Chang 等人[47]提出了一种静态磨损均衡算法。在他们的算法中，块被分成多个集合，每个集合与块擦除表中的一个位相关联。首先，下注中的所有位都是“0”。如果集合中的一个成员在该时间间隔内被擦除，则它的相关位转换为 1。还会记录任何时间间隔内的擦除总数。如果擦除次数与 BET 中 1 的数量之比达到预定阈值，则随机选择对应位仍然为 0 的集合。后来，所有

该集合中的有效数据被移动到一个自由块集合，并且前一个集合被擦除以备将来使用。

王和黄[46]观察到，由于操作系统在更高的抽象层次上了解文件（例如，文件类型数据所属的、使用它们的应用程序等。），这些信息可以用来给低级别的FTL提供提示。FTL使用这些提示来推断文件的更新频率和新近性，从而执行更好的损耗均衡，因为块分配是以这样一种方式完成的，即年轻块被分配给热数据，而旧块被分配给冷数据。

Chang和Huang[23]提出了一种损耗均衡算法，旨在减少较老（即具有高擦除计数的块）块的损耗。他们的技术跟踪块的擦除最近度，并且每当老块的擦除最近度变得比平均值高预定阈值时，具有低更新最近度的逻辑块被重新映射到这些老块。该算法动态调整阈值，以平衡损耗均衡的收益和成本。

王和黄[49]提出了观察损耗均衡(OWL)，旨在通过监控时间局部性和块利用率来主动避免擦除的不均匀性。他们的技术使用一个名为块访问表(BAT)的表，其中一个块是逻辑块(而不是闪存块)。BAT存储最近被重写的逻辑块的访问频率。使用BAT，OWL算法对逻辑块的数据进行排序，并相应地分配闪存块。排名用于预测逻辑块在不久的将来的相对访问频率。通过这种方式，OWL以主动的方式将数据放入合适的块中，以实现损耗均衡。

Jeong等人[93]表明，通过用较低的擦除电压缓慢擦除闪存块，可以提高闪存的耐久性。基于此，他们提供多种具有不同工作电压和速度的写入和擦除模式，以延长闪存的使用寿命，同时对应用吞吐量的影响最小。在软件级别，垃圾收集器和磨损均衡器被修改以利用这些写/擦除模式。例如，不是对所有块使用相同的耐久性（如在基线闪存中），而是他们的磨损均衡器使用有效耐久性（如他们的寿命延长方法所实现的）来在闪存块之间平均分配有效磨损。

Im和Shin[124]提出了一种基于PCM存储的损耗均衡技术。他们的技术计算PCM页面上的写入次数，如果一个逻辑页面频繁更新，他们的技术会为其分配更多的物理页面，以平衡所有页面上的写入。因此，基于更新频率，逻辑页面具有分配给它们的不同数量的物理页面。他们还表明，由于损耗均衡，他们的技术可以保持额外写入的数量很少。

3.3 挽救故障块和容忍故障内存系统的“原始寿命”是由块的第一次故障决定的。通过容忍几个块的故障和/或回收，可以显著提高该寿命有问题的街区。为此已经提出了几种技术，我们现在讨论其中的几种。

刘等人[78]提出了基于容错的技术来提高基于闪存的固态硬盘缓存的寿命。由于直写固态硬盘缓存中的错误数据可以通过访问硬盘来恢复，因此他们的一项技术将无法纠正的错误转换为缓存未命中，从而从硬盘中获取有效数据。另一种技术利用固态硬盘缓存容量的一小部分，在闪存达到磨损阈值时增加纠错码强度，因此固态硬盘缓存继续以降低的容量运行。

蔡等人[81]提出了一种通过降低原始误码率来提高闪存寿命的技术，即使闪存经受了远远超过其额定寿命的高P/E周期。他们的技术定期读取闪存中的每一页，使用简单的ECC纠正其错误，并在页面累积比简单ECC纠正的错误更多的错误之前，将页面重新映射到不同的位置或在其原始位置重新编程。因此，通过解决在闪存中形成最主要错误类型的保留错误，闪存的寿命得以提高[34]，[35]。

王和黄[48]提出了一种在闪存中打捞坏块以延长其寿命的方法。他们的方法基于这样的观察：坏块中的许多页面可能仍然是健康的，因此，在几个页面出现故障时丢弃一个块会导致浪费和很短的生命周期。他们的方法将一组坏块的健康页面组合在一起，形成一组更小的虚拟健康块，可用于存储冷数据。

Maddah等人[51]提出了一种物理块备用方案，当闪存或PCM存储器由于写入耐久性限制而出现故障时，该方案会延迟故障块的退出。当达到其写入耐久性极限时，动力系统控制模块或闪存模块显示“卡在”故障，这意味着它卡在0或1，仍然可以读取但不能重新编程[35]。因此，错误的发生变得依赖于数据；只有当一个不同的位值被写入一个错误的单元时，错误才会出现。基于此，在对块的不成功写入时，他们的方案不会立即停用块。相反，从备用块池中临时借用一个备用块。稍后，在原始(故障)块上再次尝试写操作，如果数据与该块上的现有数据相同，则写操作可能会成功。在这种情况下，上面的备用块被返回到池中。当故障块显示故障超过阈值时，它最终被归类为坏块并退出使用。实际上，对于相同的寿命改进，他们的方案降低了备用块的需求，或者对于相同的备用池容量实现了更高的寿命。

Yoon等人[119]提出了一种细粒度的重映射技术来防止非易失性存储器的错误。传统上，当一个块累积的磨损故障超过可以纠正的数量时，它将被禁用并重新映射。他们的技术利用磨损的存储块中仍然起作用的单元，使用高度冗余的代码来存储重定向地址。备用区由操作系统在主内存中动态创建。当重新映射的块本身失败时，它将被进一步重新映射，这可能会创建链式重新映射。为了避免这种情况，他们的技术将最终地址写入原始的失败块。他们的细粒度重新映射技术的好处是，64 B块的失败不会导致整个4 KB页面的重新映射或禁用。

高等[122]提出了一种软硬件协同方法来容忍非易失性存储器中的故障。他们的方法通过使用垃圾收集托管语言(如Java、C#、JavaScript等)提供的内存抽象,使错误处理对应用程序透明。运行时确保内存分配从不使用失败的行,并在程序执行期间当内存行失败时移动数据。使用磨损均衡的传统纯硬件方案会延迟单个故障,但是,它们的局限性在于,由于磨损均衡,在大量写入之后,故障会在整个内存中均匀发生,从而导致碎片。相比之下,他们的技术使用低成本的“故障群集硬件”,该硬件在逻辑上将故障线路重新映射到该区域的顶部或底部边缘,以最大化可用于对象分配的连续空间。在第一次失败时,这个硬件还会安装一个指针,指向正常和失败行之间的边界。因此,他们的技术减少了碎片,提高了故障时的性能。

赵等[74]注意到,在不同的P/E周期下,闪存的原始误码率随保留时间的不同而有很大的变化。因此,用于确保可靠性的ECC过度保护了闪存,因为大多数页面显示出优于最坏情况的可靠性。为了利用剩余的纠错能力,他们提出了易出错的闪存芯片输入/输出链路的过时钟,这转化为更高的性能,并且ECC对闪存存储和控制器-闪存数据传输引起的错误负责。基于对数据读取路径上的过时钟和数据写入路径上的过时钟的研究,他们表明前者比后者更加有效和有利。这是因为,写路径上的过时钟会导致永久性数据存储错误,而读路径上的过时钟影响则更容易容忍。

伊佩克等人[116]提出了一种在发生硬故障时允许PCM容量适度下降的方法。他们的技术通过在两个有故障的PCM页面上复制单个物理页面来工作,只要在两个页面中没有有故障的字节位置。使用这种方法,物理内存的每个字节都可以由两个副本中的至少一个提供服务。因为即使在具有数百个位故障的页面中,找到这样一对页面的可能性也很高,所以与传统的错误检测技术相比,它们的技术大大提高了PCM存储器的寿命。

Sampson等人[85]提出了一些技术,这些技术在数据精度和nvm的性能、寿命和密度之间进行权衡,以适用于能够容忍数据错误的应用。他们的第一种技术通过减少用于写入MLC存储器的编程脉冲的数量来允许MLC存储器中的错误,这提供了更高的性能和密度。第二种技术使用已经耗尽硬件纠错资源的块来存储近似数据,这增加了存储器寿命。此外,为了减少失败位对整体结果的影响,优先考虑高阶位的校正。

一些多媒体和图形应用程序可以容忍微小的错误,因为这些错误不会被人类终端用户察觉[5],[137]。Fang等人[123]利用这一特性来减少对PCM的写入流量。在他们的技术中,如果最初存储的数据非常接近新的数据

要写入的数据,写操作被取消,原始存储的数据被作为新数据。这也降低了存储器的能量消耗。

3.4 保留松弛

闪存允许用写入速度和P/E(编程/擦除)周期来权衡保留期。例如,闪存可以被更快地编程,但是具有更短的保持时间保证,因此,写操作可以更快。该属性也适用于其他国家/地区。几种技术利用这一特性来优化非易失性存储器系统。我们现在讨论其中的几个。

Pan等人[80]提出了一种放宽闪存保持期的方法,以提高P/E循环耐久性和/或编程速度。为了避免在短时间内丢失数据,使用了数据刷新操作。此外,为了最小化刷新操作对正常输入/输出请求的影响,他们提出了一种调度策略,该策略进行了多种优化,例如给予输入/输出请求更高的优先级,更喜欢在固态硬盘空闲时发出刷新操作等。

史等[72]提出了一种提高闪存耐久性和写入性能的技术。基于闪存的误差模型,他们的技术在闪存寿命的不同阶段应用不同的优化。在第一阶段(即早期生命周期),保留时间和P/E周期被权衡,以通过最大化写入速度来优化性能。在第二阶段(即,中等寿命),写入操作分为热写入和冷写入,热写入加快,而冷写入减慢,以提高性能,而对耐久性影响很小。在最后一个阶段,写速度减慢以延长耐久性。他们还提出了一种智能刷新方法,只刷新接近保留时间的数据。

Liu等人[56]分析了不同的数据中心工作负载,发现对于大多数工作负载,数据会频繁更新,因此只需要保留几天。这比通常为NAND闪存指定的保留时间要短得多。可以利用保留时间的这一差距来优化固态硬盘的运行。此外,随着保留放宽,需要容忍的保留错误更少,因此,可以使用较低复杂度的ECC来保护数据,从而降低ECC开销[138]。

黄等人[79]提出了一种提高基于闪存的固态硬盘性能的技术。他们的技术根据页面的一致性和可靠性要求,有选择地用EDC(检错码)替换ECC。具体来说,当在下一个存储层写入没有备份的新数据时,他们的技术使用ECC,否则使用EDC。由于EDC的解码延迟要小得多,因此对受EDC保护的页面的读取访问被加速,这提高了性能。在数据损坏时,从较低的存储层访问页面,因此不会牺牲可靠性。

吴和张[92]提出了一种减少固态硬盘响应时间的技术。他们的技术监控工作负载的写强度,当写请求队列包含几个重叠的写事务时,它以更短的数据保留时间为代价来提高内存编程速度。后来写的时候

强度再次变低，短寿命数据被重写以确保数据完整性。他们还开发了一个调度解决方案来实现他们的写策略。

3.5 闪存输入/输出调度

吴和何[94]注意到，在闪存中，一旦向闪存芯片发出编程/擦除操作，随后的读取操作必须等到缓慢的编程/擦除操作完成。他们提出了不同的策略来暂停正在进行的程序和擦除操作，以服务于未决的读取，并在稍后恢复暂停的操作。在编程操作期间，页面缓冲区包含要写入的数据，当读取请求到达时，这些数据可能会丢失。为了解决这个问题，他们提供了一个影子缓冲区，在挂起期间存储页面缓冲区的内容。读操作完成后，它用原始数据重新加载页面缓冲区。

在基于闪存的磁盘缓存中，错位写入会增加垃圾收集的开销，还会降低闪存的性能。为了解决这个问题，Kgill等人[91]提出将基于闪存的磁盘高速缓存分成单独的读写区域(例如90%的读取区域和10%的写入区域)。读取关键闪存块位于读取区域，可能仅在读取未命中时逐出闪存块和页面。写入区域捕获对闪存的所有写入，并执行错位写入。磨损均衡适用于全球所有区域。将闪存划分为读/写区域减少了垃圾收集需要考虑的块数，从而显著减少了读、写和擦除操作的数量。

Lee等人[58]提出了一种半抢占式垃圾收集方案来提高闪存中垃圾收集的性能。有效页面的移动操作包括页面读取、数据传输、页面写入和元数据更新。如果源块和目标块在同一个平面上，数据传输操作可以由回拷贝操作代替。基于此，他们的方案决定了可能的先占点，并且只允许在这些点上先占，以最小化预占开销，因此被称为半先占。预处理允许为队列中的未决输入/输出请求提供服务，从而缩短响应时间。此外，如果传入的请求访问GC进程正在处理的同一页面，它可以被合并，如果它访问不同的页面，但与当前请求的类型相同(例如，读后读)，它可以被流水线化。他们已经表明，他们的垃圾收集方案对于严重突发和以写为主的工作负载特别有用。

王等[57]提出了一种基于闪存的I/O调度器。固态硬盘通过利用固态硬盘固有的并行性来提高性能。调度器推测性地将整个固态硬盘空间划分为许多子区域，并将每个子区域与专用调度子队列相关联。传入的请求被放入与其访问地址相对应的子队列中。这有助于同时执行请求，从而增强并行性。他们的调度程序还对同一子队列中的挂起请求进行排序，以创建顺序性并减少随机写入对性能和生存期的有害影响。

3.6 持久存储系统设计的编程模型和应用编程接口

非易失性存储器的非易失性使得能够设计能够经受程序和系统故障的持久存储系统[139]。已经提出了几种技术来实现和优化它。我们现在讨论其中的几个。

Volos等人[13]提出了尼莫西妮，一个与SCMs编程的接口。尼莫西妮允许应用程序声明静态变量，其值在系统重启时保持不变。它还允许应用程序从堆中分配内存，堆由持久内存支持。对永久存储器的写入顺序由软件使用非缓存写入模式、缓存线刷新指令和存储器屏障的组合来控制。此外，它提供了直接修改持久变量的原语，并通过轻量级事务机制支持一致更新。

Coburn等人[110]提出了NV堆，一种用于提供事务语义和持久性模型的持久性对象系统。为了确保应用程序一致性，NV堆支持应用程序分离易失性和非易失性数据。例如，应用程序可以确保不存在从持久内存到动态随机存取存储器空间的指针，因此，在重新启动后可以确保持久状态的一致性。此外，应用程序可以命名堆，使使用持久媒体编程变得更容易。尼莫西妮堆和非易失堆都有助于应用程序在为堆或堆栈分配新页面时区分动态随机存取存储器、持久内存和闪存。

Narayanan和Hodson[62]提出了一种用于所有系统主存储器都是非易失性的系统的持久性方法。他们的方法使用“故障时刷新”，这意味着只有在故障时(而不是在程序执行期间)，才使用系统电源提供的小剩余能量窗口将中央处理器寄存器和高速缓存行中保存的瞬态刷新到非易失性存储器。重新启动时，操作系统和应用程序状态会像透明检查点一样恢复，因此所有状态都会在故障后恢复。因此，他们的方法提供了挂起和恢复功能，还消除了每次更新时刷新缓存行的运行时开销。

Zhao等人[133]提出了一种持久存储器设计，其采用非易失性末级高速缓存和非易失性主存储器来构建持久存储器层次结构。在他们的方法中，当一个NV高速缓存行被更新时，它代表新更新的版本，而存储在NV存储器中的干净数据代表旧版本。当脏的NV高速缓存行被逐出时，NV存储器中的旧版本将被自动更新。通过这种多版本化的持久内存层次结构，他们的方法实现了持久的就地更新，而无需日志记录或写时复制。这使得具有持久支持的系统的性能接近于没有持久支持的系统的性能。他们还开发软件接口和架构扩展，为持久内存设计提供原子性和一致性支持。

Ransford等人[96]提出了一种支持在瞬时供电计算机上执行长时间运行任务的方法，例如在非常紧张的资源限制下工作的射频识别(RFID)设备。他们的技术将一个程序转化为

可中断的计算，以便它们可以分布在多个电源生命周期中。他们的技术不断监控能源的电压，当电压下降到阈值以下时，程序的易失性阶段被写入非易失性存储器，在那里它在断电后仍然存在。在下次引导时，检查点恢复例程将状态复制回 volatile 内存，以恢复执行。

Dong 等人[3]提出使用 PCM 来提高检查点性能。传统上，硬盘用于制造检查点，然而，检查点的低带宽导致了大的过度

检查点。此外，虽然动态随机存取存储器相对较快，

其高泄漏和易挥发的特性使其不适合检查点。虽然闪存是非易失性的，但其低写入耐久性限制了检查点频率。他们使用 3D PCM 架构进行检查点，这提供了高的整体输入/输出带宽。此外，对于大规模并行进程-

ing 系统，他们提出了一种混合的本地/全局检查点机制，其中除了全局检查点，本地还会创建定期备份状态的检查点。每个节点的私有内存。可以调整创建本地/全局检查点的频率，以实现性能和弹性之间的平衡，例如，具有高瞬时故障的系统可以创建频繁的本地检查点，而只创建很少的全局检查点。此外，如果在出现故障时，本地检查点本身丢失，系统的状态可以由全局检查点恢复。

Condit 等人[103]提出了一种事务文件系统，该系统利用 SCM 的字节寻址能力来减少更新期间写入的元数据量，并通过影子更新实现一致性。他们的文件系统保证文件系统写入在刷新缓存所需的时间内在持久存储介质上变得持久，并且每个文件系统操作都以原子方式和程序顺序执行。为了订购更新，他们使用划时代酒吧。高速缓存行用纪元号标记，并且高速缓存硬件被修改以确保由于写回引起的存储器更新总是以纪元顺序发生。

3.7 NVM 存储和内存的系统级重新设计

在系统级考虑非易失性存储器的属性有助于在存储器系统/层次结构中实现多种优化，如以下技术所示。

Jung 和 Cho [106] 提出了一种系统架构，命名为 Memorage，它利用系统的持久内存资源(即。永久主存储器和永久存储设备)。众所周知，由于存储资源的过度调配，其利用率仍然很低。为了解决这个问题，memory age 折叠了主存储器和存储资源之间的传统静态边界，并允许主存储器直接从存储设备借用可访问的存储器资源来应对存储器短缺。这避免了交换页面的需要。此外，由于存储设备的寿命远高于主存储器的寿命，系统寿命受到主存储器寿命的限制。为了解决这个问题，Memorage 允许存储设备将其空闲容量捐献给主存储器，从而提供“虚拟”过度配置，而不会产生物理过度配置的成本。这有助于延长主存储器的寿命，但代价是缩短存储寿命。

表 3
利用或比较多种存储技术的技术分类

分类参考	
多种内存技术的使用	
硬盘+固态硬盘[8], [39], [44], [59], [61], [65], [66], [83], [86], [87]	
PCM+闪存+硬盘[68]	
PCM+闪存[31], [75]	
动态随机存取存储器+相变存储器或非门闪光灯	[111]
DRAM+PCM [3], [99]	
DRAM+单片机[103], [112]	
动态随机存取存储器+闪存[69]	
多种存储技术的比较	
固态硬盘 v/s 硬盘[18], [20], [25], [39], [67], [97]	
动态随机存取存储器和硬盘	[27]
动力系统控制模块硬盘[3], [68]	
STT-RAM v/s 硬盘[134]	

Balakrishnan 等人[60]提出了一种提高固态硬盘阵列可靠性的技术。他们的技术基于这样的观察，即平衡固态硬盘阵列中的写入数量可能会导致相关的故障，因为它们可能会以相似的速率耗尽擦除周期。因此，阵列可能处于多个设备已达到其擦除极限的状态，因此需要用更新的设备替换所有设备。他们的技术在阵列中不均匀地分布奇偶校验块。由于随机访问模式，奇偶校验块的更新频率高于数据块，因此奇偶校验越多的设备接收的写入越多，因此老化越快。这在驱动器上造成了年龄差异，并降低了相关故障的概率。当旧设备被新设备替换时，他们的技术会重组奇偶校验分布。

4 关于组合 OR 的研究项目 比较多种内存技术

由于每种存储技术都有自己的优点和缺点，一些研究人员建议将多种技术结合起来，将它们中最好的结合起来，而其他人则提出了一项比较研究，以提供对所涉及的权衡的见解。表 3 总结了这些技术。我们现在讨论其中的几个。

4.1 闪存+硬盘

陈等人[8]提出了一种技术，利用硬盘的低成本和固态硬盘的高速度，将两者结合起来。他们的技术根据工作负载访问模式检测性能关键的数据块，并且只移动固态硬盘中最关键的数据块。此外，语义关键的数据块(例如文件系统元数据)优先保留在固态硬盘中，以提高性能。此外，为了提高写入密集型工作负载的性能，传入的写入会缓冲到低延迟固态硬盘中。

杨和任[44]提出了一种由硬盘和固态硬盘组成的存储设计，旨在利用两者的最佳特性。在他们的设计中，固态硬盘存储读取密集型参考

块和硬盘存储当前访问的块和相应的参考块之间的差值。他们还使用一种算法,该算法在输入/输出写入时计算增量,并在输入/输出读取时将增量与参考块相结合,以连接操作系统。因此,他们的方法旨在利用固态硬盘的李泽快速读取性能和硬盘的快速顺序写入性能,同时避免慢速固态硬盘写入,从而提高其性能和寿命。

Soundararajan 等人[59]提出了一种方法,通过将硬盘驱动器用作基于 MLC 的固态硬盘的永久写缓存来延长固态硬盘的寿命。他们的技术将所有写操作附加到存储在硬盘上的日志中,甚至将它们迁移到固态硬盘,最好是在后续读取之前。他们观察到,硬盘可以匹配中高端固态硬盘的顺序写入带宽。此外,典型的工作负载包含很大一部分块覆盖,因此,通过维护日志结构的硬盘驱动器,硬盘驱动器可以在其快速顺序写入模式下运行。同时,硬盘执行的写合并增加了固态硬盘观察到的工作负载的顺序性,也减少了对固态硬盘的写入,从而延长了固态硬盘的使用寿命。

对于虚拟内存管理,刘等人[65]建议将硬盘与固态硬盘集成,以克服闪存的有限耐用性问题,同时也限制了因交换以满足服务质量要求而导致的性能损失。他们的技术是将一组虚拟内存页顺序交换到硬盘上,如果它们被期望一起读取的话。此外,基于页面访问历史,他们的技术创建了跨硬盘和固态硬盘的内存不足虚拟内存页面布局。利用这一点,随机读取可以由固态硬盘提供,顺序读取由硬盘异步提供,因此两个设备的总带宽可以用来加速页面交换,同时也避免了它们各自的局限性。

Wu 和 Reddy [83]提出了一种利用固态硬盘和硬盘来提高混合存储系统性能的技术。他们的技术利用初始块分配和迁移来达到平衡状态,不同设备的响应时间相等。他们的技术跟踪不同设备的请求响应时间(即固态硬盘或硬盘),并执行分配以选择速度更快的设备,同时在设备上实现工作负载平衡。此外,他们的技术可以检测数据块是否是冷/热的,冷数据仅在后台迁移。Kim 等人[39]提出了一种混合硬盘-固态硬盘设计,该设计利用了这两种介质的互补特性,在给定的成本预算下提供高性能和服务差异化。他们的技术使用固态硬盘和硬盘性能的统计模型来做出动态请求分区决定。由于随机写入会导致固态硬盘碎片化,并增加固态硬盘的垃圾收集开销和延迟,因此他们的技术会定期将一些页面从固态硬盘迁移到硬盘,以便部分写入可以重定向到硬盘,从而降低随机性。他们还开发了一个模型来为给定的工作负载找到最经济的硬盘/固态硬盘配置使用混合整数线性规划(ILP)。

4.2 闪存+动力系统控制模块

孙等[31]提出使用 PCM 作为 NAND 闪存存储系统的日志区。动力系统控制模块日志区域

支持就地更新,避免需要过时的日志记录。他们的方法减少了对闪存的读取和擦除操作,从而提高了闪存存储的寿命、性能和能效。由于 PCM 的字节可寻址特性,读取操作的性能也得到了提高。他们还提出了一些技术来确保 PCM 日志区域在闪存之前不会磨损。

4.3 PCM+闪存+硬盘

Kim 等人[68]评估了 PCM 在存储体系中的潜力,考虑了其成本和性能。他们观察到,尽管基于材料级特性,写入闪存比写入 PCM 慢,但基于系统级特性,由于功率限制等原因,写入基于闪存的固态硬盘可能比基于 PCM 的固态硬盘快。基于这一认识,他们研究了两个存储使用案例:分层和缓存。对于分层,他们对由闪存、硬盘和 PCM 组成的存储系统进行建模,以确定在成本限制内提供最佳性能的设备类型组合。他们观察到,相变材料可以提高分层存储系统的性能,某些组合(例如 30%相变材料+ 67%闪存+ 3%硬盘的组合)可以提供比没有相变材料的组合更高的性价比。对于缓存,他们比较了 PCM 和闪存的总输入/输出时间和读取延迟。他们观察到,闪存和 PCM 固态硬盘的组合可以提供比仅闪存配置更好的读取延迟和聚合输入/输出时间。

4.4 动态随机存取存储器+相变存储器

Qureshi 等人[99]提出了一种混合主存储器系统,其中动态随机存取存储器被用作相变存储器的“页面高速缓存”,以结合动态随机存取存储器的延迟优势和相变存储器的容量优势。他们提出了几种策略来减轻 PCM 存储器的写开销,并增加其寿命。出现页面错误时,从硬盘取出的页面只写入动态随机存取存储器。只有当从动态随机存取存储器中取出该页面并标记为脏时,该页面才会写入动力系统控制模块。此外,在高速缓存行的粒度上跟踪对页面的写入,并且仅回写页面中修改的行,以减少对主存储器的有效写入次数。此外,为了实现磨损均衡,每页中的行以旋转的方式存储在 PCM 中。

4.5 动态随机存取存储器+单片机

Bailey 等人[112]为应用程序提供了一个带有标准 get/put 接口的持久化、版本化的键值存储。他们的方法将永久数据保存在单片机中,并在单片机上使用动态随机存取存储器作为薄层,以解决其性能和损耗限制。线程维护和操作动态随机存取存储器中的本地易失性数据,只有通信的持久状态被写入单片机,从而避免了 SCM 的慢写瓶颈。SCM 的字节可寻址特性用于非易失性存储器中的细粒度事务,快照隔离用于支持并发性、一致性、可恢复性和版本控制,例如,在断电时,以前提交给 SCM 的数据可以在恢复时访问,并将处于一致状态。

Payer 等人[87]提出了一种在内存层次结构的同一级别上利用低成本、高容量硬盘和高成本、低容量固态硬盘的技术。他们的技术允许使用低性能固态硬盘或高性能固态硬盘。对于低性能固态硬盘，可执行文件和程序库被移动到固态硬盘，剩余文件被移动到硬盘；此外，随机访问的文件被移动到固态硬盘，具有混合或连续访问模式的文件被移动到硬盘。当使用吞吐量接近硬盘的高性能固态硬盘时，最常用的文件将被移动到固态硬盘，其余文件将被移动到硬盘。

4.6 固态硬盘

Narayanan 等人[25]在 2009 年比较了硬盘和基于闪存的固态硬盘的性能、能耗和成本。对于一系列数据中心工作负载，他们分析了固态硬盘对磁盘的完全替换，以及固态硬盘作为磁盘和动态随机存取存储器之间中间层的使用。他们观察到，由于每美元固态硬盘的容量较低，用固态硬盘替换磁盘并不是他们工作负载的最佳解决方案。他们发现固态硬盘的容量/成本需要提高 3 到 3,000 倍，才能与磁盘媲美。

Albrecht 等人[97]在 2013 年进行了类似的研究，发现由于闪存价格下降以及磁盘性能相对稳定，盈亏平衡点一直在转移，闪存现在在更大范围的工作负载中变得更加经济。

4.7 动态随机存取存储器和硬盘

考尔菲尔德等人[27]比较了几种存储器技术，如动态随机存取存储器、非易失性存储器和硬盘驱动器。他们衡量这些技术对输入/输出密集型、数据库和内存密集型应用程序的影响，这些应用程序具有不同的延迟、带宽要求和访问模式。他们还研究了将内存技术连接到主机系统的不同选项的效果。他们观察到，非易失性存储器在应用级和原始输入/输出性能方面都有很大的提升。对于某些应用，相变材料和 STT-RAM 可以在性能上比硬盘提高一个数量级。

5 未来的挑战和结论

在一个数据爆炸的时代，对存储系统的存储和处理需求正在急剧增加。虽然很明显，传统的内存技术远远不能满足未来计算系统的需求，但目前的非易失性存储器技术也无法满足未来系统的性能、能效和可靠性要求。我们相信，要应对这些挑战，需要在从设备级到系统级的多个抽象层上对非易失性存储器进行有效管理。

例如，在设备层面，3D 设计可以带来更密集的外形、更小的占地面积和更低的延迟[21]。此外，制造更大容量的供应链管理原型及其大规模生产将推动对它们的进一步研究。虽然 MLC NVM 比 SLC NVM 提供了更高的密度和更低的价格，但它的续航能力一般是两个-

比 SLC NVM 低三个数量级，比如 70 nm SLC 闪存的续航时间在 100 K 周期左右，2 位 2x nm MLC 闪存的续航时间在 3 K 周期左右，而对于 3 位 MLC 闪存，这个数值只有几百个周期[100]。由于对存储器容量的日益增长的需求可能需要使用多层非易失性存储器，因此需要有效的软件方案来减轻非易失性存储器的过写。此外，非易失性存储器通常被认为不受辐射引起的软错误的影响，然而，由于热噪声[138]或由于电阻漂移[120]存储在多层电容器相变存储器单元中的值随时间的变化，随机位反转可能在非易失性存储器中出现软错误。在不久的将来，除了性能和能量，研究人员还需要研究诸如非易失性存储器的软错误恢复能力等问题。

在架构级别，写最小化、损耗均衡和容错技术的协同集成将有助于实现 NVM 存储系统生命周期的数量级改进。在系统级，通过硬件/软件合作在单个地址空间中统一内存/存储管理可以进一步最小化两级存储模型的开销。由于从毫瓦级手持系统到兆瓦级数据中心和超级计算机等各种规模和形状的计算系统都依赖于高效的存储系统，并呈现出不同的约束和优化目标，因此考虑它们的独特特性对于设计针对不同平台和用途优化的技术来说将是极其重要的。最后，这些技术还需要具有很高的成本竞争力，以证明它们在商品市场上的使用是合理的。

在这篇文章中，我们提出了一个技术概况，利用非易失性存储器的主存储器和存储系统。我们还讨论了组合或比较多种内存技术的技术，以研究它们的相对优点，并将它们的优点结合在一起。我们根据几个关键参数对这些技术进行了分类，以突出它们的相似性和差异性，并确定主要的研究趋势。希望这项调查将激发新的想法，在未来的计算系统中充分利用非易失性存储器的潜力。

感谢

这项工作得到了美国能源部科学办公室高级科学计算研究所的支持。

参考

- [1] (2015). [在线]. 可用: www.youtube.com/yt/press/statistics.html
- [2] (2013). [在线]. 可用: <http://goo.gl/qwyFHe>
- [3] X. 董, N. Muralimanohar, N. Jouppi, R. Kaufmann 和 Y. 谢, “利用 3D PCRAM 技术降低未来数据库云服务器系统的检查点开销”, 载于 Proc. 高性能表单. 电脑. Netw., 储存肛门., 2009 年, 第 57:1 - 57:12 页。
- [4] B. 吉里达尔、西斯拉克、杜加尔、德累斯顿斯基、陈, R. 帕蒂, b. 霍尔德里, c. 查克拉巴蒂, t. 穆奇和 d. 布劳, “探索动态随机存取存储器组织以获得高能效和高响应性的超大规模存储器”, 载于《程序设计》。里面的糖膏剂高性能。电脑., Netw., 储存肛门., 2013 年, 第 23:1 - 23:12 页。
- [5] 南米塔尔和维特尔, “分析和提高图形处理器能效方法的调查”, ACM Comput. 《调查》, 第 47 卷, 第 2 期, 第 19:1-19:23 页, 2015 年。
- [6] J. 维特尔和米塔尔, “非易失性存储器系统在超大规模高性能计算中的机遇”, Comput. Sci. 英格., 第 17 卷, 第 2 期, 第 73-82 页, 2015 年 3-4 月。

- [7] 南米塔尔, 维特尔和李, “管理嵌入式动态随机存取存储器和非易失性片上高速缓存的体系结构方法综述”, IEEE Trans. 并行发行版. 系统., 第 26 卷, 第 6 期, 第 1524-1537 页, 2014 年 6 月。
- [8] F. 陈, 库法蒂和张, “海斯托尔:在高性能存储系统中充分利用固态硬盘”, 载于 Proc. 里面的糖膏剂超级计算., 2011 年, 第 22-32 页。
- [9] (2012). 闪存驱动器取代了脸书亚马逊的磁盘。[在线]。可用:www.wired.com/2012/06/flash-数据中心/
- [10] (2011). 认识一下戈登, 世界上第一台闪存超级计算机。[在线]。可用:www.wired.com/2011/12/gordon-super-计算机/
- [11] F. 舒 • € • rmann, F. Delalondre, P. S. Kumbhar, J. Biddiscombe, 米 (meter 的缩写)) 吉拉、塔奇拉、库里奥尼、梅兹勒、莫尔扬、芬克斯、弗朗切斯基尼、杰曼、施奈德巴赫、芬德和费奇, “为科学计算重定输入/输出基准:在 IBM BlueGene/Q 超级计算机中利用存储类内存”, 载于 Proc. 里面的糖膏剂超级计算., 2014 年, 第 331-347 页。
- [12] (2014). 橡树岭将获得下一代超级计算机。[在线]。可用:http://goo.gl/d315UD
- [13] H. 《尼莫西妮:轻量级持久记忆》。阿奇特. 新闻, 第 39 卷, 第 1 期, 2011 年, 第 91-104 页。
- [14] 南陈, 吉本斯和纳特, “对相变存储器数据库算法的再思考”, 载于《过程》。第五届双年度会议。创新数据系统。决议, 2011 年, 第 21-31 页。
- [15] 页 (page 的缩写) 帕万, 贝兹, 奥利佛和扎诺尼, “闪存单元概述”, 程序. IEEE, 第 85 卷, 第 8 期, 第 1248-1271 页, 1997 年 8 月。
- [16] 南米塔尔, “相变存储器电源管理技术调查”, 国际. J. Comput. 辅助工程. 技术., 2014.
- [17] 米 (meter 的缩写)) Kryder 和 C. Kim, “硬盘之后——接下来会发生什么?” IEEE Trans. 马格尼., 第 45 卷, 第 10 期, 第 3406-3413 页, 2009 年 10 月。
- [18] F. 陈, 库法蒂和张, “理解基于闪存的固态驱动器的内在特性和系统含义”, ACM SIGMETRICS Perform. 评价. 修订版, 第 37 卷, 第 1 期, 第 181-192 页, 2009 年。
- [19] 名词 (noun 的缩写) 阿格拉瓦尔、普拉巴卡兰、沃博、戴维斯、马纳塞和帕尼格拉希, “固态硬盘性能的设计权衡”, 载于《程序设计》。USENIX • 安努. 技术. 糖膏剂, 2008 年, 第 57-70 页。
- [20] 米 (meter 的缩写)) 桑维多, 朱福荣, 卡尔尼和塞林格, “与非闪存及其在存储架构中的作用”, Proc. IEEE, 第 96 卷, 第 11 期, 第 1864-1874 页, 2008 年 11 月。
- [21] 米 (meter 的缩写)) Poremba, S. Mittal, D. Li, J. S. Vetter, 和 Y. Xie, “destination:一种用于建模新兴 3D NVM 和 eDRAM 缓存的工具”, 载于 Proc. 德斯. 汽车. 测试 Eur., 2015 年, 第 1543-1546 页。
- [22] A. 金古普塔和乌尔高恩卡, “DFTL:采用基于需求的页面级地址映射选择性缓存的闪存转换层”, 载于 Proc. ACM Int. 糖膏剂阿奇特. 支持专业程序. 郎. 操作系统., 2009 年, 第 229-240 页。
- [23] 长度-张炳良和黄立中, “块映射固态盘的低成本损耗均衡算法”, 《SIGPLAN 通知》, 第 46 卷, 第 5 期, 第 31-40 页, 2011 年。
- [24] 南-李伟文和穆恩, “基于闪存的数据库管理系统的设计:页内记录方法”, 载于 Proc. ACM SIGMOD Int. 糖膏剂管理. 数据, 2007 年, 第 55-66 页。
- [25] D. Narayanan, E. Thereska, A. Donnelly, S. Elnikety 和 A. Rowstron, “将服务器存储迁移到固态硬盘:权衡分析”, 在 Proc 中. 欧元. 糖膏剂电脑. 系统., 2009 年, 第 145-158 页。
- [26] J. 罗, 可汗, 赵, 谢, 和穆特鲁, “一个有效的存储和内存的硬件/软件协同管理的例子”, 载于 Proc. 车间节能设计., 2013.
- [27] A. m. 考尔菲尔德, j. 科伯恩, t. 莫洛夫, a. 德, a. 阿克, j. 何, a. 贾加塞森, R. K. 古普塔, a. 斯纳弗利和 s. 斯旺森, “理解新兴非易失性存储器对高性能、IO 密集型计算的影响”, 载于《Proc》。里面的糖膏剂高性能. 电脑., Netw., 储存肛门., 2010 年, 第 1-11 页。
- [28] R. 法肯塔尔、北川、大冢、普拉尔、米尔斯、筑地、贾瓦尼法德、特德罗、筑地、施巴哈拉和 G. 嘘, “一个 16 Gb 的随机存取存储器, 200 兆字节/秒的写入和 1gb/秒的读取在 27 纳米技术,” 在 Proc. IEEE Int. 固态电路会议., 2014 年, 第 338-339 页。
- [29] 米 (meter 的缩写)) Wu 和 W. Zwaenepoel, “eNVy:一个非易失性的主存储器存储系统”, ACM SigPlan notifications, 第 29 卷, 第 11 期, 第 86-97 页, 1994 年。

- [30] 长度——张炳良, “关于大规模闪存存储系统的有效损耗均衡”, 载于 Proc. ACM Symp. 应用计算., 2007 年, 第 1126-1130 页。
- [31] G. 孙, 朱有洲, 陈德牛, 谢有洲, 陈有洲和李海峰, “一种用于性能, 能耗和寿命改进的混合固态存储体系结构”, 载于 Proc. 里面的辛普. 高性能. 电脑. 阿奇特., 2010 年, 第 1-12 页。
- [32] 南闪存输入输出架构的重构设计. 阿奇特. 列特., 2014, Doi: 10.1109/ LCA. 2014. 2329423.
- [33] Y. 蔡, 穆特鲁, 哈拉奇, 昂萨尔, 克里斯塔尔, 麦, 等, “多层存储与非闪存的相邻单元辅助纠错”, 载于《计算机程序设计》。里面的糖膏剂 Meas. 建模组件. 系统., 2014 年, 第 491-504 页。
- [34] Y. 蔡, 哈拉奇, 穆特鲁和麦, “多层与非闪存中的错误模式: 测量, 表征和分析”, 载于 Proc. 德斯., Autom. 测试 Eur., 2012 年, 第 521-526 页。
- [35] Y. 蔡, 穆特鲁, 哈拉奇和麦, “多层与非快闪存储器中的程序干扰: 表征, 建模和缓解”, 载于《程序设计》。里面的糖膏剂电脑. 德斯., 2013 年, 第 123-130 页。
- [36] F. 陈, 罗, 张, “CAFTL: 一种内容感知的闪存转换层, 可延长基于闪存的固态硬盘的使用寿命”, 载于 Proc. USENIX 会议. 文件存储技术., 第 11 卷, 2011 年。
- [37] F. 陈, 李若望, 张, “开发基于闪存的固态硬盘内部并行性在高速数据处理中的重要作用”, 载于 Proc. 里面的辛普. 高性能. 电脑. 阿奇特., 2011 年, 第 266-277 页。
- [38] B. 凯利, j. 怀尔斯, d. 迈耶, k. 贾米森, k. 弗雷泽, t. 迪根, D. 斯托登、列斐伏尔、费斯塔和沃菲尔德, “分层: 虚拟化非易失性存储器上的可扩展高性能存储”, 载于《过程》。USENIX 会议. 文件存储技术., 2014 年, 第 17-31 页。
- [39] Y. 金、古普塔、乌尔高恩卡、伯曼和西瓦苏布拉马-尼亚姆, “混合存储: 一种结合固态硬盘和硬盘的高性价比高性能存储系统”, 载于 Proc. 里面的辛普. 模特, 肛门. 模拟. 电脑. 远程办公. 系统., 2011 年, 第 227-236 页。
- [40] 南基于闪存的固态硬盘能效和性能的综合研究. 阿奇特., 第 57 卷, 第 4 期, 第 354-365 页, 2011 年。
- [41] 米 (meter 的缩写)) 郑, 崔, 沙勒夫和坎德米尔, “三重阵列: 一种用于高性能存储系统的基于非固态硬盘的自主全闪存阵列”, 载于《程序设计》。第 19 个 Int. 糖膏剂阿奇特. 支持专业程序. 郎. 操作系统., 2014 年, 第 441-454 页。
- [42] 南周, 朴, 吴, 金, 易, 和甘格, “主动磁盘遇到闪存: 智能固态硬盘的一个案例”, 在 Proc. 里面的糖膏剂超级计算., 2013 年, 第 91-102 页。
- [43] 页 (page 的缩写) 黄, 万, 周, 黄, 李, 王, “提高固态硬盘的有效容量和寿命”, 载于《计算机科学与技术》杂志. 里面的糖膏剂 Netw. 阿奇特. 存储, 2013 年, 第 50-59 页。
- [44] 杨和任军, “智能现金: 固态硬盘和硬盘的智能耦合阵列”, 载于 Proc. 里面的辛普. 高性能. 电脑. 阿奇特., 2011 年, 第 278-289 页。
- [45] Y. 王, 巴森, 杜特, 邵振中, “元治愈: 一种用于 NAND 闪存存储系统中元数据的可靠性增强策略”, 载于 Proc. 德斯. 汽车. 糖膏剂, 2012 年, 第 214-219 页。
- [46] C. 王和王文福, “声表面波: 系统辅助的与非门闪存器件写入耐久性的损耗均衡”, 载于 Proc. 德斯. 汽车. 糖膏剂, 2013 年, 第 1-9 页。
- [47] Y. ——张学良、谢俊伟、郭铁伟, “闪存存储系统的耐久性增强: 一种有效的静态磨损均衡设计”, 载于《程序设计》。德斯. 汽车. 糖膏剂, 2007 年, 第 212-217 页。
- [48] C. 王和王文福, “通过挽救坏块来延长 nand 闪存的寿命”, 载于 Proc. 糖膏剂德斯., Autom. 测试 Eur., 2012 年, 第 260-263 页。
- [49] C. 王和王文福, “观察性磨损均衡: 一种有效的闪存管理算法”, 载于《过程》。德斯. 汽车. 糖膏剂, 2012 年, 第 235-242 页。
- [50] J. 廖福政, 李立林, 肖光国, “闪存中的自适应损耗均衡”, Comput. 阿奇特. 列特., 2014, Doi: 10.1109/ LCA. 2014. 2329871.
- [51] R. 数据依赖备用来管理好于坏的数据块. 阿奇特. 列特., 第 12 卷, 第 2 期, 第 43-46 页, 2013 年。

- [52] 长度 M. Grupp, A. M. Caulfield, J. Coburn, S. Swanson, E. Yaakobi, 页 (page 的缩写) 西格尔和沃尔夫, “表征闪存:异常、观察和应用”, 载于《程序》。里面的辛普。微芯片。 , 2009 年, 第 24-33 页。
- [53] 南 Boboila 和 P. Desnoyers, “闪存驱动器中的写入耐久性:测量和分析”, 载于 Proc. USENIX 会议。文件存储技术。 , 第 10 卷, 2010 年。
- [54] 长度 m. 格鲁普, J. D. 戴维斯和 s. 斯旺森, “与非闪存暗淡的未来”, 载于 Proc. USENIX 会议。文件存储技术。 , 2012。
- [55] Y. 张, Arpaci-Dusseau, R. H. Arpaci-Dusseau, “基于闪存的固态硬盘的去间接性与无名写入”, 载于 Proc. USENIX 会议。文件存储技术。 , 2012。
- [56] R.-刘少生、杨振林和吴文伟, “通过保留松弛优化基于 NAND 闪存的固态硬盘”, 载于《Proc》。USENIX 会议。文件存储技术。 , 2012。
- [57] H. 王, 黄, 何, 周, 李, 何, “一种新的固态硬盘 I/O 调度器, 提高了固态硬盘的性能和寿命”。辛普。海量存储系统。技术。 , 2013 年, 第 1-5 页。
- [58] J. 李彦宏, 李彦宏, 希普曼, 口头, 王飞飞和金俊杰, “固态硬盘的半抢占式垃圾收集器”, 在 Proc. 里面的辛普。表演。肛门。系统。软。 , 2011 年, 第 12-21 页。
- [59] G. Soundararajan, V. Prabhakaran, M. Balakrishnan 和 t. 沃泊, “使用基于磁盘的写缓存来延长固态硬盘的寿命”, 载于 Proc. USENIX 会议。文件存储技术。 , 2010 年, 第 10 卷, 第 101-114 页。
- [60] 米 (meter 的缩写) 巴拉克里希南, a. 卡达夫, v. 普拉巴卡兰和 d. 马尔基, “差分 RAID:重新思考固态硬盘可靠性的 RAID”, ACM Trans.《存储》, 第 6 卷, 第 2 期, 第 4 页, 2010 年。
- [61] J. Matthews, S. Trika, D. Hensgen, R. Coulson 和 K. Grimsrud, “英特尔睿频存储器:主流计算机系统存储层次结构中的非易失性磁盘高速缓存”, ACM Trans. 存储, 第 4 卷, 第 2 期, 第 4:1-4:24 页, 2008 年。
- [62] D. 纳拉亚南和霍德森, “全系统持久性”, 在 Proc. ACM SIGARCH Comput. 阿奇特。新闻, 第 40 卷, 第 1 期, 第 401-410 页, 2012 年。
- [63] C. 计算机固态硬盘的性能是带宽、并发性、设备结构和系统组织的函数。阿奇特。新闻, 第 37 卷, 第 3 期, 第 279-289 页, 2009 年。
- [64] H. Kim 和 S. Ahn, “BPLRU:一种改进闪存随机写入的缓冲区管理方案”, 载于 Proc. USENIX 会议。文件存储技术。 , 2008 年, 第 8 卷, 第 1-14 页。
- [65] K. 刘, 张, 戴维斯, 江, “固态硬盘与硬盘协同耦合实现服务质量感知虚拟内存”, 载于《过程》。里面的辛普。表演。肛门。系统。软。 , 2013 年, 第 24-33 页。
- [66] 科尔兹达斯和维格拉斯, “刷新存储层” VLDB 基金会, 第 1 卷, 第 1 期, 第 514-525 页, 2008 年。
- [67] 南 Park 和 K. Shen, “基于闪存的固态硬盘上科学 I/O 工作负载的性能评估”, 载于 Proc. IEEE Int. 糖膏剂集群计算。研讨会, 2009 年, 第 1-5 页。
- [68] H. “评估企业存储系统的相变存储器:缓存和分层方法的研究”, 载于《程序设计》。USENIX 会议。文件存储技术。 , 2014 年, 第 33-45 页。
- [69] A. Badam 和 V. S. Pai, “固态硬盘:混合固态硬盘/随机存取存储器内存管理变得容易了”, 载于 Proc. USENIX 会议。Netw. 系统。德斯。实施, 2011 年, 第 16-16 页。
- [70] Y. 哦, 蔡志勇, 李德全和诺思, “减少缓存以获得更好的性能:平衡混合存储系统中的缓存大小和闪存缓存的更新成本”, 载于 Proc. USENIX 会议。文件存储技术。 , 第 12 卷, 2012 年。
- [71] T. 普里切特和 m. 托特索迪, “SieveStore:一种高选择性、集成级的磁盘高速缓存, 以实现高性价比”, 载于 Proc. 里面的辛普。电脑。阿奇特。 , 2010 年, 第 163-174 页。
- [72] 长度史, 邱光国, 赵美明, 薛俊杰, “错误模型引导的闪存联合性能和耐久性优化”, IEEE 转译。电脑。-援助德斯。集成电路系统。 , 第 33 卷, 第 3 期, 第 343-355 页, 2014 年 3 月。
- [73] J. 杨, N. Plasson, G. Gillis, N. Talagala, S. Sundararaman, 和 R. 伍德, “HEC:提高基于闪存的高性能缓存设备的耐用性”, 载于 Proc. 里面的系统。Storage Conf. , 2013 年, 第 10 页。
- [74] K. 赵, 文卡塔拉曼, 张, 李, 郑, 张, “超时钟固态硬盘:

超越闪存芯片输入/输出时钟规格的安全运行”, 载于《中国科学》杂志。里面的辛普。高性能。电脑。阿奇特。 , 2014 年, 第 536-545 页。

- [75] 长度史, 李俊杰, 薛春生, 周, “用于高能效和高性能系统的混合非易失性磁盘缓存”, ACM Trans. 德斯. 汽车. 电子. 系统., 第 18 卷, 第 1 期, 第 8:1-8:23 页, 2013 年。
- [76] 米 (meter 的缩写)) 《闪存层:一种轻量级、一致且持久的存储缓存》, 载于 Proc. 欧元. 糖膏剂电脑. 系统., 2012 年, 第 267-280 页。
- [77] R., 马尔莫尔, 兰加斯瓦米, 孙达拉拉曼, 塔拉加拉, 赵, “主机端闪存缓存的写入策略”, 载于 Proc. USENIX 会议. 文件存储技术., 2013 年, 第 45-58 页。
- [78] R.——刘少林、杨振林、李振华、陈光毅, “DuraCache:一种使用 MLC NAND 闪存的耐用 SSD 缓存”, 载于 Proc. 德斯. 汽车. 糖膏剂, 2013 年, 第 166 页。
- [79] 页 (page 的缩写) 黄, 苏伯迪, 何世和, 周, “FlexECC:部分放松多层固态硬盘的 ECC 以获得更好的缓存性能”, 载于 Proc. USENIX • 安努. 技术. 糖膏剂, 2014 年, 第 489-500 页。
- [80] Y. 潘光东, 吴清泉, 张天祥, “准非易失性固态硬盘:利用闪存的非易失性来提高企业应用的存储系统性能”, 载于《中国科学》杂志. 里面的辛普. 高性能. 电脑. 阿奇特., 2012 年, 第 1-10 页。
- [81] Y. 蔡., 穆特鲁, 哈拉奇, 安 • 克里斯塔尔, 安塞尔和麦, “闪存纠正和刷新:提高闪存寿命的保留感知错误管理”, 载于《程序设计》. 里面的糖膏剂电脑. 德斯., 2012 年, 第 94-101 页。
- [82] 长度史, 薛, 胡, 曾文成, 周, 沙, “通过智能牺牲高速缓存减少闪存上写活动”, 载于《计算机程序设计》杂志. 五大湖交响乐团., 2010 年, 第 91-94 页。
- [83] X. 吴和雷迪, “在混合存储系统中利用并发来提高延迟和吞吐量”, 载于 Proc. 里面的辛普. 模特, 肛门. 模拟. 电脑. 远程办公. 系统., 2010 年, 第 14-23 页。
- [84] 动词 (verb 的缩写) 帕皮拉和查克拉巴蒂, “闪存的能量感知错误控制编码”, 在 Proc. 德斯. 汽车. 糖膏剂, 2009 年, 第 658-663 页。
- [85] A. 桑普森、纳尔逊、施特劳斯和塞泽, “固态存储器中的近似存储”, 载于《程序》. 里面的辛普. 微芯片., 2013 年, 第 25-36 页。
- [86] H. 基于与非门闪存的混合存储的最佳节能技术. 消费者. 电子., 第 58 卷, 第 3 期, 第 841-848 页, 2012 年 8 月。
- [87] H. 帕耶, 桑维多, 班迪奇和基尔希, “组合驱动:在异构存储设备中优化成本和性能”, 在 Proc. 将固态存储器集成到存储层次结构的研讨会, 2009 年, 第 1 卷, 第 1-8 页。
- [88] 南一种用于闪存存储系统的过程感知热/冷识别方案. 消费者. 电子., 第 56 卷, 第 2 期, 第 339-347 页, 2010 年 5 月。
- [89] C. 王, 马, 孟, 金, 恩格尔曼, “将固态硬盘聚合存储作为内存分区暴露在超大规模计算机中”, 载于 Proc. 里面的并行发行版. 流程. 辛普., 2012 年, 第 957-968 页。
- [90] 米 (meter 的缩写)) Gamell, I. Roderio, M. Parashar 和 S. Poole, “在具有深度内存层次的系统上探索数据密集型科学工作流的能量和性能行为”, 载于 Proc. 里面的糖膏剂高性能. 电脑., 2013 年, 第 226-235 页。
- [91] T. 罗伯特和穆奇, “改进基于与非门闪存的磁盘高速缓存”, 在 Proc. 里面的辛普. 电脑. 阿奇特., 2008 年, 第 327-338 页。
- [92] 吴和张, “WAWR:使用按需快速写入和重写来缩短 SSD 响应时间”, IEEE Trans. 电脑., 第 63 卷, 第 10 期, 第 2500-2512 页, 2014 年 10 月。
- [93] J. 郑先生、韩先生、李先生、金先生、郑先生、韩先生、李先生
J. Kim, “使用动态编程和擦除扩展的基于 NAND 闪存的存储系统的寿命改进”, 载于 Proc. 第 12 届 USENIX 会议. 文件存储技术., 2014 年, 第 61-74 页。
- [94] G. 吴和何, “通过 NAND 闪存编程和擦除暂停来减少 SSD 读取延迟”, 载于 Proc. FAST, 2012 年。
- [95] B. 范 • 埃森, 谢海明, s. 艾姆斯, r. 皮尔斯和 m. 戈卡莱, “DI-MMAP—一个用于核心外数据密集型应用的可扩展内存映射运行时”, 集群计算., 第 18 卷, 第 1 期, 第 15-28 页, 2013 年。
- [96] B. 《纪念品:对射频识别规模设备上长时间运行计算的系统支持》, 载于 Proc. 阿奇特. 支持程序. 郎. 操作系统., 2011 年, 第 159-170 页。
- [97] C. 《商人》, 斯托克利, 瓦利吉, 拉贝尔,
名词 (noun 的缩写) 史晓龙和施勒克, “骏利:云存储工作负载的最优闪存配置”, 载于 Proc. USENIX • 安努. 技
- 术. 糖膏剂, 2013 年, 第 91-102 页。

- [98] 名词 (noun 的缩写) 刘, 柯普, 卡恩斯, 卡罗斯, 格里德, A. Crume 和 C. Maltzahn, “论突发缓冲区在领导级存储系统中的作用”, 载于 Proc. 辛普. 海量存储系统. 技术., 2012 年, 第 1-11 页。
- [99] 米 (meter 的缩写) 使用相变存储技术的可扩展高性能主存储系统。里面的辛普. 电脑. 建筑师., 2009 年, 第 24-33 页。
- [100] 南李金堂, 金金凯, 金俊杰, “使用恢复感知动态节流对基于闪存的固态硬盘进行生命周期管理”, 载于 Proc. FAST, 2012 年。
- [101] D. 李、维特尔、马林、麦考迪、西拉、刘和 W. 于, “确定字节可寻址非易失性存储器在极端规模的科学应用中的机会”, 载于 Proc. 里面的并行发行版. 流程. 辛普., 2012 年, 第 945-956 页。
- [102] R. 刘树生、沈德年、杨振林、俞树声、王振宇, “NVM 二重奏: 统一工作存储器和持久存储体系结构”, 载于 Proc. 阿奇特. 支持程序. 郎. 操作系统., 2014 年, 第 455-470 页。
- [103] J. 康迪特、南丁格尔、弗罗斯特、伊佩克、李、伯格和库切, “通过字节可寻址的持久存储器实现更好的输入输出”, 载于《程序设计》. ACM SIGOPS 第 22 届交响乐. 操作系统. 《原则》, 2009 年, 第 133-146 页。
- [104] E. Giles, K. Doshi 和 P. Varman, “使用 WrAP 弥合持久性和易失性存储器之间的编程鸿沟”, 载于 Proc. 里面的糖膏剂电脑. 前沿, 2013 年, 第 30 页。
- [105] E. 李, 柳, 张和汉, “快捷方式-JFS: 一种用于相变存储器的高效写日志文件系统”, 载于《程序设计》. 辛普. 海量存储系统. 技术., 2012 年, 第 1-6 页。
- [106] J. 荣格和赵树理, “记忆: 新兴的基于持久内存的可延展主内存和存储架构”, 载于《过程》. 里面的糖膏剂超级计算., 2013 年, 第 115-126 页。
- [107] A. 阿克, 考尔菲尔德, 莫洛夫, 古普塔和斯旺森, “缟玛瑙: 一种原型相变存储阵列”, 载于《程序设计》. 第三届 USENIX 会议. 热门话题存储文件系统., 2011 年, 第 2 页。
- [108] Z. 刘, 王, 李, 维捷尔, 余, “基于相变存储器的快速磁盘输入输出持久写缓存”, 载于《计算机科学与技术》杂志. 里面的辛普. 模特, 肛门. 模拟. 电脑. 远程办公. 系统., 2012 年, 第 451-458 页。
- [109] A. 考尔菲尔德、德、科伯恩、莫洛、古普塔 南斯旺森, “莫奈塔: 面向下一代非易失性存储器的高性能存储阵列架构”, 载于 Proc. 里面的辛普. 微芯片., 2010 年, 第 385-395 页。
- [110] J. 科伯恩, 考尔菲尔德, 阿克, 格鲁普, 古普塔, R. Jhala 和 S. Swanson, “NV-Heaps: 利用下一代非易失性存储器使持久对象变得快速和安全”, 载于 Proc. ACM SIGARCH Comput. 阿奇特. 新闻, 第 39 卷, 第 1 期, 2011 年, 第 105-118 页。
- [111] J.C. Mogul, E. Argollo, M. A. Shah 和 P. Faraboschi, “对 NVM+ DRAM 混合主存储器的操作系统支持”, 载于 Proc. 第 12 次 Conf. 热门话题操作系统., 2009 年, 第 14 页。
- [112] K. 贝利、霍恩亚克、塞泽、格里布尔和利维, “探索具有键值存储的存储类内存”, 载于《过程》. 车间互动非挥发性物质/闪存操作系统. 《工作量》, 2013 年, 第 4 页。
- [113] 南陈佩利和温氏, “记忆的持久性”, 载于《程序》杂志. 里面的辛普. 电脑. 阿奇特., 2014 年, 第 265-276 页。
- [114] 南 Kannan, A. Gavrilovska 和 K. Schwan, “降低最终用户设备中非易失性堆的持久性成本”, 载于 Proc. 里面的辛普. 高性能. 电脑. 阿奇特., 2014 年, 第 512-523 页。
- [115] 南 Kannan, A. Gavrilovska, K. Schwan 和 D. Milojicic, “使用 NVM 作为虚拟内存优化检查点”, 载于 Proc. 里面的辛普. 并行发行版. 流程., 2013 年, 第 29-40 页。
- [116] E. 伊佩克, j. 康迪特, E. B. 南丁格尔, d. 伯格和 t. 莫斯比罗达, “动态复制存储器: 从纳米级电阻存储器构建可靠的系统”, 载于 Proc. 阿奇特. 支持程序. 郎. 操作系统., 2010 年, 第 3-14 页。
- [117] 南谢切特, 罗赫, 斯特劳斯和伯格, “使用 ECP, 而不是 ECC, 电阻性存储器中的硬故障,” 在 Proc. 里面的辛普. 电脑. 阿奇特., 2010 年, 第 141-152 页。
- [118] 名词 (noun 的缩写) 吴海成、斯里尼瓦桑、里弗斯和李海光, “更安全: 记忆的故障错误恢复”, 载于《程序设计》. 里面的辛普. 微芯片., 2010 年, 第 115-124 页。
- [119] D. 尹洪浩, 穆利马诺哈尔, 张俊杰, 朱皮平, 和埃雷兹, “保护非易失性存储器免受硬错误和软错误的影响”, 载于《程序设计》. 里面的辛普. 高性能表单. 电脑. 阿奇特., 2011 年, 第 466-477 页。

- [120] 米 (meter 的缩写)) 阿瓦斯蒂, m. 谢夫戈尔, k. 苏丹, b. 拉金德兰, r. 巴拉苏布拉蒙尼安和 v. 斯里尼瓦桑, “易错新兴记忆的高效擦洗机制”, 载于 Proc. 里面的辛普。高性能。电脑。阿奇特。 , 2012 年, 第 1-12 页。
- [121] J. 陈, 文卡塔拉玛尼, 黄, 等, “重复使用: 将有故障的模块重新循环以延长使用寿命”, 载于 Proc. 里面的糖膏剂可靠的系统。Netw。 , 2012 年, 第 1-12 页。
- [122] T. 高、斯特劳斯、布莱克本、麦金利、伯格和拉鲁斯, “使用托管运行时系统容忍可穿戴存储器中的漏洞”, 载于《程序设计》。糖膏剂程序员。郎。德斯。实施, 2013 年, 第 297-308 页。
- [123] Y. 方, 李海峰, 李晓峰, “软相变存储器: 通过近似写入提高视频应用中相变存储器的能效和寿命”, 载于 Proc. IEEE 亚洲测试标准。 , 2012 年, 第 131-136 页。
- [124] 南基于相变存储器的存储设备磨损水平的差异化空间分配。消费者。电子。 , 第 60 卷, 第 1 期, 第 45-51 页, 2014 年 2 月。
- [125] 莫拉鲁、安德森、卡明斯基、托利亚、阮冈纳赞和宾克特, “字节可寻址非易失性主存储器的一致、持久和安全的内存管理”, 载于《程序设计》。第一届 ACM SIGOPS Conf. 及时结果操作系统。 , 2013 年, 第 1:1-1:17 页。
- [126] 南非易失性字节可寻址存储器的一致和持久的数据结构。USENIX 会议。文件存储技术。 , 2011 年, 第 61-75 页。
- [127] X. 吴和雷迪, “SCMFS: 存储类内存的文件系统”, 载于 Proc. 里面的糖膏剂高性能。电脑。 , Netw。 , 储存肛门。 , 2011 年, 第 39 页。
- [128] 南 Sardashti 和 D. Wood, “UniFI: 利用非易失性存储器实现统一的容错和空闲功率管理技术”, 载于 Proc. 里面的糖膏剂超级计算。 , 2012 年, 第 59-68 页。
- [129] F. Aouda, K. Marquet 和 G. Salagnac, “瞬时供电系统的程序状态到 NVRAM 的增量检查点”, 载于 Proc. 里面的辛普。可重构通信。-Centric Syst. -片上, 2014 年, 第 1-4 页。
- [130] 南 R. Dulloor, S. Kumar, A. Keshavamurthy, P. Lantz, D. Reddy, R. 桑卡兰和杰克逊, “持久存储的系统软件”, 在 Proc. 欧元。糖膏剂电脑。系统。 , 2014 年, 第 15:1-15:15 页。
- [131] 米 (meter 的缩写)) 库雷希, 弗朗西斯基尼, 贾格莫汉和拉斯特斯, “预置: 通过利用写时间的不对称性来改善相变存储器的性能”, 载于《程序设计》。里面的辛普。电脑。阿奇特。 , 2012 年, 第 380-391 页。
- [132] D. 查克拉巴蒂, 博姆和班达里, “阿特拉斯: 为非易失性存储器一致性撬锁”, 载于《程序设计》。里面的糖膏剂面向对象程序设计。系统。郎。Appl, 2014, 第 433-452 页。
- [133] J. 赵, 李树生, 尹德华, 谢永年, 和朱新平, “窑: 在有和没有持续支持的系统之间缩小性能差距”, 在 Proc. 里面的辛普。微芯片。 , 2013 年, 第 421-432 页。
- [134] H. 金, 安, 刘, 蔡, 韩, “非易失性存储器的内存文件系统”, 载于《程序设计》。自适应收敛系统。 , 2013 年, 第 479-484 页。
- [135] 南坎南, 加夫里洛夫斯卡, 施万, 米洛吉契奇和塔尔瓦, “使用主动非易失性随机存取存储器进行输入输出分段”, 载于《程序设计》。《Petascale 数据分析: 挑战与机遇》, 2011 年, 第 15-22 页。
- [136] Z. 王, 易, 刘若兰, 董, 陈, “持续性跨行为记忆”, 计算机。阿奇特。列特。 , 2014, Doi: 10.1109/LCA.2014.2329832。
- [137] A. 潘德, V. Ahuja, R. Sivaraj, E. Baik 和 P. Mohapatra, “4G 网络中视频传送的挑战和机遇”, IEEE 多媒体, 第 20 卷, 第 3 期, 第 88-94 页, 2013 年 7-8 月。
- [138] 南米塔尔和维特尔, “计算系统可靠性建模和改进技术综述”, IEEE Trans. 并行发行版。系统。 , 2015, Doi: 10.1109/TPDS.2015.2426179。
- [139] J. Kubiawicz, D. Bindel, Y. Chen, S. Czerwinski, P. Eaton, R. Gummadi, S. Rhea, H. Weatherspoon, W. Weimer, C. Wells, 和 B. Zhao, “OceanStore: 一种全球规模持久存储的体系结构”, 《ACM Sigplan 通知》, 第 35 卷, 第 11 期, 第 190-201 页, 2000 年。



斯巴塞米塔尔在印度 IIT 获得电子和通信工程学士学位，在爱荷华州立大学获得计算机工程博士学位。他目前在 ORNL 大学担任博士后研究助理。他的研究兴趣包括非易失性存储器、存储系统能效、高速缓存和 GPU 架构。他是 IEEE 的成员。



杰弗里·维特尔获得了佐治亚理工学院的博士学位。他是 ORNL 和 GT 的联合任命。在 ORNL，他是一名杰出的 R&D 员工，也是未来科技集团的创始人。在 GT，他是计算科学与工程学院的联合教授，美国国家科学基金会 (NSF) 跟踪 2D 实验计算设施的项目主管，该设施使用图形处理器进行大规模异构计算，并且

英伟达 CUDA 卓越中心主任。他的研究兴趣包括大规模多线程处理器、非易失性存储器和异构多核处理器。他是 IEEE 的高级成员。