

请参见本出版物的讨论、统计和作者简介，网址为：<https://www.researchgate.net/publication/319529366>

Emerging NVM: A Survey on Architectural Integration and Research Challenges

《美国计算机学会电子系统设计自动化学报》2018年1月

DOI: 10.1145/3131848

引文

24

读

3,211

4名作者，包括：



Jalil Boukhobza

Université de Bretagne Occidentale

117 PUBLISHES 384 引文



Stéphane Rubini

Université de Bretagne Occidentale

63 PUBLISHES 267 引文



本出版物的一些作者也在从事这些相关项目：



项目切达，一个实时调度分析工具 [View project](#)



莫斐斯欧洲项目 [View project](#)

本页之后的所有内容都是由上传的 [Jalil Boukhobza](#) 2017年12月19日。

用户请求增强下载的文件。

新兴的非虚拟制造: 建筑集成与研究挑战综述

UMR 布列塔尼大学理工学院，邮编 6285；香港理工大学理学院理学院理学院理学院理学院
理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院
理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院
理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院
理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院
理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院理学院
邵子力，香港理工大学

近年来,人们对非易失性存储器的兴趣激增。凭借密度和功耗等诸多优势, NVM 正在内存层级中开辟一席之地,并可能最终改变我们对计算机架构的看法。出现了许多非易失性存储器,如磁阻随机存取存储器(MRAM)、相变随机存取存储器(PCM)、电阻随机存取存储器(ReRAM)和铁电随机存取存储器(FeRAM),每种存储器都有其独特的特性和特定的挑战。科学界在将这些技术集成到内存层级结构中方面做了大量的工作。随着许多公司宣布即将将大规模生产 NVM, 我们认为是时候后退一步, 讨论与 NVM 集成相关的大量文献了。本文综述了将非易失性存储器集成到存储器层次结构中的最新工作。特别地, 我们介绍了四种类型的非易失性存储器, 即 MRAM、PCM、ReRAM 和 FeRAM, 并从水平或垂直的角度研究了将它们集成到存储器层次结构中的不同方法。这里, 水平集成意味着新存储器被放置在与现有存储器相同的层, 而垂直集成意味着新存储器被交错在两个现有层之间。此外, 我们描述了每种 NVM 技术的挑战和机遇。

• 信息系统~存储级内存硬件~非易失性内存硬件~内存和密集存储软件及其工程~内存管理

附加关键词和短语:非易失性存储器、MRAM、PCM、FeRAM、ReRAM、存储器、主存储器。

自动呼叫管理参考格式:

贾利勒·布霍布扎,斯特凡·鲁比尼,陈仁海,2017.新兴的NVM:关于架构集成和研究挑战的调查. ACM Trans. 嵌入. 电脑. 系统. 第9、4、39条(2017年3月), 36页。
DOI: <http://dx.doi.org/10.1145/0000000.0000000>

1. 介绍

容量、性能、能效和可扩展性是内存层次结构的设计者必须处理的关键因素，以便满足当前数据密集型应用程序不断增长的需求。

作者地址:法国布列斯特勒戈戈大道 20 号, 西布列塔尼大学 UMR 分校, 邮编:6285; 邵先生就职于香港九龙红磡香港理工大学计算学系嵌入式系统及 CPS 实验室; 陈, 天津大学计算机科学与技术学院认知计算与应用天津重点实验室; 电子邮件: {boukhobza, rubini} @ univ-Brest . fr; cszlshao@comp.polyu.edu.hk; crh6250790@gmail.com.

允许免费制作部分或全部作品的数字或硬拷贝供个人或课堂使用，前提是拷贝的制作或分发不是为了盈利或商业利益，并且拷贝在第一页或显示屏的初始屏幕上显示本通知以及完整的引文。必须尊重除 ACM 之外的其他人拥有的本作品组件的版权。允许带学分摘要。以其他方式复制、重新发布、在服务器上发布、重新发布到列表或在其他作品中使用本作品的任何部分需要事先获得特定许可和/或支付费用。可向美国纽约州纽约市佩恩广场 2 号 701 套房 ACM 公司出版部申请许可，传真：(1) 212 869-0481，或 permissions@acm.org。

2010 年 ACM 1539-9087/2010/03-ART39 15.00 美元

DOI:<http://dx.doi.org/10.1145/0000000.0000000>

音量。如今，以不同形式产生的数字数据量呈指数级增长[阮冈纳赞, 2011; 2008 年冬季]。对于晶体管集成, 摩尔定律给出了一个令人印象深刻的增加过程的例子, 产生的数据量以更快的速度增长。在最近的一项研究中, EMC 预测从 2009 年到 2020 年, 每年创建的数字数据量将增长 44 倍。这主要是由于产生数据的设备(以及设备中的传感器)数量的增长——迄今为止, 大约有 110 亿台设备连接到互联网。事实上, 随着社交网络、视频回放和事务处理的日益普及, 应用程序变得越来越数据密集型, 需要大量内存来存储和处理这些活动生成的数据。

性能。高效处理生成的数据已经成为一个重大的经济和社会问题。更好的数据分析意味着更好的结果、过程和决策。它可以帮助努力产生新的想法和解决方案, 或者尝试更准确地预测未来事件(天气预报、疾病传播等)。已经付出了很多努力来为这些大量数据的在线和离线操作提供更多的处理能力[Mutlu 2015]。然而, 从性能的角度来看, 这给内存层次结构带来了很大的压力, 以弥补与处理元素之间的历史差距。

能量。除了性能和容量之外, 能效也是设计内存系统时需要考虑的关键指标。事实上, 计划于 2020 年推出的数据库云服务器计算系统最具挑战性的技术创新之一是功耗的管理和优化[Kogge 等人, 2008]。美国的数据中心消耗了该国 1.5% 以上的能源, 预计这一比例将每年增加 18% [Zhang 等人, 2010]。这种能量的很大一部分来自记忆系统。存储系统的能耗估计占典型数据中心总能耗的 20% 到 40% 之间[Carter 和 Rajamani, 2010]。其他研究给出了不同的数字, 例如, 动态随机存取存储器 (DRAM) 主存储器子系统消耗 HPC (高性能计算) 节点的 30% 到 50%, 尽管这个比例取决于 DRAM 的容量和配置[维特尔和米塔尔 2015]。

可扩展性。DRAM 技术在尺寸、容量和性能方面都经历了令人印象深刻的改进; 然而, 许多研究预测, DRAM 的扩展趋势将在未来 5 到 10 年内达到平稳。动态随机存取存储器单元需要足够大才能进行可靠的感测, 但这与特征尺寸缩小的趋势背道而驰。因此, 高密度内存带来了指数级的成本损失(例如, 在 2008 年, 使用 1 个内存 8GB 成本为 212 美元/GB, 而 2*4GB 内存成本为 50 美元/GB, 4*2GB 内存成本为 15 美元/GB [Mogul 等人, 2009])。动态随机存取存储器单元也需要定期刷新, 因此即使不执行读/写操作也要消耗功率; 这种功耗已被证明是峰值功率的 19% 至 31% 之间[Mogul 等人, 2009 年]。与尺寸非常小的特征的干扰问题和可靠性相关的其他问题给未来的动态随机存取存储器设计带来了很大的困难[Yoongu 等人, 2014; 维特尔和米塔尔 2015]。因此, 增加基于电荷的技术(如 DRAM) 以及闪存的规模变得越来越困难[Mutlu 2015]。

在这种背景下, 新兴的非易失性固态存储器有望彻底改变存储器的层次结构。它们可以提供非常高的密度、几乎为零的静态功耗和高耐用性。许多新兴的非易失性存储器技术

正在被广泛研究, 如相变存储器(称为 PRAM 或 PCM)、电阻存储器(ReRAM)、自旋扭矩转移存储器(STT-RAM)和铁电存储器(FeRAM)。几家公司已经宣布即将大规模生产这种技术。

最受欢迎的老牌 NVM 是 NAND 闪存。它利用了 21 世纪初智能手机使用的激增, 现在在许多应用领域无处不在, 如嵌入式系统和高性能计算。它已经成为部分弥合动态随机存取存储器和存储性能之间差距的不可或缺的技术。

就像十年前的闪存一样, 非易失性存储器正吸引着人们的极大兴趣, 关于如何将不同的技术集成到存储器层次结构中的问题, 人们正在进行大量的工作。寻求大规模生产 nvm 的不同公司发布的众多公告证明, 有必要后退一步, 讨论和分类在最先进的工作中已经研究过的集成选项。

这项调查有三个目的: 1) 阐明女企业家的特点; 2) 根据架构集成选项对最先进的作品进行分类——更准确地说, 是横向或纵向集成。水平集成意味着将 NVM 放在与现有存储器技术相同的级别(例如, 在具有 SRAM 的高速缓存中), 而垂直集成意味着新存储器在两个现有级别之间交错(例如, 在 DRAM 和传统存储设备之间); 3) 讨论准备使用这些非易失性存储器所涉及的主要挑战。

最近还有其他关于新兴非女性职业经理人的调查。在 [于和陈 2016] 中, 作者对不同类型的 NVM 进行了主要的技术和低级描述。其他一些非常有趣的调查, 如 [Mittal 等人, 2015b] 的调查, 更多地是关于内存层次结构的特定级别(在这种情况下, 关于缓存), 或者关于所采用的软件优化 [Mittal 和 Vetter, 2016]。其他调查, 如 [夏等人, 2015 年] 的调查是详尽的, 与一项技术(此处为 PCM)相关。与之前的调查不同, 我们的调查侧重于四种 NVM 技术的架构集成问题: PCM、MRAM、FeRAM 和 ReRAM1。

论文的其余部分组织如下: 第 2 节奠定了讨论非易失性存储器集成的基础。第 3、4、5 和 6 节分别讨论了 PCM、MRAM、FeRAM 和 ReRAM 技术。在每一节中, 我们定义了其中的一种技术, 并讨论了它的集成选项。我们将突出一些有代表性的研究工作中讨论的主要问题。

2. 背景

在本节中, 我们首先介绍内存层次结构, 并讨论集成选项。然后, 我们介绍闪存集成的例子, 它被认为是最后一种主要的破坏性存储技术。最后, 我们将介绍 NVM 属性和集成选项。

1 请注意, 在本次调查中, 不会调查 Domain Wall Memory (DWM), 我们建议读者参考 [Mittal 2016] 中有趣的调查

2.1 内存层次结构概述

存储器是计算机系统的关键组成部分，随着时间的推移，它们的功能和与系统的集成都有了很大的发展。实际上，已经开发了大量的设备和技术。

现代计算机的设计引入了内存分层管理管理的第一个部门：支持数据高速移动的快速内存，它是运行应用程序不可或缺的，被称为主内存或主内存；和用于大容量存储的存储器。后者提供了前者，这得益于较低成本的数据存储属性，即所谓的二级内存或存储内存。这种内存的硬件分类已经转化为操作系统级别的不同管理策略，因为不同的服务管理这些内存。实际上，特定的服务管理主内存，并在不同的应用程序进程之间共享该资源，而存储系统作为外围设备管理，更准确地说，作为块设备管理。

给定内存技术在内存层次金字塔中的位置(请参见 Fig. 1)反映到主处理单元的距离。它在带宽和持久性方面的性能越好，它相对于处理单元的架构越接近；它的性能越差，它相对于处理单元的架构就越远。

从体系结构的角度来看，人们可能会考虑仅根据其在技术进程、带宽和耐用性方面的特征，将给定的非易失性存储器技术插入存储器层级金字塔中的任何位置。从数量的角度来看，整合将取决于一些成本考虑。一般来说，给定内存技术的性能越大，成本越高，越接近集成的 CPU。例如，集成的一个例子是缓存级别。静态随机存取存储器是离中央处理器更近的存储器。

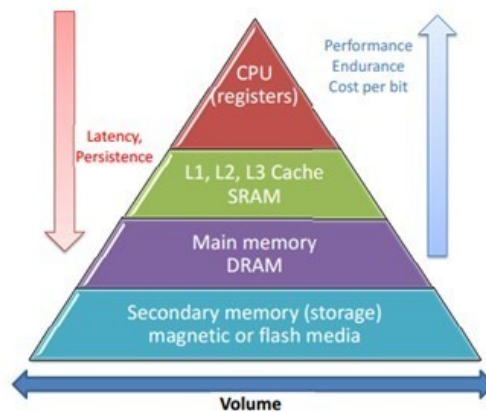


图 1。记忆等级金字塔。

然而，从应用程序和操作系统的角度来看，情况是不同的。例如，从操作系统的角度来看，即使计算机中集成了许多内存组件，但主要还是以下三个组件：

- 1) 高速缓存系统, 可以激活或停用, 但不提供明确的管理, 因为中央处理器高速缓存由特定的硬件组件管理。
- 2) 主存储器, 它是字节可寻址的, 在操作系统中作为具有专用服务的资源来管理。关于后者, 系统依赖一些硬件组件来加速其管理(例如, TLB 的翻译后备缓冲器)。
- 3) 最后是存储系统, 无论是传统磁盘还是基于闪存的存储驱动器(固态驱动器或固态硬盘)。从操作系统的角度来看, 存储设备是作为外围设备管理的块寻址设备。

从架构的角度来看, 可以执行 NVM 的集成

垂直或水平。

垂直集成意味着集成的非易失性存储器将在存储器层次结构中向下移动给定的存储器技术并取代它。在这种情况下, 使用的访问方法与替换的内存技术相同。例如, 在动态随机存取存储器被替换的情况下, 非易失性存储器以与动态随机存取存储器相同的方式被访问(从应用层)。

水平集成意味着在层次结构中补充给定的内存技术[维特尔和米塔尔 2015]。在水平集成中, 新集成的非易失性存储器的接口方式与同级现有存储器相同。例如, 在主内存中进行水平集成的情况下, 需要一些特定的应用程序编程接口, 以便应用程序可以利用内存异构性。一些标准化组织如 SNIA 已经开始在整个“非易失性存储器编程模型(NPM)”[SNIA 2015]中研究这个问题。该文档声称这种新的体系结构需要一些特定的应用编程接口, 但是也支持标准的 POSIX 应用编程接口。这意味着一个系统需要提供两个 API, 一个允许程序员知道 NVM 的存在, 而另一个(标准 POSIX)隐藏了内存异构性。

无论是水平集成还是垂直集成, 出于性能优化的原因, 硬件(控制器)和/或软件(操作系统和/或应用程序)组件都必须知道新的内存堆栈。

为了完全集成 nvm, 必须考虑硬件集成和软件集成。前者回答了关于横向或纵向集成以及 NVM 应该如何与系统其余部分接口的问题, 而后者则与系统和应用程序升级有关, 以充分利用额外的内存技术。

2.2 闪存, NVM 的先驱

闪存一直是一项颠覆性技术, 因为它允许人们讨论操作系统上的标准存储集成, 即块输入/输出层[bjrling 等人, 2013; 欧阳等[2011]]。

闪存显示高密度, 空闲时耗电少, 访问时间比硬盘快, 尤其是随机读取。但是要有效地集成到给定的存储系统中, 必须考虑其特性: 读/写性能不对称、有限的寿命、写/擦除操作粒度不对称等等。

2.1.1. 闪存集成。闪存可以通过三种方式集成到存储服务器中[Roberts 等人, 2009; Boukhobza 2013]: 1) 作为内存系统的扩展(水平集成), 2) 作为闪存所在的存储加速器

作为存储频繁访问的代码和数据的缓存进行管理,从而减少硬盘的输入/输出数量(垂直集成),或者 3)作为替代存储设备,闪存固态硬盘被用作硬盘的替代或补充(再次水平集成)。在以上三个选项中,有两个是水平集成,但被放置在两个不同的级别:主内存或辅助内存。从应用的角度来看,在闪存与 DRAM 集成的情况下,集成对于应用是透明的,因为闪存是通过专用硬件来管理的,以补充 DRAM(例如, NVDIMM[黄和蒋 2014; 陈等[2016]]。在水平集成的第二种情况下,闪存通过传统的存储系统接口展示给系统的其余部分,因此可以像硬盘驱动器一样通过标准文件系统接口进行访问。闪存还通过 PCIe 接口进行了集成,通过利用多通道和其他多芯片的内部并行性以及平面并行性,为性能提升提供了更大的空间。通过 PCIe 的这种集成可以是存储堆栈兼容的,因此被视为存储设备,也可以是不兼容的,因此需要一些特定的驱动程序和软件工具供应用程序访问。后者的目标是通过专门的软件消除传统的存储堆栈延迟。

出于成本考虑(闪存大约比硬盘贵一个数量级),基于闪存的存储系统在中期内不会取代磁盘系统[Gurumurthi 2009]。因此,磁盘存储系统不会从内存层次结构中消失。

2.1.2. 闪存属性和管理。与传统的磁存储系统相比,与非门闪存具有一些独特的特性,这些特性促使科学界为它们的集成(水平和垂直)设计特定的解决方案。以下是取自[Boukhobza 2013; Boukhobza 和 Olivier 2017]:

- 随机读取:闪存完全是基于电子的组件,没有移动的机械部件。与硬盘相反,读取访问延迟不再取决于数据的物理位置,因此随机读取的性能与顺序读取相同。与传统硬盘相比,这使得闪存的读取操作更加高效和可预测。硬盘在随机访问的情况下性能不佳,导致在操作系统级别实施了许多机制来尽量缩小这种性能差距:输入/输出调度器、页面缓存等。在用户层,许多应用程序也假设有一个底层硬盘,因此试图最大化顺序请求,同时最小化随机请求,如数据库管理系统(如排序算法)和大数据应用程序(如 MapReduce)。在闪存集成的情况下,重新审视了这些机制,以充分利用闪存性能,科学界为此做出了重大贡献。
- 输入/输出性能不对称:闪存的另一个特点是,与硬盘不同,读取通常比写入快得多。这主要是由于基本闪存单元的电特性:编程该单元以达到稳定状态比简单地读取它需要更多的时间。许多闪存专用的缓冲系统机制被设计用来吸收写操作,从而提高性能。

- 块中的顺序写入: 为了避免写入错误(由于某些电干扰属性), 必须在一个块中顺序执行写入, 尤其是对于 MLC(多级单元) 技术。纠错码(ECC)在硬件中实现以应对这种错误。
- 低功耗: 即使基于闪存的存储系统的功耗取决于其内部架构和应用的工作负载[Park 等人, 2011], 它仍然比硬盘更节能。随着高性能计算路线图整合了越来越多的能效限制, 这样的存储系统技术将比以往任何时候都更受追捧。
- I/O 接口: 为了便于闪存的采用, SSD 作为替代技术的集成依赖于使用类似于硬盘的 I/O 接口。从用户的角度来看, 这个特性可以被认为是一个优点, 因为不需要修改操作系统(驱动程序除外)。然而, 这可能会给人一种错觉, 认为人们必须像处理硬盘一样处理闪存, 这远非事实。

- 耐冲击性: 这一特性是缺少机械零件的又一后果。

在设计基于闪存的解决方案时, 必须处理一组给定的约束。

这些制约因素可归纳如下:

- 写入前擦除限制: 这是闪存最关键的限制之一。不能直接覆盖数据。该约束意味着, 如果给定页面的数据需要就地修改, 则应该首先擦除整个块(参见前面的约束)。这是很费时间的; 因此, 数据通常通过映射方案被不适当地更新(在旧数据无效之后)。这种映射方案对固态硬盘的性能有很大影响。除了映射方案之外, 闪存控制器还必须实现垃圾收集器来回收以前无效的数据。
- 写入/擦除粒度不对称: 写入在页面上执行, 而擦除操作在块上实现。闪存块由给定数量的页面组成, 大小在 2 到 8kb(2 的幂)之间。
- 擦除/写入周期数有限: 对于单级单元(SLC), 平均写入/擦除周期数约为 105, 对于多级单元(多层单元)约为 104, 对于三级单元约为 5000。达到最大擦除周期数后, 给定的存储单元可能不再保留数据。由于对耐久性的这种限制, 闪存控制器实施磨损均衡技术来平衡闪存块的磨损。

因此, 为了与传统存储接口(就像硬盘驱动器一样)集成, 闪存控制器应该至少实现上述三种服务: 映射方案、垃圾收集器和磨损均衡器。这些机制对性能有很大的影响。

2.3 NVM, 为什么和如何?

从体系结构和系统的角度来看, 可以在内存的三个主要子类之一的内存层次结构中插入一个新的 NVM: 处理器缓存、主内存或存储系统。nvm 可以垂直或水平集成。

表 1. 根据最先进的研究，nvm 的特征 2[维特尔和米塔尔 2015；米塔尔等人。2015b 夏等 2015；王等 2014aSuresh 等人，2014 年；Baek 等人，2013 年；Meena 等人，2014 年]

	静态随机存取存储器	动态随机存取存储器	硬盘驱动器	“与非”闪存	STT-RAM	ReRAM	脉冲编码调制	FeRAM
单元大小 (F2)	120-200	60-100	不适用的	4-6	6-50	4-10	4-12	6-40
写耐力	1016	>1015	> 1015 (pb: 机械零件)	104-105	1012-1015	108-1011	108-109	1014-1015
读取延迟	~ 0.2-2 纳秒	~10ns	3-5 毫秒	15-35 岁	2-35ns	~10ns	20-60 纳秒	20-80 纳秒
写入延迟	~ 0.2-2 纳秒	~10ns	3-5 毫秒	200-500 s	3-50 纳秒	~50ns	20-150 纳秒	50-75 纳秒
泄漏功率	高的	中等	(机械零件)	低的	低的	低的	低的	低的
动态能源 (拆装)	低的	中等	(机械零件)	低的	低/高	低/高	中/高	低/高
成熟度	成熟的	成熟的	成熟的	成熟的	测试芯片	测试芯片	测试芯片	制造

2.2.1. 作为一个存储系统。从存储系统的角度来看，NVM 可以横向集成。因此，类似于闪存，应该实现一些约束管理机制并将其抽象到更高的层次，以便新的基于 NVM 的存储系统可以集成到传统的存储软件堆栈中。NVM 也可以垂直集成，例如基于 PCIe 的闪存设备。另一种可能性是让一些混合设备(如硬盘)集成闪存，但这次是闪存以外的非易失性存储器技术。

2.2.2. 作为主要记忆。由于许多非易失性存储器技术具有非常吸引人的性能特征(参见 Table 1)，它们可以集成到内存层次结构中比存储系统更高的级别。非易失性存储器具有一些特殊的特征，例如读取性能高于写入性能，耐久性根据技术而变化，并且受写入操作的影响比受读取操作的影响更大，以及大多数非易失性存储器在被访问时仅使用电能的事实。当集成非易失性存储器作为动态随机存取存储器或在动态随机存取存储器旁边以具有一些混合主存储器时，必须处理所有这些特征，并且必须包括一些新的服务。

同样，NVM 可以与 DRAM 垂直集成，也可以水平集成[李等，2012]。在第一种情况下，动态随机存取存储器可以被视为非易失性存储器的高速缓存，以减少对主存储器的访问延迟。NVM 也可以与内存水平集成在同一总线上。数据放置问题可以由硬件管理，从而抽象到操作系统，或者由操作系统管理，抽象到应用程序。最后，操作系统可以为应用程序提供一些接口，以便于决定应用层的位置。在这种情况下，在处理内存异构时，操作系统支持是一个关键问题。

2.2.3. 作为处理器缓存。从 CPU 缓存的角度来看，一级缓存一般访问频率较高，因此需要非常低的延迟和高的耐用性。这对于大多数国家职业妇女来说很难实现，即使其中一些人可以被认为是这种融合的候选人。另一方面，末级高速缓存的设计更多是为了减少片外数据移动，因此必须具有高密度(以实现大容量)。集成在中央处理器缓存中的非易失性存储器更多地用于末级缓存，而不是其他级别的缓存[Mittal 等人，2015b]。然而，即使在这种情况下，nvm 也会吸收大量的写操作，因此问题

2 价值观不是严格的价值观，而是对特征的总体看法。

必须减轻磨损。此外，该技术与制造工艺的兼容性也是一个非常重要的问题。

Table 2 总结了关于存储器层次结构的 NVM 集成的最先进的工作。如前所述，NVM 可以与其他技术水平集成；在这种情况下，无论是在硬件、操作系统还是应用级别，都需要提供一些数据迁移机制。NVM 可以通过向下移动另一个内存来垂直集成，也可以替代现有的内存技术。在 Table 2 中，在处理器高速缓存的情况下，“替换”行列出了所有级别的高速缓存都被 NVM 替换的研究。

表 2. 非易失性存储器集成的最新研究分类

		MRAM	ReRAM	脉冲编码调制	FeRAM
缓存(不同级别)	水平的	[obolil 等人, 2015 年; 李等 2012; Syu 等人, 2013 年; 李等 2014; 吴等 2009; Jadidi 等人, 2011 年; 李等 2011; 王等人, 2014 年; Komalan 等人 2014; 程等 2016]	[王等 2014aKomalan 等人, 2013 年; 米塔尔和维特尔 2015]	[吴等 2009; Joo 等人, 2010 年]	
	垂直的	[obolil 等人, 2015 年; 尼森等人, 2014 年; 孙等人, 2009 年; 吴等 2009; Samavatian 等人, 2014 年; Smullen 等人, 2011 年; Jog 等人, 2012 年; 周等 2009bGoswami 等人, 2013 年; Yazdanshenas 等人, 2014 年; Ahn 等人, 2012 年; Rasquinha 等人, 2010 年; Park 等人, 2012 年; 陈等 2013; Kwon 等人, 2014 年; Jokar 等人, 2016 年; 尼森等人, 2015 年; 程等 2016]	[董等 2013; 王等 2013; Jokar 等人, 2016 年]	[吴等 2009]	
	更换	[obolil 等人, 2015 年; Smullen 等人, 2011 年; 孙等人, 2011 年; 郭等, 2010; Goswami 等人 2013; 王等 2015]	[董等 2013]		
主存储器	水平的	[杨等 2013; Suresh 等人, 2014 年; 魏等[2015]	[Hassan 等人, 2015 年; 魏等[2015]	[Dhiman 等人, 2009 年; Park 等人, 2010 年; Bock 等人, 2011 年; 苏雷什等人 2014; 周等。 2009a 孙等人, 2015 年; 魏等, 2015; 李等人 2014; 萨尔科德和 Asadi 2016 魏等人。 2015; Kannan 等人 2016; Dulloor 等人 2016; 吴等 2016; 李等 2012; Oikawa 2014; 高等人, 2015 年]	[Doh 等人, 2007 年; 苏雷什等人 2014]
	垂直的	[苏雷什等人, 2014 年]		[库雷希和斯里尼瓦桑, 2009 年; Suresh 等人, 2014 年; Awad 等人, 2016 年; 吴等人, 2016 年]	[Suresh 等人, 2014 年; Jung 等人 2010]
	更换	[Kultursay 等人, 2013 年; 王等 2014; 金等 2014]	[徐等 2013; 徐等 2015]	[Lee 等人, 2009 年; 陈等 2012; Park 等人, 2015 年]	[Baek 等人, 2013 年]
仓库	水平的	[Lee 等人, 2014 年]	[Tanakamaru 等人, 2014 年; 孙等人, 2014 年; 藤井等人, 2012 年]	[孙等 2010; 考尔菲尔德等人, 2010 年; Park 等人, 2010 年]	[Yoon 等人, 2008 年]
	垂直的	[康等 2015]		[刘等 2011; 康等 2015]	
	更换	[Lee 等人, 2014 年]	[Jung 等人, 2013 年]	[Akel 等人, 2011 年; Kim 等人, 2014 年]	[Baek 等人, 2013 年]

应该注意的是 Table 2 研究最多的选项是处理器缓存中 MRAM 的垂直集成和主内存中 PCM 的水平集成，这两项技术是最先进的研究工作。请注意，对于许多与相变材料(或相变材料)相关的研究，所提出的机制也适用于相变材料(或相变材料)。

在本文的其余部分，我们将研究以下每一种非易失性存储器：PCM、MRAM、FeRAM 和 ReRAM，以及它们的集成。

3. 脉冲编码调制

正如我们在 Table 1, 相变存储器 (PCM)，也称为相变随机存取存储器 (PRAM)，在 CMOS 制造工艺中具有小尺寸单元和优异的可扩展性，相对快速的随机存取，适度的写入吞吐量，良好的保持时间，与 NAND 闪存相比非常好的耐久性，以及无擦除编程(不需要擦除操作来更新数据)[Qureshi 等人, 2011]。

3.1 基本概念

3.1.1 PCM 技术。PCM 使用硫属化合物合金来实现存储单元。一个 PCM 单元通常使用一个薄的硫属化物层，如 $\text{Ge}_2\text{Sb}_2\text{Te}_5$ (GST) [薛等, 2011] 和两个在两侧包裹硫属化物的电极，此外还有一个加热器(见 Fig. 2)。PCM 是一种电阻式非易失性存储器，利用不同的电阻来表示位信息。

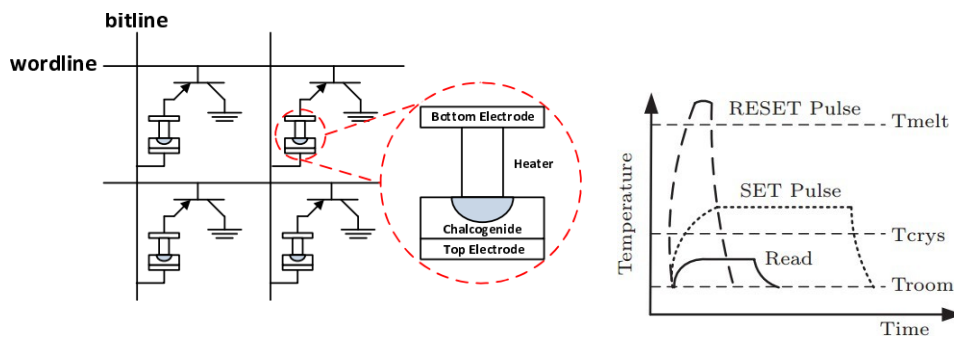


图 2。PCM 单元阵列[薛等, 2011]]和 RESET/SET 温度[夏等, 2015]

这种硫属化合物材料经受由电引发的快速非晶到晶的相变过程。

相位取决于电流脉冲产生的热量。短而高的电压脉冲导致非晶态(复位，位 0)，此时 GST 被加热到熔化温度以上 (T_{melt} in Fig. 2). 另一方面，低电压的长脉冲导致结晶状态(置位，位 1)。在这里，商品及服务税被加热到结晶温度 (T_{crys}) 以上，但低于熔化温度(参见 Fig. 2). 因此，写性能由更长的操作即 SET 决定。一旦建立了一个阶段，就可以在不干扰这个阶段的情况下读取它；材料在置位和复位阶段的电阻之比在 102 和 104 之间 [Kim 2008]。

由于两种状态电阻值之间的差异很大，因此可以通过利用中间电阻值来存储多位(多级单元)，从而提高相变材料的存储密度。几项最新研究

报道了每单元 3 比特的密度[Wu 等人, 2015]。然而, 由于材料成分的波动, 中间值的存储需要一个基于少量 SET 操作的迭代编程和验证过程, 直到达到目标电阻值[Jiang 2014]。

3.1.2. 写耐力。PCM 的主要问题是写入耐久性, 因为 PCM 单元只能支持有限数量的写入操作(通常为 108 个)。这主要是由于反复施加在相变材料上的热应力。实际上, 热膨胀和收缩降低了电极存储接触, 这导致单元中写入电流的可靠性降低。但 PCM 的续航能力比闪存好(105 个周期左右), 比 DRAM 差(1015 个周期)。在决定将这种技术集成到内存层次结构中时, 该属性至关重要。

关于动力系统控制模块写入耐久性的另一个问题与复位最佳电流有关。最新研究证实, 使用高于最佳值的电流值会导致存储单元快速退化, 从而降低其耐久性。如[薛等 2011; Kim 和 Ahn 2005], 写入能量增加两倍会导致耐久性降低大约 50 倍。

3.2 集成选项

PCM 的写入特性缩小了其在不同层次的存储器中的使用范围。写耐久性和高写延迟是主要问题, 因为它们不允许例如在处理器高速缓存的第一级中集成, 因为写流量非常大[Mittal 等人, 2015b]。在最先进的工作中, 主要的集成可能性是作为主存储器或存储系统。在这两种情况下, 动力系统控制模块可以垂直或水平集成。

3.2.1. 主存储器中的动力系统控制模块。相变存储器作为主存储器有三种集成可能性: 1) 作为动态随机存取存储器的替代品, 2) 与动态随机存取存储器水平放置, 3) 与动态随机存取存储器垂直放置。

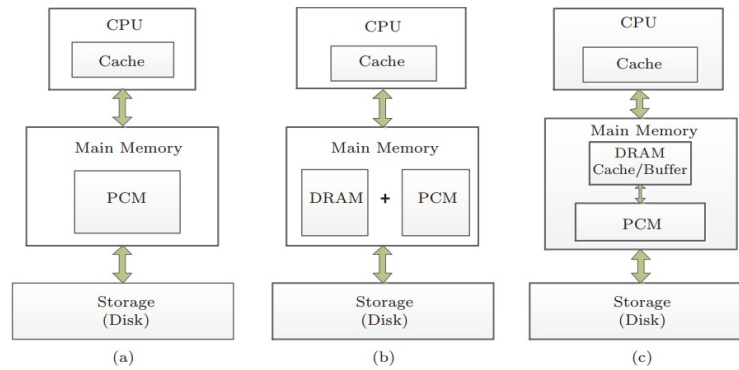


图 3。三种 PCM 集成方案[夏等 2015]

Fig. 3 显示了不同的集成选项。当 DRAM 和 PCM 被放置在同一级别时, 根据访问模式的数据放置是由应用程序(例如, 通过特定的 API, 如在[Kannan 等人 2016]中)在操作系统级别实现的[Lee 等人 2014; Salkhordeh 和 Asadi, 2016 年; Wei 等人, 2015 年), 或内存控制器级别[Dhiman 等人, 2009 年; 库雷希

和 Srinivasan 2009 年; Lee 等人, 2009 年]。事实上, 主要目标是通过在动态随机存取存储器上移动写密集型工作负载/数据来隐藏写操作的高延迟, 并依靠相变材料的非易失性来节省能源[Dhiman 等人, 2009 年; 孙等人, 2015 年]。例如, 在[Wei 等人, 2015]中, 作者使用 NVM 空间来存储文件系统元数据。他们的系统称为 FSMAC, 将数据和元数据输入/输出流从文件系统级别解耦到硬件, 从而实现高性能增益。PCM 也用于检查点的主存储器[Oikawa 2014; 高等[2015]和容错问题[李等 2012]。

在动态随机存取存储器被集成为相变存储器的高速缓存的情况下, 其目标是吸收写操作以减少整体写延迟, 并解决相变存储器的写耐久性问题。事实上, 一些实验结果[Qureshi 和 Srinivasan 2009]表明, 就内存而言, 给定一个小的 DRAM 缓冲区(占 PCM 的 3%), 可以实现速度的三倍提高, 以及寿命的三倍延长。当然, 性能取决于写入模式。[Awad 等 2016]强调了页面预取和页面替换策略等操作系统机制对这种 PCM 集成的高度影响。采用水平和垂直集成, [吴等人 2016]使用一小部分 DRAM 来加速对 NVM 的访问, 同时将另一部分水平放置在 PCM 中。

[Park 等 2015]选择了图 3 的选项(a); 他们用 PCM 代替了 DRAM, 同时试图为 LPDDR2 接口集成优化行缓冲器的使用。

3.2.2. 存储系统中的 PCM。与 NAND 闪存相比, PCM 具有执行字节随机存取和直接就地更新的能力; 这使得它成为闪存的严重竞争对手。

基于 PCM 的存储系统已经定型。例如, Onyx [Akel 等人, 2011]是一种基于 PCM 的存储阵列, 在 PCIe 进行了接口, 事实证明, 对于某些特定的工作负载模式(不规则和读取占主导地位), 它比最先进的固态硬盘具有更好的性能。

许多研究都集中在存储系统中 PCM 的垂直集成上。例如, 在 [Sun 等人, 2010 年]中, PCM 被用于固态存储并与闪存集成, 以吸收日志更新, 类似于缓存。目标是 1) 减轻闪存的突发更新, 这会降低性能并影响耐用性; 2) 提高性能, 因为 PCM 允许就地更新; 以及 3) 节约能源, 因为 PCM 比 DRAM 更节能。其他研究使用了相同的原理, 即在固态硬盘内部使用相变存储器来补充或替换动态随机存取存储器[刘等人, 2011], 目的是在相变存储器非易失性的帮助下, 优化性能, 避免固态硬盘缓存因电源故障而丢失数据。

根据最先进的研究和原型, 基于 PCM 的存储系统显示出巨大的前景, 因为它们可以大幅减少延迟并增加存储系统的带宽, 以实现高性能输入/输出密集型计算[考尔菲尔德等人, 2010 年]。

如何设计这种类型的非易失性存储器以充分利用其性能是一个关键问题。许多原型存储系统都基于 PCIe, 作者声称这种接口会成为瓶颈, 因为它会限制可实现的性能(例如, 在延迟方面)。

从操作系统的角度来看, 存储系统软件堆栈的许多层都是在考虑硬盘驱动器的情况下设计的, 因此可以容忍几微秒的延迟。现在情况不再如此, 需要努力

使存储软件堆栈适应即将到来的新 NVMs [Lee 等人, 2014]。许多问题需要回答, 例如从底层混合体系结构中抽象出什么到操作系统, 如何连接不同的存储系统, 以及应该使用什么协议/机制。为此, 已经提出了一些建议, 例如, 在 [Park 等人, 2010] 中, 作者提出了在 Linux 操作系统中 PCM 的两种可能的集成:

- 1) 在主存储器中同时具有动态随机存取存储器和相变存储器, 并根据输入/输出类型执行数据放置。与全动态随机存取存储器相比, 这将节省大量能量。根据所执行的进程的分段来执行页面放置。
- 2) 作为传统存储系统的补充存储。这里, 数据放置基于将小规模随机访问数据放入 PCM。

本研究提出了一个具体的解决方案, 如何在 Linux 上接口混合内存系统。

3.3 开放式研究问题

[夏等 2015] 观察到, 文献中已经讨论了许多优化问题, 目的是使该技术在不同系统中可行。主要问题涉及优化写入延迟, 以便可以使用 PCM 代替 DRAM, 优化耐用性和节能。

3.3.1. 优化写入延迟。由于写入延迟是由相变材料的某些物理特性造成的, 因此应在其他级别执行优化。

事实上, 第一种优化技术包括优化写命令延迟。许多选择是可能的, 例如: (1) 减少写入位: 该解决方案包括仅写入修改位 [周等人, 2009a; 杨等 2007; 董等人, 2015 年]; (2) 通过更好地利用 SET 和 RESET 操作属性来增加写入位的并行级别 [岳等 2013a; Du 等人, 2013 年]; (3) 隐藏读操作优化的写延迟: 例如, 这是通过在读操作到达时暂停长时间的写操作, 然后在读操作完成后恢复写操作来实现的 [Qureshi 等人, 2010]。

3.3.2. 优化续航能力。相变材料作为动态随机存取存储器替代品的主要问题之一是写入耐久性, 这是由于相变材料的持续加热。已经提出了许多技术来解决这个问题。

优化耐久性的解决方案可以分为三个子类: (1) 通过作用于电气特性来提高耐久性, 例如通过以最佳复位电流为目标, 因为过度编程对耐久性有非常坏的影响。例如, 在 [姜等 2012] 中, 作者提出了一种弹性复位, 它降低了复位电流, 从而对存储器寿命产生了积极的影响。(2) 第二种选择包括最小化写操作的数量; 这可以通过插入吸收频繁写入的 DRAM 缓冲器来实现。提出了其他解决方案, 例如避免写入未修改的存储位 [Lee 等人, 2009; 周等, 2009 年 a; 杨等 [2007] 甚至避免从缓存中写入无用数据 [Bock 等, 2011]。还探索了其他技术, 如数据压缩 [孙等, 2011]。(3) 最后, 全球 PCM

通过磨损平衡技术可以提高耐用性。磨损校平机主要有两类：依赖记忆块状态的磨损校平机，例如切换热数据（频繁更新）和冷数据[Zhou 等人，2009a Yun 等人，2012 年]；以及独立于它们的访问模式随机交换数据的磨损校平机，例如那些周期性地这样做的磨损校平机，例如在[Zhou 等人，2009a 库雷希和斯里尼瓦桑，2009 年]。

3.3.3. 书写干扰。当单元间距离过小时，PCM 会受到写干扰。复位操作产生的高温会通过热传导导致相邻电池发生相变。这可能会诱发寄生 SET 操作[Russo 等人，2008 年]。这种效应可以将技术规模限制在 20 纳米以下。最先进的解决方案包括施加最小的单元间空间，以及设计写入策略和数据编码技术以避免写入干扰[Jiang 等人，2014；王等 2015b 王等 2015c]。

3.3.4. 节能。支持非易失性存储器作为动态随机存取存储器的替代或补充的关键论点之一是关于它们的能量行为。的确，静态能量可以忽略不计。不幸的是，相变材料表现出非常高的写入能量（高于动态随机存取存储器）。已经提出了许多技术来解决这个问题。

有两种主要方法可以降低写操作的能耗：

(1) 减少写入次数，以及 (2) 减少写入操作的能耗。前面引用的技术包括避免更新所有位、避免不必要的写入等，不仅降低了写入延迟，还降低了写入操作的能耗。至于第二类解决方案，它们主要包括利用写不对称 (SET 和 RESET) [徐等 2009；Mirhoseini 等人，2012 年；陈等 2012；岳等朱 2013b] 为了减少能量，因为 RESET 操作比 SET 操作消耗更多的能量[夏等 2015]。

在 MLC PCM 中，单元线的写入以复位操作开始，随后是可变数量的置位操作。这个数字取决于要存储的值 (P&V 过程)。在 [Jiang 2014] 中，作者提出在不同的时间在不同的单元之间调度 RESET 操作，以减小提供 RESET 操作能量的电荷泵的尺寸。通过减小它们的尺寸，可以减少浪费的功率。这种优化依赖于这样一个事实，即根据目标值，要执行的 SET 操作的数量可能不同。结果，一些复位操作可以容易地在时间上偏移，而不会影响单元线写操作持续时间。

3.4 PCM 成熟度和市场状况

许多公司已经投入大量精力，使 PCM 在市场上可用。2004 年 8 月，纳米芯片授权将相变材料技术用于微机电系统 (微电子机械系统) 探针存储设备。2006 年 9 月，三星宣布了一款使用二极管开关的 512 Mb 尺寸的原型设备。相当高的密度值得注意。原型机的电池尺寸只有 46.7 纳米，比当时商用的闪存设备还要小。相比之下，当时唯一生产的 MRAM 和 FeRAM 设备的大小约为 4 Mb。三星原型 PCM 设备的高密度表明，PCM 可能是闪存的可行竞争对手，而不像其他设备那样局限于利基角色[Athmanathan 等人，2016]。

三星发布公告后, 英特尔和意法半导体也相继发布了公告, 它们在 2006 年(在英特尔开发者论坛上)展示了自己的 PCM 设备。他们想出了一个 128 Mb 的设备, 意法半导体开始制造它作为概念验证。2008 年 2 月, 英特尔和意法半导体披露了第一个多级 (MLC) PCM 阵列原型。原型在每个物理单元中存储了两个逻辑位; 实际上, 256 Mb 的内存存储在一个 128 Mb 的物理阵列中。这意味着, 除了正常的两种状态(完全无定形和完全结晶)之外, 额外的两种不同的中间状态代表不同程度的部分结晶, 允许在同一物理区域中存储两倍多的位。

2011 年 6 月, IBM 宣布他们已经创建了一个稳定、可靠、高性能和稳定性的多位 PCM。同样在 2008 年 2 月, 英特尔和意法半导体运送了他们第一个相变材料产品的原型样品——一种 90 纳米、128 兆字节的产品, 他们称之为阿尔维斯通。

2010 年 4 月, Numonyx 宣布了 128-Mbit NOR 兼容相变存储器的 Omneo 系列, 而三星宣布将在 2010 年秋季之前发货的智能手机中使用多芯片封装 (MCP) 的 512 Mb PCM。2012 年 7 月, 美光宣布为移动设备提供 PCM, 这是生产中的第一个 PCM 解决方案。然而, 美光在 2014 年 1 月将所有 PCM 零件撤出市场。2016 年, IBM 证明了他们可以在每个单元中存储两位, 并且正在实现每个单元三位的目标[Athmanathan 等人, 2016]。

4. MRAM

磁性或磁阻随机存取存储器 (MRAM) 是一种更成熟、更有前途的非易失性存储器, 它具有以下非常吸引人的特性: 高密度、快速读取、低泄漏和高耐久性(参见 Table 1 对于 MRAM 的 STT-RAM 变体)。

4.1 基本概念

4.1.1. MRAM 科技。MRAM 是基于一种特殊材料的磁性。与标准的动态随机存取存储器技术相反, 信息载体是磁隧道结 (MTJ), 而不是电荷。可以使用电信号来控制 and 感测磁取向 [Kultursay 等人, 2013]。

MTJ 实现了存储功能。它由被氧化物(隧道)阻挡层分隔的两个铁磁层组成。其中一个铁磁层具有固定的磁化方向, 称为参考(或钉扎)层, 而另一个铁磁层具有可变的磁化方向, 称为自由层。

当两层具有相同的磁化方向时, MTJ 表现出低电阻; 这称为并行状态, 表示逻辑 0(参见 Fig. 4)。当磁化方向彼此相反时, 我们称之为反平行状态, 其中电阻高于并行状态。这表示逻辑 1。

在被广泛研究和本文关注的自旋扭矩转移随机存储器 (STT-RAM) 中, 自由层的磁化方向根据施加在源极线 (SL) 和位线 (BL) 之间的电压而变化。因此, 当在 S1 和 BL 之间施加高正电压差时, 写入逻辑 0, 而如果施加高负电压差, 则写入逻辑 1。为了确保状态反转, 必须为给定的保持已发出的电流

称为写脉冲宽度的持续时间(介于 2ns 和 12ns 之间)及其幅度应等于阈值电流(介于 100 a 和 1000 a 之间)。

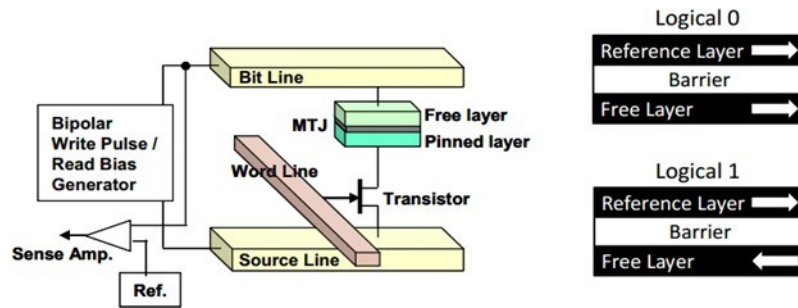


图 4。一种存储单元[薛等 2011]和操作[库尔图说等 2013]。

STT-RAM 与最初的 MRAM 电池的不同之处在于它如何设定 MTJ 的方向:在以前的 MRAM 技术中,方向是使用外部磁场设定的,而在 STT-RAM 则使用极化电子的电流。

为了进行读取操作,在 S1 和 BL 之间施加非常小的电压,这导致电流流过 MTJ。该电流的值与取决于磁化强度的 MTJ 电阻有关。将该电流与参考电流进行比较,然后从 MRAM 单元中读取值 0 或 1。

另一种有前途的 MRAM 技术是自旋轨道转矩随机存取存储器。此处的结使用垂直磁化,并增加了一个额外的端子来分隔读取和写入操作的路径。这种实现允许更好地优化两个(读和写)路径,因为它们有不同的物理要求。

4.1.2. MRAM 耐力。对于 STT-RAM,已经估计了 10^{15} 个写入周期的耐久性值。然而,在其他一些研究中,最佳耐力测试得出的值小于 4×10^{12} [Mittal 等人, 2015b 怀 2008]。

事实上,MRAM 主要受到两个不可靠性来源的影响:(1) MTJ 电池容易出现热不稳定性,这可能会导致数据丢失。失败的概率取决于结的实现、温度、易受攻击时间(写入和最后一次读取数据之间的时间)以及存储单元的数量。随着技术规模的缩小,热不稳定性增加。(2) 高写入电流对存储单元施加应力并使结退化,然后随着时间的推移限制它们的完整性。这个问题会影响到写电流较高的 STT ram。

此外,由于读取操作需要电流,这也会引起磁化强度的变化,称为读取干扰。由于写电流与面积成比例,特征尺寸的缩小导致写电流的减小,而读电流必须保持足以驱动读出放大器[Wang 等人, 2015a]。因此,读取干扰率随着存储器密度而增加。STT 随机存取存储器也比软件随机存取存储器更敏感,因为写和读路径是相同的。

4.2 集成选项

从图中可以看出 Table 1, STT-MRAM 的写入耐久性与动态随机存取存储器和静态随机存取存储器大致相同。这使得它很适合集成于

这些层次的记忆。但是续航距离 1015 的理想值相差甚远。然而,正如我们将在接下来的章节中看到的,已经做出了一些努力来增加耐久性,例如,通过折衷保留时间。

从性能的角度来看,即使 STT-RAM 的读取性能与静态随机存取存储器相当,并且在某些情况下比动态随机存取存储器更好,但它在写入延迟和能量方面是落后的。

大多数最先进的工作在主存储器或高速缓冲存储器级别集成了 STT-MRAM,无论是补充给定级别的层次结构还是替换它。

4.2.1. 片内高速缓存中的 MRAM。在最先进的工作中,和主要通过两种方式集成到片上高速缓存中:(1)在高速缓存层次的所有级别中,如 [Li 等人, 2012; Komalan 等人, 2014 年; 尼森等人, 2015 年], 或 (2)在缓存 (LLC) 的最后一级,如 [Syu 等人, 2013 年; 程等 [2016]。第二组解决方案背后的理念是,一级缓存不能容忍 STT-RAM 的大写入延迟,因为这一级的访问非常频繁。

MRAM 在各级缓存中。许多研究建议在所有级别的缓存中使用 STT-MRAM。在 [尼森等人, 2014 年] 中,作者强调了用 STT-RAM 取代 L1 和 L2 的静态随机存取存储器缓存在静态能量方面可以预期的收益。研究表明,尽管测试工作负载的性能相当,但 STT-RAM 大幅降低了静态能量,同时消耗了更多动态能量。在 [Komalan 等人 2014] 中也得出了相同的结论,其中 STT-RAM 通过扩展未命中状态保持寄存器来应对 NVM 的写延迟,从而取代了 L1 指令高速缓存。

多级高速缓存提供了缓解读干扰问题的能力。对于作为缓存的集成,在 [Wang 等人, 2015] 中,提出了一种读取和恢复策略,通过该策略,在处理未命中之后,恢复缓冲器在后台刷新源块。基于当读取电流与写入电流方向相反时会发生读取干扰的观察,他们还建立了两次读取序列,这使得检测受干扰的存储单元成为可能,从而仅刷新这些存储单元。由于读取干扰管理,这两种技术优化了性能并降低了能量开销。

[李等 2014] 的作者设想在一级高速缓存系统上将与 SRAM 水平集成。在这样的混合缓存中 [孙等, 2009; 吴等 2009; Jadidi 等人, 2011 年; Li 等人, 2011], 系统根据访问模式迁移来自这两种技术的数据,主要是为了减轻高写操作延迟。[Li 等人 2014] 的作者提出了一种编译器辅助的方法来最小化迁移开销。在 [王等 2015] 中,作者提出了一个完整的集成设计,作为实时应用的暂存存储器。

如前所述,非常有前途的软件随机存取存储器技术大大减少了访问时间和能源消耗。它甚至被用于 L1 混合缓存。在 [obril 等人, 2015] 中,作者研究了静态随机存取存储器、软件随机存取存储器和 STT-RAM 的许多配置。评估的主要指标是面积、执行时间和能量。该研究最后得出了一些有趣的结论,例如,对于低功耗设备来说, SOT-RAM 可能是一级缓存的最佳选择,但是对于小内存块来说,从面积和性能来看,静态随机存取存储器仍然是最佳解决方案。它们甚至是混合配置的许多情况

例如用于 L1 指令高速缓存的软件随机存取存储器和用于 L1 数据高速缓存的静态随机存取存储器，而用于 L2 高速缓存的软件随机存取存储器是最佳解决方案。

末级缓存中的 MRAM。将 MRAM 引入 LLC 的研究中使用的主要技术是 STT-RAM。

在[Wu 等人, 2009]中, 作者提出了一个非常有趣的研究, 当将一些非易失性存储器(如和 PCM)与一些具有易失性存储器(如 eDRAM)的新技术集成时, 探索高速缓存的体系结构。在这项研究中, 作者评估了许多不同的体系结构, 分为两种类型的混合缓存, 即:缓存间级混合缓存体系结构(LHCA)和缓存内或基于区域的混合缓存体系结构(RHCA)。在前一种情况下, 每一级都严格地由一种存储技术组成(纵向集成), 而在后一种情况下, 每一级都可能由具有不同技术的区域组成(横向集成)。研究表明, 与三级静态随机存取存储器高速缓存相比, LHCA 的 IPC(每个周期的指令数)可以提高 7%, 而 RHCA 在相同的面积限制下, IPC 可以提高 12%。

关于水平集成, 对于给定级别的缓存, 可以集成两种内存技术, 系统应该应用一些自适应数据放置机制, 以充分利用每种集成技术。例如, 在[Wang 等人, 2014]中, 系统试图通过检测 LLC 上的访问模式来减少对写的写入。

遵循对给定级别的缓存进行分区的思想, 在[Samavatian 等人, 2014]中, 作者通过调整保留期将 L2 缓存划分为 STT-RAM 的两个区域。他们使用低保留分区来吸收频繁写入的数据, 使用高保留区域来存储不太频繁写入和非常频繁读取的数据。降低保持率允许写入能量和 STT-RAM 单元的面积减小。

4.2.2. 主记忆中的 MRAM。尽管大多数关于集成 MRAM 的最先进的工作都集中在片上高速缓存子系统上, 但一些研究探索了集成 MRAM 作为 DRAM 的替代或补充。

在[Kultursay 等 2013]中, 作者研究了 STT-RAM 对主记忆的完全替代。事实上, 如前所述, 动态随机存取存储器受到两个主要因素的限制:可扩展性和功耗。作者首先比较了动态随机存取存储器和非优化的 STT-RAM 技术。他们表明, STT-RAM 简单地替换 DRAM 并不能带来好的性能; 相反, STT-RAM 主存储器的表现非常糟糕, 因为它的写操作延迟和能量消耗。然后, 作者对 STT-RAM 进行了一些优化。一种优化涉及利用其感测放大器和行缓冲器的去耦结构的特性。这允许它在存储单元上执行选择性和部分写入, 从而减少写入操作的数量, 从而减少延迟和能量。另一个优化包括为写操作绕过行缓冲区, 因为观察到写操作的命中数比读操作低得多。在给定数量的工作负载和多核处理器上进行此类优化的结果表明, STT-RAM 内存在实现与动态随机存取存储器相当的性能的同时, 大幅降低了能耗(约 60%)。

在[Yang 等人, 2013]中研究的另一种方法包括将与 DRAM 水平集成, 同时实施一种方案来引导写密集型数据

到 DRAM。将两种内存技术集成到主内存中的主要目的是降低能耗水平。

尽管先前的研究显示将集成到主存储器中有很好的结果,但正如[王等 2014]所指出的,芯片的内部结构与当前的 DRAM 接口不兼容。因此,为了解决这个问题,作者提出了一些技术来使 LPDDR3 的设计与 MRAM 架构兼容。

在[Jin 等人, 2014]中提到了与记忆的面积、能量和潜伏期有关的其他问题。作者提请读者注意 STT-RAM 单元的大小,这可能是获得类似动态随机存取存储器密度的严重障碍。此外,据报道,STT-MRAM 的写入能力对于许多应用程序域来说太大了。作者建议修改一些关键参数,如热稳定性因子、临界电流和保留时间,以减轻 STT-RAM 电池的能量和延迟特性,使其更适合作为动态随机存取存储器的替代品。

4.2.3. 存储系统中的 MRAM。有几项关于将 MRAM 集成到存储系统设备中的研究。李等人研究了存储系统中的相变材料和 STT-RAM 集成[李等人, 2014 年]。这项有趣的研究通过集成这些非易失性存储器强调了内核输入/输出软件堆栈的作用。此外,他们还比较和分析了内核机制(如预读)和访问方法(如直接和同步 I/o)与基于 NVM 的存储系统的效率。在[康等 2015]中,作者提出了存储系统中 NVM 的垂直集成。这项研究集中在脉码调制,但他们的技术也可以在 STT-RAM 工作。其思想是在存储系统中将 NVM 用作缓存时使用保留放宽,以减少延迟,因为缓存中的大多数数据在进入缓存后会在短时间内被逐出。

4.3 开放式研究问题

在许多关于在给定缓存级别使用 STT-RAM 的研究中,作者试图通过一组给定的优化来提高其性能或适应 MRAM 的约束。这些总结在[Mittal 等人, 2015b]:

4.3.1. 放松非波动性。放松 STT-RAM 的非波动性[Smullen 等, 2011; Jog 等人, 2012 年; 孙等人, 2011 年; 郭等(2010):这里的想法是使用具有不同保留期的的不同区域(水平集成)或水平(垂直集成),从而导致不同的写入延迟和能耗属性。

4.3.2. 最小化或避免写入。最小化或避免写入[Joo 等人, 2010; Bishnoi 等人, 2014 年; 周等 2009bGoswami 等人, 2013 年; Yazdanshenas 等人, 2014 年; Ahn 等人, 2012 年; 孙等人, 2009 年; Rasquinha 等人, 2010 年; Jung 等人, 2013 年; Park 等人, 2012 年; Cheng 等人(2016):该优化的主要目标是减少在上更新的比特数。这可以通过不同的技术来实现:例如,在写入数据之前读取数据并只写入不同的位,在更新之前使用缓存来执行比较,对数据进行编码以减少位修改,等等。

4.3.3. 提高内存寿命。通过使用损耗均衡来提高内存寿命:对于闪存,损耗均衡技术可用于提高内存寿命,同时应对工作负载的时间和空间局部性[Mittal 2013; 陈等 2013; 王等 2014aJokar 等人, 2016 年]。

4.3.4. 解决 0/1 不对称。由于 MTJ 开关时间的不对称性，执行 1 到 0 转换所需的时间大于执行 0 到 1 转换所需的时间，因此在操作之前将所有单元预设为 0 允许减少有效写入等待时间[Kwon 等人，2014]。

4.4 MRAM 成熟度和市场状况

公司已经开发 MRAM 很多年了。2003 年，采用 180 纳米光刻工艺制造的 128 千比特 MRAM 芯片问世。第二年，Infineon 推出了一款 16 兆比特的原型机。2005 年，霍尼韦尔发布了一份使用 150 纳米光刻工艺的 1 兆拉德硬 MRAM 的数据表，随后是运行在 2 千兆赫的 MRAM 存储单元。索尼宣布第一个实验室生产的 STT-MRAM 利用自旋极化电流通过隧道磁阻层写入数据。这种方法比传统的 MRAM 消耗更少的能量并且更具可扩展性。同年，飞思卡尔半导体公司展示了一种 MRAM，它在写入周期中使用氧化镁阻挡层并改善了位电阻，从而降低了所需的写入电流。

东芝和 NEC 在 2006 年宣布了一款 16 兆比特的 MRAM 芯片，采用了新的功率分流设计。它实现了 200 兆字节/秒的传输速率，34 纳秒的周期时间，是所有 MRAM 芯片中性能最好的。日立和东北大学在 2009 年展示了一个 32 兆比特的 STT-RAM，然后在 2010 年宣布了一个多层次的 MRAM。2012 年 11 月，Everspin 推出了 64Mb STT-MRAM(Everspin 2015)。

5. FERAM

铁电随机存取存储器 (FeRAM 或 FRAM) 是一种比其他非易失性存储器 (闪存除外) 更早进入大规模生产阶段的非易失性存储器。事实上，如今 FeRAM 被用于许多产品，主要是嵌入式领域[藤崎琴音，2013]，例如，数据记录器和无线网卡等汽车设备，并有望成为可穿戴电子设备非易失性存储的良好候选产品[文卡特什和辛格，2015]。

FeRAM 表现出一些非常吸引人的特性，使其成为未来 NVM 最强的早期候选之一[Burr 等人，2008 年]，即低延迟 (低至 DRAM 的延迟)，据报道为 20ns [Burr 等人，2008 年]，低功耗，无擦除写操作，以及直接 CMOS 集成的可能性[Kato 等人，2005 年]。

5.1 基本概念

铁电存储器的单元类似于动态随机存取存储器的单元，因为它由一个晶体管和一个铁电电容器 (FCAP) 组成，也称为 1T1C 结构[Kryder 等人，2009; Kanaya 等人，2004 年]。铁电材料由铅 (Zr_x, Ti_{1-x})O₃ (PZT 或锆钛酸铅) 在两个金属电极之间形成。FCAP 的特征是两个剩余的可逆偏振态。由于铁电材料与动态随机存取存储器中的介电材料相比表现出的磁滞特性，FCAP 电荷保持其极性而不需要电源。这解释了它的非波动性。此外，FeRAM 不需要刷新操作；因此，它比动态随机存取存储器消耗更少的功率。

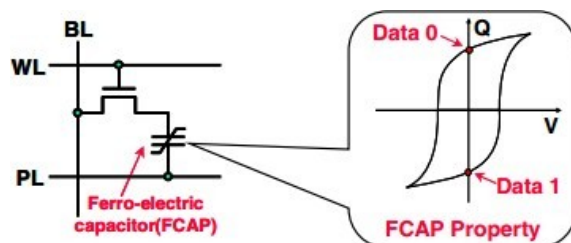


图 5. 一种铁电存储器单元

写操作是通过迫使脉冲通过 FCAP 的板线(PL)为 0, 并通过位线(BL)为 1 来实现的(参见 Fig. 5). 与 NAND 闪存相反, FeRAM 不需要非对称电压和/或用于写操作的特殊高电压。对于 FeRAM, 使用的电压与 V_{cc} 相同, 并且可以非常小。

如前所述, FeRAM 被视为最早取代和/或补充动态随机存取存储器和存储系统技术的成熟非易失性存储器之一。然而, 经过多年对该技术的研究, FeRAM 细胞似乎仍然难以扩展。事实上, 它们的尺寸仍然明显大于千兆级集成的动态随机存取存储器单元。因此, 正如将在下一节中看到的, 最近关于 FRAM 集成的工作解决了 FeRAM 是否可以用于嵌入式系统(它们已经部署在微处理器中)或者在混合系统中数量有限, 或者作为小容量高速缓存的问题。

5.2 集成选项

在关于 FeRAM 性能的最新研究中, 给出了许多不同的性能数据。然而, 许多人认为它的读写操作或多或少是对称的。与 MRAM 和 PCM 不同, FeRAM 在读和写操作之间没有表现出强烈的不对称性。FeRAM 的性能接近 DRAM, 但似乎仍然落后。相比之下, 它在功耗方面要优于 DRAM。关于耐力, 报告的数值在 10^{12} 到 10^{15} 之间。

FeRAM 的集成可以从不同的角度进行讨论:

- 从耐用性的角度来看, FeRAM 可以取代 DRAM 或闪存等存储系统技术。

- 从性能的角度来看, FeRAM 可以补充 DRAM 并取代存储系统技术。

- 从密度的角度来看, FeRAM 还不能取代 DRAM 或存储系统技术, 因为它对千兆位的可扩展性仍然是一个悬而未决的问题。由于无法证明 FeRAM 具有高密度, 因此与闪存相比, 它的价格仍然非常高, 尽管嵌入式领域的一些应用正在推动这项技术。

尽管如此, FeRAM 已经在最先进的工作中得到研究, 事实上, 由于 FeRAM 背后的物理已经被很好地掌握, 如果发现一些具有更好可扩展性的新材料, 它的性能可以很容易地提高。

一些 FeRAM 集成命题存在于最先进的研究中。据我们所知, 在所有的研究中, 它都集成在动态随机存取存储器和存储器之间

系统。这种集成要么是横向的，用于以 FeRAM 取代或补充 DRAM 和存储系统的方案；或者垂直，包括在 DRAM 和存储系统之间插入 FeRAM，通常作为元数据高速缓存[Doh 等人，2007；Yoon 等人，2008 年]。

5.2.1. 主存中的 FeRAM。我们没有找到一套详尽的关于使用 FeRAM 作为主要内存替代或补充的最先进的工作。这可能是由于以这种方式集成时可能导致的性能下降。

已经确定了两项主要研究。在[Suresh 等人，2014]中，作者评估了许多不同的 NVM 技术，其中包括 FeRAM。它已被研究作为动态随机存取存储器的替代物，也作为形成混合主存储器的补充存储器。测试了铁电存储器的性能，并与其他非易失性存储器，即 STT-RAM 和相变存储器进行了比较。模拟表明，与动态随机存取存储器相比，单独使用铁电存储器会降低性能，但在某些情况下，它会导致能耗的小幅提高。然而，对于显示大静态能量的工作负载，具有 DRAM 和 FeRAM 的混合存储器显示出更好的能量效率，即使它降低了性能。与其他国家妇女相比，野生动物在这项研究中表现较差。

另一项关于非易失性存储器集成的有趣研究（在本研究中称为存储类存储器——单片机）已经进行[Baek 等人，2013]。在这项研究中，针对不同的工作负载测试了几种体系结构（请参见 Fig. 6）。这包括作为存储系统的经典动态随机存取存储器+闪存、动态随机存取存储器+单片机、仅单片机和单片机+闪存。一个 32MB 的 FeRAM 原型被用作单片机。这项研究的第一个观察结果是，在大多数情况下，单片机+闪存的性能最差。这仅仅意味着用铁电存储器直接替换动态随机存取存储器可能不是一个好主意，因为铁电存储器的延迟很高（与动态随机存取存储器相比）。另一方面，依赖于工作量，仅配置管理产生了一些好的结果。作者更进一步，提出了一些优化配置管理的配置。其想法是就地更新文件和元数据，而不是像内存/存储传统系统那样有多个副本。作者提出了一个统一的文件/数据管理系统，称为对象管理。传统的文件和内存 API 之间建立了桥梁，由配置管理器实现对象管理。通过这种新的体系结构，作者展示了与简单地重用传统操作系统内存和存储堆栈的配置管理相比，一些显著的性能增强。

之前研究的作者明确提到，即使 FeRAM 的能耗属性优越，其性能也落后于 DRAM。然而，如果要在高性能计算中使用，FeRAM 需要展示可扩展性，这是它远远不能做到的。

5.2.2. 存储系统中的 FeRAM。从存储系统的角度来看，FeRAM 在最先进的研究中主要是水平集成的，除了在变色龙[Yoon 等人，2008 年]中，它是水平和垂直集成的，在[Baek 等人，2013 年]中，作者主要集中在用 FeRAM 替换主存储系统，有或没有 DRAM。

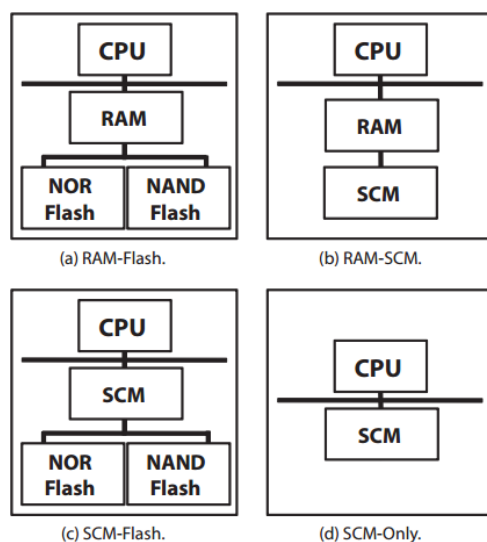


图6. 在[Baek 等人, 2013]中测试了内存层次结构

如上所述, 由于集成和可伸缩性问题, FeRAM 的使用一直很少。除了[Baek 等人, 2013]之外, 大多数最先进的研究主要使用 FeRAM 来优化对文件系统的数据访问[Jung 等人, 2010; Doh 等人, 2007 年]或 FTL [Yoon 等人, 2008 年]元数据。

在 Chameleon [Yoon et al. 2008] 中, 想法是将 (1) FTL 元数据 (如映射表) 存储在 FeRAM 中, 以及 (2) 使用 FeRAM 块和页面缓冲区来加速对底层闪存的访问。在 [Doh 等人, 2007] 中, 主要思想是将闪存文件系统 (米女) 元数据存储到 FeRAM 中, 而不是将数据存储在闪存中, 并在装载时将其复制到 DRAM 中。在测试的工作负载上观察到了非常好的性能。

在 FRASH [Jung 等人, 2010] 中, 作者试图将持久性引入内存数据结构, 这样就有可能加速一些操作, 例如装载时间, 并通过将文件系统数据结构存储在 NVM (FeRAM) 中来提供对它们的更快访问。对于第一个目标, 目标是克服动态随机存取存储器的易失性, 并通过将数据结构存储在 FeRAM 中来避免昂贵的装载时间数据结构。对于第二个目标, FRASH 在 FeRAM 中存储一些存储系统结构, 例如页面元数据。事实证明, 与经典的闪存文件系统 (如 YAFFS 或其他单片机集成解决方案) 相比, FRASH 具有良好的性能。

5.3 FeRAM 成熟度和市场状态

许多公司都在提供基于 FeRAM 的产品。其中有 Ramtron, 产品种类繁多; 带有基于 FeRAM 的微型计算机的德州仪器; 还有东芝, 它在 2009 年以 130 纳米技术推出了 128 兆字节的 FeRAM [Shiga 等人, 2009]。

目前的大部分 FeRAM 技术是由无晶圆厂半导体公司 Ramtron 开发的。一个主要的被许可方是富士通, 它经营着可能是最大的具有 FeRAM 能力的半导体代工生产线。自 1999 年以来, 富士通一直使用这条生产线来生产独立的铁电存储器, 以及带有嵌入式铁电存储器的专用芯片 (如智能卡芯片)。日本富士通公司

在 2010 年之前为 Ramtron 生产设备。自 2010 年以来, Ramtron 的制造商一直是德州仪器 (TI) 和 IBM。至少从 2001 年开始, TI 就与 Ramtron 合作, 在改进的 130 纳米工艺中开发 FeRAM 测试芯片。2005 年秋季, Ramtron 报告称, 他们正在评估使用 TI 的 FeRAM 工艺制造的 8 兆位 FeRAM 原型样品。富士通和精工爱普生在 2005 年合作开发了一种 180 纳米的 FeRAM 工艺。2012 年, Ramtron 被赛普拉斯半导体收购。

据报道, 三星、松下、Oki、东芝、英飞凌、海力士、赛默特、剑桥大学、多伦多大学和大学间微电子中心(比利时 IMEC)也有 FeRAM 研究项目。

自 2016 年以来, TI 最新 MSP430 微控制器一直使用 FeRAM 以统一的内存方式进行代码和数据存储[Wong 2016]。

6. RERAM

电阻式随机存取存储器背后的基本思想是, 它们是以电阻状态的形式存储逻辑状态的存储器[桑德胡 2013]。由于其特性, ReRAM 应该集成许多应用领域, 如消费电子、个人计算机、医疗应用、空间和汽车应用[Meena 等人, 2014]。事实上, ReRAM 与传统的半导体制造工艺兼容。ReRAM 是一项非常有前途的技术, 因为它的功耗比 PCM 低, 密度比 MRAM 高(参见 Table 1)。ReRAM 在 10 纳米的制造技术方面表现出良好的稳定性[Akinaga 和 Shima 2010]。从耐用性的角度来看, 它比闪存好一到五个数量级(参见 Table 1)。从可扩展性和制造的角度来看, ReRAM 可以被认为是通用存储器的一个重要候选; 然而, 它的耐久性仍然有限, 不能替代 SRAM 或 DRAM。然而, 目前对它进行了非常广泛的研究, 许多科学障碍可能在未来几年得到解决。

6.1 基本概念

ReRAM 属于忆阻器器件的范畴。忆阻器是具有可变电阻的双端电阻器件。电阻值根据所施加电压的大小和极性以及施加电压的持续时间而变化[William 2008]。忆阻器的主要特点是断电时电阻值不变; 这为它提供了非易失性。

忆阻器的概念是 1971 年由 Leon Chua 发明的[Chua 1971]。它被称为电子学的第四个基本构件(与电阻器、电容器和电感器一起)。2008 年, 惠普实验室的一组研究人员证明了这一点[Sturkov 等人, 2008]: “忆阻器是存储电阻的收缩, 因为这正是它的功能……无限存储电阻值的能力意味着忆阻器可以用作非易失性存储器” [William 2008]。在忆阻器中, 低电阻值被视为逻辑 1, 而高电阻值被视为逻辑 0。为了读取存储单元值, 感测逻辑将流过所选忆阻器单元的电流与该值进行比较

流经参考电阻, 两者上施加的电压相同。

ReRAM 电池是一种双端金属-绝缘体-金属(MIM)器件, 由夹在两个电极金属导体之间的绝缘或电阻材料(即氧化物)组成(参见 Fig. 7)。

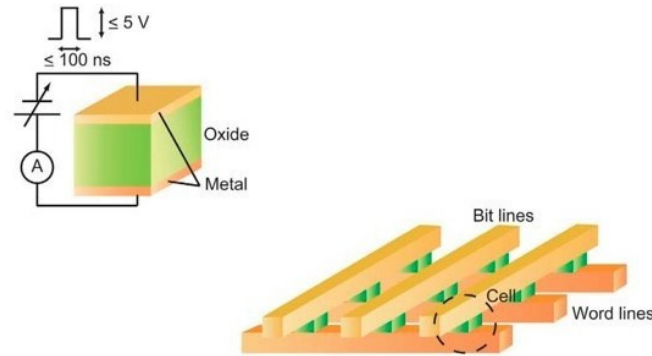


图 7。ReRAM 单元 3

ReRAM 背后的基本原理是导电细丝的形成(低电阻状态)和破裂(高电阻状态), 其目的是通过电阻氧化物层分流顶部和底部电极(M) (参见 Fig. 8). 根据其开关行为, ReRAMs 可分为以下两种类型: 1) 单极器件, 其中开关方向取决于施加电压的幅度[Kryder 等人, 2009; 桑德胡 2013; William 2008]和 2) 双极器件, 其中开关方向取决于电压极性。在双极器件中, 开关机制基于氧空位的产生, 氧空位在顶部和底部电极之间形成电通路。TiO_x 材料已被广泛用作顶部和底部电极材料[Choi 等, 2005; 藤本等人, 2006 年; Hosoi 等人, 2006 年]。许多其他材料已被用于最先进的工作, 如氧化锆[李等人, 2006 年]和氧化镍[吉本斯和比德尔 1964]。

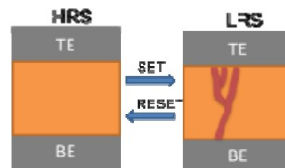


图 8。氧化物基随机存取存储器中的随机存取存储器单元和细丝形成

导电桥随机存取存储器 (CBRAM), 另一项有前途的技术, 是一种固体电解质存储器, 其中固体电解质材料夹在惰性电极(阴极)和可氧化电极(阳极)之间。

6.2 集成选项

由于其集成能力和性能, ReRAM 在许多研究中被认为是存储器层次结构中的潜在替代和/或补充存储器。据报道, ReRAM 的续航时间在 10^{10} 至 10^{12} 之间[Shang 等人, 2011 年; Kim 等人, 2011 年; Eshraghian 等人, 2011 年], 这意味着它

³ 来源: <http://www.dobreprogramy.pl/Crossbar-pokazal-terabajtowy-czip-pamieci-ReRAM-niebawem-新闻, 45456.html>

如果没有特定的优化，很难集成到缓存级别。已经针对水平和垂直集成研究了 ReRAM。

6.2.1. 缓存内存中的 ReRAM. ReRAM 的持久性可能会对其在低内存级别(如缓存)的集成造成障碍，这种可能性已经在许多研究中进行了调查。

在[董等 2013]中，作者研究了许多 nvm，重点是一个 ReRAM 案例研究。提出了不同的模型，允许为集成随机存取存储器和其他非易失性存储器进行设计空间探索。这些模型是通过一个名为 NVSim 的模拟器实现的[董等，2012]。[Dong 等人，2012]的作者比较了在高速缓存(传统 SRAM)堆栈中垂直集成 ReRAM 时的性能差异。测试了许多场景：将 ReRAM 放在所有传统缓存级别，然后放在两个更高的级别(L2 和 L3)，然后只放在更高的级别(L3)。在每种情况下，作者都注意到能量增益和性能下降(在读取性能方面，ReRAM 与 SRAM 相当，但在写入性能方面明显较慢)。人们需要根据应用所需的能量/性能权衡，在所选架构上做出妥协。在之前的一项研究中，作者依靠磨损均衡技术来部分解决耐久性问题[Schechter 等人，2010；王等[2013]。在[Mittal 和 Vetter 2015a]中，作者展示了 L2 缓存中的 ReRAM 与 AYUSH 系统中的 SRAM 的水平集成。为了解决随机存取存储器的耐久性问题，经常重复使用的数据被存储在静态随机存取存储器中。

针对更具体的应用领域，在内存层次结构中也引入了 ReRAM。在[Komalan 等人，2013]中，它被集成到 L0(循环缓冲器[Uh 等人，1999])和/或 L1 高速缓存级别，用于超低功耗嵌入式指令存储器，专用于具有特定读取主导工作负载(循环主导代码)的低功耗无线/多媒体应用。当集成到 FPGAs 中作为静态随机存取存储器的替代品时，ReRAM 也表现出良好的性能，以改善电源泄漏和快速开/关操作等问题[clermidi 等人，2014]。

6.2.2. 主内存中的 ReRAM. 至于 PCM，ReRAM 已经集成在主存储器中，以补充 DRAM。例如，在[Hassan 等人，2015]中，作者考虑了带有 DRAM 的水平方案中的 ReRAM。他们提出了应用层面的数据放置和迁移策略，并在关键价值数据存储上进行了验证。

在[Xu 等 2015]中，基于没有选择晶体管(OT1R)的单元阵列结构，对 8Gb 的 ReRAM 芯片进行了建模。该芯片与 DDR3 接口兼容，并受到与相同容量的动态随机存取存储器相同的功率和面积预算的限制。作者通过微架构级增强和数据编码优化了写入延迟。他们的模拟显示，与使用动态随机存取存储器芯片的主存储系统相比，性能下降不到 10%。

6.2.3. 存储系统中的随机存取存储器. 许多研究人员发现，ReRAM 耐力不能适应低水平的记忆层次；因此，ReRAM 被认为是闪存未来的替代或补充。

在[Tanakamaru 等 2014]中，作者研究了由小容量的 ReRAM 和大容量的闪存组成的混合存储系统。ReRAM 的集成是横向的，目标是存储镜像信息、奇偶校验和错误记录，因为 ReRAM 是非易失性的、页面可重写的，并且比闪存具有更好的性能和耐用性。拟议的目标

统一固态存储系统 (USSS) 实现了反向镜像、移位镜像、错误减少合成和页面 RAID, 旨在提高内存的可接受原始误码率 (ABER)。

在 [Jung 等人, 2013] 中, 作者强调了设计大规模交叉开关 ReRAM 的主要问题, 例如写干扰和读检测裕度 [Jung 等人, 2013]。作者认为, 如果不考虑这个问题, 对于大规模的 ReRAM 来说, 检测裕度是不可接受的。作者提出了一些基于新的访问策略的微体系结构解决方案来解决这个问题。他们还提出了一种宏架构设计, 并评估了 8 GB ReRAM 与两种闪存配置 (集成为闪存的替代品) 相比的性能。所提出的体系结构提供了高达 8 倍的带宽和 66% 到 88% 的延迟。

其他研究使用水平集成在固态设备中的 ReRAM, 以吸收一些特定的模式, 并从密集和随机的碎片数据更新中释放闪存。在 [Sun 等 2014] 和 [Fujii 等 2012] 中, 作者提出了将 ReRAM 与 NAND 闪存横向集成在一起; 这个想法是将 ReRAM 用于高度写入和随机的数据, 以优化整体性能。

6.3 公开的研究问题

与 ReRAM 相关的开放式研究问题和挑战已在 [薛等, 2011] 和 [杨和 Williams, 2013] 中进行了讨论, 并在本节中进行了总结。

6.3.1. 转换耐力。开关耐久性与器件在其性能下降到不可容忍的阈值之前可逆地接通和断开的周期数有关。文献报道的耐久值远远没有达到金属氧化物忆阻器件的极限。关键在于开关材料的选择。已经研究了许多开关材料, 例如 Pt/TaO_x/Ta、TiO_x、HfO_x [Yee 等人, 2010] 和 WO_x [Chien, 2010]。一些显示了接近 100 亿个开关周期的耐久性, 这仍然远远低于所寻找的大约 10¹⁵–10¹⁶ 的通用存储器耐久性。

6.3.2. 设备产量。为了使 ReRAM 在制造上可行, 应该提高器件产量, 因为据报道它很低 [Yang 等人, 2010]。决定这一产量的关键参数之一是器件内部开关通道的控制 [Yang 和 Williams 2013]。一些最先进的解决方案通过植入纳米级开关中心提高了产量 [杨等, 2010]。

6.3.3. 运作能量。尽管与闪存和相变存储器等其他非易失性存储器相比, 切换随机存取存储器单元所需的能量非常小, 但降低其值以提高存储密度仍然非常重要。切换所需的能量大于读取所需的能量。降低开关能耗的主要解决方案有两种: (1) 降低结尺寸, 或 (2) 优化开关材料 [苗等, 2011]。

6.3.4. 电铸工艺。电铸是施加高电压或电流, 需要一次进行, 以便能够切换电阻存储单元。这一过程导致电子传导率降低 [薛等, 2011; 杨等 [2009]。此后, 根据所用材料的不同, 该装置的运行特性也有很大的不同。在计算电路应用中限制这种差异非常重要 [杨等, 2009]。

6.3.5. 潜路径和非线性。潜行路径是交叉开关结构中的意外路径，可能会导致写和读失败[Jung 等人, 2013]。这是由于在 ReRAM 交叉结构上没有晶体管或任何存取器件造成的。防止潜路径是实现大规模 ReRAM 存储器的必要特性[陈等, 2011a 陈等 2011b]。为了在不引入接入设备的情况下解决潜通路问题，需要针对 on 状态的高度非线性的电流-电压关系[Yang 和 Williams 2013]。事实上，非线性越高，存储单元越好，因为潜在电流显著降低[Niu 等人, 2012]，并且写电流和能量消耗的缩放越好。如[Jung 等人, 2013]中所讨论的，最先进的潜在路径解决方案可能会引入一些其他问题，例如写干扰和读检测裕量，这些问题可能会阻止 ReRAM 扩展，从而成为替换存储系统的候选。

为了进行 ReRAM 的设计空间探索，开发了许多工具，例如 NVSim [Dong 等人, 2012]，并且建立了模型来研究和优化许多参数，例如面积、延迟、能量[Xu 等人, 2011]和可靠性[Xu 等人, 2014]，或者研究实现多级单元 ReRAM 的权衡[Xu 等人, 2013]。

6.4 ReRAM 成熟度和市场状态

ReRAM 仍在开发中，一些原型已经由 Rambus 公司和 Crossbar 公司提供。2012 年 2 月，Rambus 收购了一家名为 Unity Semiconductor 的 ReRAM 公司，松下在 2012 年 5 月推出了 ReRAM 评估套件。评估套件基于氧化钽 1T1R 存储单元架构。2013 年，Crossbar 推出了一款芯片形式的 ReRAM 原型，其大小约为一枚邮票，可以存储 1 TB 的数据。2013 年 8 月，Crossbar 公司声称，他们的 ReRAM 芯片的大规模生产计划在 2015 年进行。同年，英特尔和 Micron 推出了一款与 ReRAM 相似的所谓 3D Xpoint 内存，该内存由 Crossbar 开发。2016 年，西部数字在闪存峰会上宣布，打算将 3D 电阻式 RAM (ReRAM) 作为存储级内存 (SCM) 用于其未来的专用超快固态硬盘[Shikov 2016]。

7. 结论

在本文中，我们介绍了四种非常有前途的非易失性存储器技术：PCM，MRAM、FeRAM 和 ReRAM，每一个都有它的优点和缺点。这些非易失性存储器处于不同的开发阶段，其中一些已经在制造中（例如一些微型计算机中的铁电存储器），另一些仍处于原型阶段。这些非易失性存储器已经在最先进的研究中进行了讨论，并根据许多因素进行了集成，例如耐久性、写入延迟、能量和工艺制造兼容性。

已经提出了两种整合国家风险管理的主要方式：横向或纵向。水平集成包括补充内存层次金字塔中现有的内存级别，从而为存储数据提供至少两种选择（通常没有冗余）。这种情况下的数据放置遵循一些与模式相关的属性，主要基于两个特征：操作类型和访问模式。在垂直集成中，在内存层次结构中插入了一个附加层。插入的级别充当缓存或缓冲区，以应对两种内存技术之间的性能差距。

从应用甚至操作系统的角度来看, 集成主要有两种选择。首先, 可以对操作系统和/或应用程序隐藏引入的内存, 就像存在不同级别的缓存内存一样, 其中操作系统知道缓存的存在, 但不知道有多少级别, 也没有能力管理缓存, 因为这种管理是在硬件中完成的。第二个选项包括使新集成的内存对操作系统可见, 并最终对应用程序可见, 以便可以选择存储数据的内存。

NVM 为优化计算系统带来了新的研究机会。持续开发 NVM 的特定特性应该是一个有趣的方向。例如, 与动态随机存取存储器相比, 所有非易失性存储器技术都具有较高的静态功率, 但写入性能和耐久性较差。因此, 需要开发更有效的硬件/软件技术来利用这些特征。此外, 将非易失性存储器集成到存储器/存储系统中可能会引入新的异构流量。例如, 输入/输出流量可以传递到内存系统。艾尔。2016]. 开发有效的架构技术来处理这种异构性, 同时对现有系统进行少量修改, 将是一个挑战。最后, 通过 NVM 集成, 可以构建统一的内存/存储系统。研究如何在软件级别充分利用这种统一的内存/存储系统应该是很有趣的。

参考

- J. 安凯财。2012. 低功耗 STT-RAM 缓存的低位缓存。在国际标准学会会报, 480-483。
- A. 阿克尔, 考尔菲尔德, 莫洛夫, 古普塔, 斯旺森。2011. 编玛瑙: 一种原型相变存储阵列。在《热存储学报》中。
- H. 秋泽, H. Shima。2010. 基于金属氧化物的电阻随机存取存储器。IEEE 会议录, 98, 2 (2010), 2237-2251。
- A. Athmanathan, M. Stanisavljevic, N. Papandreou, H. Pozidis, E. Eleftheriou。2016. 多级单元相变存储器: 一项可行的技术, 电气和电子工程师协会电路和系统新兴和精选主题杂志, 6, 1 (2016), 87-100。
- A. Awad, S. Blagodurov, Y. Solihin。2016. 基于非易失性存储器的存储器扩展的写感知管理。在国际科学学会会刊上。第九条
- 南白先生, 蔡先生, 李博士, 诺先生。2013. 面向存储级内存的高能效和高性能软件架构。嵌入。电脑。系统。12, 3 (2013). 第八十一条。
- R. 比什诺伊, 奥博里尔, 易卜拉希米, 塔霍里。2014. 为实现低功耗, 在 STT-MRAM 避免不必要的写操作。《国际量子电动力学会议录》, 第 548-553 页。
- 米 (meter 的缩写) 比约林、庞奈、布格尼姆、达扬。2013. 阻塞设备接口的必要死亡。《CIDR 论文集》。
- 南博克, 查尔德斯, 梅尔赫姆, 莫斯, 张。2011. 分析无用回写对 PCM 主存续航和能耗的影响。《国际摄影测量与空间学会学报》, 第 56-65 页。
- J. Boukhobza。2013. 云中闪烁: 给 NAND 闪存存储系统带来一些启示。2013. 面向云环境的数据密集型存储服务, IGI 全球, 241-266。
- J. Boukhobza, P. Olivier, 2017 年。闪存集成, 爱思唯尔, ISBN 9781785481246。
- G. 伯尔、库迪、斯科特、林、戈帕拉克里希南、谢诺伊。2008. 存储级存储器的候选设备技术概述。52, 4 (2008), 449-464。
- J. 卡特, 2010 年。设计高效服务器和数据中心, 计算机, 43, 7 (2010), 76-78。
- A. m. 考尔菲尔德, j. 科伯恩, t. 莫洛夫, a. 德, a. 阿克尔, j. 何, a. 贾加塞桑, R. K. 古普塔, a. 斯纳夫利, s. 斯旺森。2010. 了解新兴非易失性存储器对高性能、IO 密集型计算的影响。《最高法院学报》, 第 1-11 页。
- J. 陈, 蒋瑞中, 黄海辉, 文卡塔拉马尼。2012. 对非易失性主存储器的能量感知写入, ACM SIGOPS 操作系统评论, 45, 3 (2012), 48-52。
- Y. 陈, 李, 张, 皮诺。2011a. 3D-HIM: 用于双极 RRAM 设计的 3D 高密度交错存储器。2011 年《纳米研究学报》第 59-64 期。
- Y. 陈, 李, 陈友友, 皮诺。2011b. 3D-ICML: 具有交错互补存储层的 3D 双极随机存取存储器设计。日期记录, 1-4。

- Y. 陈, 王文伟, 李洪海, 许春华, 张永年, 温文伟. 2013. 基于多级自旋转扭矩随机存取存储器单元的片上高速缓存及其优化. 技术. 电脑. 系统. 9, 2 (2013).
- R. 陈, 邵志忠, 李铁生. 2016. 弥合大数据工作负载的输入/输出性能差距: 一种基于 NVDIMM 的新方法. 在微学报, 1-12.
- H. 程等, 2016. 高效非对称末级缓存的循环块感知包含属性. 《ISCA 论文集》, 第 103-114 页.
- W. 钱钟书. 2010. WOX • RRAM 的单极开关行为. IEEE 电子设备快报, 31, 2 (2010), 126-128.
- B. J. Choi 等人, 2005 年. 《原子层沉积二氧化钛薄膜的电阻转换机制》, 应用物理, 98, 033715 (2005).
- 长度蔡美儿. 1971. 忆阻器——缺失的元素. 电路理论, 18, 5 (1971), 507-519.
- F. 克莱米迪、约万诺维奇、昂卡拉亚、奥赫什海赫、托马斯、图尔基伊尔马兹、维亚内洛、门户网站、米 (meter 的缩写)) 博克特. 2014. 电阻存储器: 哪些应用? . 日期记录.
- G. 阿尤布, 罗辛. 2009. 一个混合的 PRAM 和 DRAM 主存储系统. 在...里
发援会议事录, 664-469.
- 周一岳、蔡志勇、李德全、诺思. 2007. 利用非易失性内存提高闪存文件系统性能. 在 EMSOFT 学报, 164-173.
- W. 董等. 2015. 使用位翻转和循环移位最小化基于非易失性存储器的主存储器的更新位. 载于《HPC 会议记录》, 第 290-295 页.
- X. 董, 徐振华, 谢, N. P. Jouppi. 2012. NVSim: 新兴非易失性存储器的电路级性能、能量和面积模型. IEEE Trans. 论集成电路和系统的计算机辅助设计, 31, 7 (2012), 994-1007.
- X. 董, N. P. Jouppi, Y. Xie. 2013. 一种用于评估新兴存储器层次结构的电路架构协同优化框架. 《国际摄影测量与空间学会学报》, 第 140-141 页.
- Y. 杜, 周, 柴尔德斯, 莫塞, 梅尔赫姆. 2013. 平衡 PCM 单元编程的位映射. 《ISCA 论文集》, 第 428-439 页.
- 南杜罗尔、罗伊、赵、孙达拉姆、萨蒂什、桑卡兰、杰克逊、施万. 2016. 异构内存系统中的数据分层. 欧洲系统学报. 第十五条
- K. Eshraghian 等人, 2011 年. 忆阻器金属氧化物半导体内容可寻址存储器 (MCAM): 未来高性能搜索引擎的混合架构. 关于超大规模集成电路系统, 19, 8 (2011), 1407-1417.
- Everspin 技术公司, 2015 年. 检索自 2016 年 11 月 <https://www.everspin.com/64mb> 自旋转扭矩 MRAM DDR 3 dram 兼容.
- H. 藤井, 宫崎骏, 日本, 日本, 日本, 日本. 2012. 通过数据碎片抑制, x11 性能提升, x6.9 续航能力增强, 3D TSV 集成混合 ReRAM/MLC NAND 固态硬盘能耗降低 93%. 在 VLSIC 学报, 134-135.
- 米 (meter 的缩写)) 藤本, H. Koyama, Y. Hosoi, K. Ishihara, S. Kobayashi 2006. 二氧化钛/氮化钛纳米晶体薄膜的高速电阻转换, 日本. J. Appl. Phys. 45, 8/11 (2006), L310-L312.
- Y. 藤崎琴音. 2013. 新兴新型固态非易失性存储器综述, 日本应用物理杂志, 52 (2013).
- 南高, 何, 徐. 2015. 未来混合存储系统的实时内存检查点. 在...里
《国际科学学会学报》, 第 263-272 页.
- J. 吉本斯 • 比德尔. 1964. 固体电解质氧化镍薄膜的开关特性. 7, 11 (1964), 785-790.
- 名词 (noun 的缩写) 陶, 冰夷. 2013. 使用电阻存储器的吞吐量内核架构的功率性能协同优化. 《HPCA 论文集》, 第 342, 353 页.
- X. 郭, , 索亚塔, 2010. 电阻计算: 通过基于 STT- MRAM 的计算, 避免低泄漏的电源墙. 《ISCA 论文集》, 第 371-382 页.
- 南古鲁穆提. 2009. 云计算领域的存储架构, IEEE Micro, 29, 6 (2009), 68-71.
- A. 哈桑, H. Vandierendonck, D. S. Nikolopoulos. 2015. 混合内存层次结构上的高效内存数据存储. 在《达蒙论文集》中.
- Y. Hosoi 等人, 2006 年. 高速单极开关电阻随机存取存储器 (RRAM) 技术, IEDM 会议录, 30 . 7 . 1 - 30 . 7 . 4.
- Y. 怀. 2008. 自旋转移力矩 MRAM (STT-MRAM): 挑战和前景, 美国天文物理学会通报, 18, 6 (2008), 33-40.
- H. 黄福生, 姜天成. 2014. 基于闪存的 NVDIMM 的设计与实现. 《美国国家海洋与大气管理局学报》第 1-6 期.
- A. 贾迪迪, M. Arjomand, H. Sarbazi-Azad. 2011. 通过自适应线替换实现混合缓存架构的高耐久性和高性能设计. 《ISLPED 论文集》, 第 79-84 页.
- 长度姜, 张勇, 杨军. 2012. 基于相变存储器的低功耗长寿命多层陶瓷电容器的弹性复位. 《国际船舶和港口设施保安规则汇编》, 第 39-44 页.
- 长度姜, 张勇, 杨军. 2014. 减轻超密集相变存储器中的写干扰. 在...里
DSN 会议记录, 216-227.
- 长度姜, 赵 b, 杨军, 张勇. 2014. 一种低功耗、可靠的相变电荷泵设计

- 记忆。《ISCA 论文集》，第 397-408 页。
- Y. 金, 谢哈布, 荣格。2014. STT-MRAM 替代主存的面积、功耗和延迟考虑。在记忆论坛会议录中。
- A. Jog, A. K. Mishra, x. 丛, x. 袁, V. Narayanan, R. Iyer, C. R. Das。2012. 高速缓存恢复: 设计易失性 STT-RAM 高速缓存, 以提高 CMPs 的性能。发援会议事录, 第 243-252 页。
- 米 (meter 的缩写)) R. Jokar, M. Arjomand, H. Sarbazi-Azad。2016. 红杉: 一种基于高耐久性非易失性存储器的高速缓存体系结构, IEEE 超大规模集成电路系统交易, 24, 3 (2016), 954-967。
- Y. 朱德牛, 董晓东, 孙光国, 张恩昌, 谢永年。2010. 相变存储器高速缓存的能量和耐久性感知设计。在日期的程序, 136-141。
- J. 郑永元、金东、申洪全。2010. 利用混合文件系统中的存储类内存进行分层存储。关于存储 6, 1 (2010), 第 3 条。
- J. 中田荣格, 吉本, 川口。2013. 利用额外的全零数据标志的高能效自旋转扭矩内存缓存。《国际量子电动力学会议录》, 216-222 页。
- 米 (meter 的缩写)) 荣格、沙尔夫、坎德米尔。2013. 一个大规模存储级 RRAM 系统的设计。《国际科学学会会刊》第 103-114 页。
- H. Kanaya 等人, 2004 年。一种 $0.602 \mu\text{m}^2$ 的依假链单元结构, 通过一次掩模蚀刻工艺形成, 用于 64 Mbit FeRAM。在 VLSI 技术学报, 150-151。
- D. 康、白思豪、蔡俊杰、李德华、诺思和穆特鲁。2015. 非易失性存储器的遗忘高速缓存管理, 载于《MSST 论文集》, 第 1-13 页。
- 南坎南, 加夫里洛夫斯卡, 施万。2016. pVM: 持久虚拟内存, 用于高效的容量扩展和对象存储。欧洲系统学报。第十三条
- Y. 加藤, 山田, 岛田。2005. 使用电荷补偿技术的 $0.18\text{-}1\text{m}$ 无损读出铁电存储器。选举。戴夫。52, 12 (2005), 2616-2621。
- K. 金, 安圣杰。2005. 可制造高密度 PRAM 的可靠性研究, 可靠性。在...里《IRPS 论文集》, 第 157-162 页。
- K. 金。2008. 未来存储技术: 挑战与机遇。《超大规模集成电路学报》, 第 5-9 页。
- H. 金, 塞沙德里, 迪基, 赵。2014. 评估企业存储系统的相变存储器: 缓存和分层方法的研究。存储 10, 4 (2014), 第 15 条。
- Y. Kim 等人, 2011 年。双层 RRAM, 续航能力无限, 切换极为均匀。在...里 VLSIT 会议录, 第 52-53 页。
- 页 (page 的缩写) Kogge 等人, 2008 年。数据库云服务器计算研究: 实现数据库云服务器系统的技术挑战, 美国国防高级研究计划局信息处理技术办公室。
- 米 (meter 的缩写)) P. Komalan 等人, 2013 年。用于嵌入式平台的基于 NVM 的混合指令存储器组织的设计探索, 嵌入式系统的设计自动化, 17, 3-4 (2013), 459-483。
- 米 (meter 的缩写)) 科马兰、戈麦斯·佩雷斯、坦利亚多、拉加万、哈特曼、卡特霍尔。2014. 通过 MSHR 增强技术探索基于非易失性存储器的高速缓存的可行性。在日期的程序中, 第 21 条。
- 米 (meter 的缩写)) H. Kryder, C. S. Kim。2009. 硬盘之后——接下来是什么?, IEEE Trans. 关于磁学, 45, 10 (2009), 3406-3413。
- E. Kultursay, M. Kandemir, A. Sivasubramaniam, O. Mutlu。2013. 评估 stt-ram 作为一种节能的主存储器替代方案。在《国际船舶和港口设施保安规则》第 256-267 页。
- K. 具有冗余块的非对称写体系结构: 一种高写速度的 STT-MRAM 高速缓存体系结构。关于超大规模集成电路系统, 22, 4 (2014), 712-720。
- B. 李, 伊佩克, 穆特鲁, 伯格。2009. 将相变存储器设计为可扩展的 dram 替代方案。《ISCA 论文集》, 第 2-13 页。
- H. 李等人, 2010 年。亚纳秒开关速度和高耐久性的基于 HfOx 的电阻存储器过复位问题的证据和解决方案。《IEDM 论文集》, 19. 7. 1-19. 7. 4。
- D. Lee 等人, 2006 年。用于非易失性电阻存储器应用的掺杂二元金属氧化物的优异的均匀性和可再现的电阻切换特性。《IEDM 论文集》, 30. 8. 1-30. 8. 4。
- E. 李, 洪恩, 柳先生, 诺先生。2014. 非易失性存储器存储的实证研究: 一个操作系统的视角和启示。在《吉祥物学报》中。405-410。
- 南李, 亨, 诺。2014. 《时钟-DWF: 混合相变存储器和动态随机存取存储器结构的写历史感知页面替换算法》, IEEE Trans. 计算机, 63, 9 (2014), 2187-2200。
- D. 李、维特尔、马林、麦考迪、西拉、刘、于。2012. 识别极端规模科学应用中字节可寻址非易失性存储器的机会。《国际公共政策学会会报》, 第 945-956 页。
- J. 李, 薛春华, 徐玉英。2011. 基于 STT-RAM 的 CMPs 能效混合缓存。在超大规模集成电路-系统芯片学报, 31-36。
- X. 李, 卢 k, 王, 周。2012. NV-process: 一种基于非易失性存储器的容错流程模型。《亚太空间科学协会会议录》第 1 条。
- Y. 李, 陈友友, 琼斯。2012. 一种对抗非易失性存储器不对称的软件方法。

- 《ISLPED 论文集》，第 191–196 页。
- 李清，李建利，石立林，赵明敏，薛建中，何玉英。2014. 面向高效嵌入式系统的编译器辅助的基于 STT-RAM 的混合高速缓存，IEEE 超大规模集成电路系统事务，22，8 (2014)，1829–1840。
- Y. 刘，周春华，程晓明，带 PCM 的混合固态硬盘。2011. 在《美国国家医学科学院院刊》第 1–5 期。
- J. 米娜，斯泽，吕德，曾。2014. 新兴非易失性存储器技术概述，纳米尺度研究快报，9，526 (2014)，1–33。
- F. 苗等，2011. 纳米级传导通道的解剖揭示了高性能忆阻器的机制。23 (2011) 5633–5640。
- A. 米尔霍西尼，波特魔芋，库尚法尔。2012. 基于编码的相变存储器能量最小化。发援会议事录，第 68–76 页。
- 南米塔尔。2013. 使用高速缓存着色来减轻非易失性高速缓存中的组间写变化，爱荷华州立大学理工学院。有代表性的
- 南米塔尔，维特尔。2015a. AYUSH: 延长静态随机存取存储器和非静态随机存取存储器混合高速缓存寿命的技术，IEEE 计算机体系结构通讯，14，2 (2015)，115–118。
- 南米塔尔，维特尔，李。2015b. 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。关于并行和分布式系统，26，6 (2015)，1524–1537。
- 南米塔尔，维特尔。2016. 将非易失性存储器用于存储和主存系统的软件技术综述，并行和分布式系统上的 IEEE 事务，27，5 (2016)，1537–1550。
- 南米塔尔。2016. 使用域墙内存构建处理器组件的技术综述。J. Emerg. 技术。电脑。系统。第 13 条第 2 款 (2016 年)，第 29 条。
- J. c . 莫卧儿，e . 阿尔戈洛，m . 沙阿，p . 法拉博希。2009. 操作系统支持 NVM+DRAM 混合主存。摄影学报。
- O. 穆特鲁。2015 年主内存扩展：挑战和解决方案方向。《下一代计算机设计的超越摩尔技术》一书的第 6 章，斯普林格出版社，127–153 页。
- D. 牛，徐春华，穆拉利马努哈，朱皮，谢。2012. 高密度交叉点电阻存储器的设计权衡。《ISLPED 论文集》，第 209–214 页。
- F. 奥博里尔，比什诺伊，易卜拉希米，塔霍里。2015. 使用片上高速缓存层次结构的软件 MRAM 的混合存储器技术的评估。论集成电路和系统的计算机辅助设计，34，3 (2015)，367–380。
- 南小川。2014. 非易失性主内存上的独立内核/进程检查点，用于快速内核恢复。在 ARCS 学报，233–244。
- X. 欧阳，内兰丝，威普费尔，布林，潘达。2011. 超越块输入/输出：重新思考传统存储原语。《HPCA 论文集》，第 301–311 页。
- J. 朴德信，李海光。2015. 具有 LPDDR2–NVM 接口的相变存储器行缓冲结构的设计空间探索。在超大规模集成电路–系统芯片学报，104–109。
- 南朴永金，乌尔高卡，李杰，徐。2011. 闪存存储能效综合研究，系统架构杂志，57，4 (2011)，354–365。
- 南帕克、古普塔、莫朱德尔、拉古纳坦、罗伊。2012. 使用 STT 磁流变液提高能效的未来高速缓存设计：器件、电路和架构。发援会议事录，492–497。
- Y. Park, S. K. Park, K. H. Park, Linux 内核支持利用相变存储器。2010. Linux 研讨会论文集，217–224。
- 米 (meter 的缩写)) 库雷希，斯里尼瓦桑，里弗斯。2009. 使用相变存储器技术的可扩展高性能主存储器系统。《ISCA 论文集》，第 24–33 页。
- 米 (meter 的缩写)) 拉斯特拉-蒙塔诺。2010. 通过写取消和写暂停提高相变存储器的读性能。《HPCA 论文集》。
- 米 (meter 的缩写)) K. Qureshi, S. Gurumurthi, B. Rajendran。2011. 相变存储器：从设备到系统，摩根和克莱普尔，计算机体系结构综合讲座。
- 页 (page 的缩写) 阮冈纳赞。2011. 从微处理器到纳米存储：重新思考以数据为中心的系统，计算机，44，1 (2011)，39–48。
- 米 (meter 的缩写)) Rasquinha, D. Choudhary, S. Chatterjee, S. Mukhopadhyay, S. Yalamanchili。2010. 使用自旋扭矩转移 (STT) 随机存取存储器的高能效高速缓存设计。《ISLPED 论文集》，第 389–394 页。
- D. 罗伯特，t . 凯格勒，t . 穆奇 2009. 将与非门闪存设备集成到服务器上，美国计算机学会通讯，52，4 (2009)，98–106。
- 单位 Russo, D. Ielmini, A. Redaelli, A.L. Lacaita。2008. 相变存储器中编程和读取性能的建模第一部分：单元优化和缩放。电气和电子工程师协会电子设备学报，55，2 (2008)，506–514。
- R. 萨阿迪·萨尔科德。2016. 一种混合动态随机存取存储器结构的操作系统级数据迁移方案。936–941。
- 米 (meter 的缩写)) H. Samavatian, H. Abbasitabar, M. Arjomand, H. Sarbazi-Azad。2014. 一种高效的图形处理器 STT-RAM 末级缓存架构。发援会议事录，1–6。

- G. 桑德胡. 2013. 新兴记忆技术领域. 在《美国国家医学科学院院刊》第 1-5 期。
- 南谢切特, 罗赫, 斯特劳斯, 伯格. 2010. 使用 ECP, 而不是 ECC, 来解决电阻存储器的硬故障. 阿奇特. 新闻 38, 3 (2010), 141-152.
- 南尼森、托里斯、萨萨特利、布托、穆萨德. 2014. 多核架构下基于磁随机存取存储器的存储器层次结构探索. 在 ISVLSI 学报, 248-251.
- 南尼森、托雷斯、萨萨特利、加马提、穆萨德. 2015. 面向处理器架构的新兴非易失性存储器技术探索流程. 在 ISVLSI 学报, 460-460.
- 南 Sheu 等人, 2011 年. 一个 4Mb 嵌入式 SLC 电阻式随机存取存储器宏, 具有 7.2ns 的读写随机存取时间和 160ns 的多层随机存取能力. 《ISSCC 论文集》, 第 200-202 页。
- H. Shiga 等人, 2009 年. 一个 1.6 GB/s 的 DDR2 128 Mb 链 FeRAM, 具有可扩展的八通道位线和传感方案. 《ISSCC 论文集》, 第 464-465 页。
- A. 希洛夫. 2016. 西部数字将使用 3D ReRAM 作为专用固态硬盘的存储级内存. ANANDTECH.
- C. W. Smullen, V. Mohan, A. Nigam, S. Gurumurthi, M. R. Stan. 2011. 为快速且节能的 STT-RAM 缓存放宽非易失性. 《HPCA 论文集》, 第 50-61 页。
- SNIA, 2015, 非易失性存储器编程模型 (NPM), 检索自 2017 年 5 月 http://www.snia.org/网站/默认/文件/technical_work/final/NMProgrammingModel_v1.1.1.pdf。
- 斯特科夫, 斯奈德, 威廉 2008. 失踪的忆阻器发现, 自然, 453 (2008), 80-83.
- G. 孙, 董晓东, 谢玉英, 李军, 陈玉英. 2009. 一种新颖的 CMPs 3D 堆叠 MRAM L2 缓存架构. 《HPCA 论文集》, 第 239-249 页。
- G. 孙, 朱友友, 陈友友, 牛, 谢友友, 陈友友, 李. 2010. 一种混合固态存储体系结构, 可提高性能、能耗和寿命. 2010. 《HPCA 论文集》, 第 1-12 页。
- G. 孙, 牛, 欧阳俊杰, 谢. 2011. 一种基于频繁值的 PRAM 内存架构. 在...里 ASP-DAC 会议录, 211-216。
- H. 孙, 宫崎骏, 柔口骏, 竹内骏. 2014. 3D-TSV 混合随机存取存储器/多层存储与非固态硬盘的高性能和高能效冷数据驱逐算法. 《电路与系统》第一卷, 第 61 卷, 第 2 期 (2014 年), 第 382-392 页。
- Z. 孙、毕晓明、李洪浩、王文荣、朱晓明、吴文伟. 2011. 采用动态刷新方案的多保留级别 STT-RAM 缓存设计. 《微学报》, 329-338 页
- Z. 孙, 贾, 蔡, 张, 居. 2015. AIMR: 具有非易失性存储器和动态随机存取存储器的混合存储器结构的自适应页面管理策略. 载于《HPCC 会议记录》, 第 284-289 页。
- A. 苏雷什、西科蒂、卡林顿. 2014. 针对高性能计算、数据密集型应用的新兴内存技术评估. 《集群学报》, 239-247 页。
- 南部逸夫, 林毅夫. 2013. 具有缓存分区和访问感知策略的 CMP 架构中的高耐久性混合缓存设计. 《GLSVLSI 学报》, 19-24 页。
- 南 Tanakamaru, M. Doi, K. Takeuchi. 2014. 与非门闪存/随机存取存储器混合统一固态存储架构. 《电路与系统》第一卷, 第 61 期, 第 4 期 (2014 年), 第 1119-1132 页。
- G. 呃, 王永年, , 金图尔卡尔, 伯恩斯, 诉曹. 1999. 有效利用零开销循环缓冲区. 在 LCTES 会议录, 10-19.
- H. 辛格·文卡特什. 2015. 框架适合像手套一样的可穿戴电子产品, 电子设计. 检索于 2016 年 11 月 <http://electronicdesign.com/记忆/框架-适合-可穿戴-电子-手套>。
- J. 维特尔, 米塔尔. 2015. 非易失性存储器系统在超大规模高性能计算中的机遇, 科学与工程计算, 17, 2 (2015), 73-82.
- J. 王, 董晓明, 谢, N. P. Jouppi, 2013. i2WAP: 通过减少组间和组内写变化来提高非易失性高速缓存的寿命. 《HPCA 论文集》, 第 234-245 页。
- J. 王, 董晓明, 谢, N. P. Jouppi. 2014. 非易失性高速缓存的耐久性感知高速缓存行管理. 阿奇特. 代码 Optim. 11, 1 (2014a)。
- J. 王, 董, 谢. 2014. 建立和优化基于 MRAM 的商品记忆. 阿奇特. 代码 Optim. 11, 4 (2014b)。
- R. 王, 姜立军, 张勇, 王立军, 杨军. 2015a 选择性恢复: 未来 STT-MRAM 的高能效读取干扰缓解方案. 发援会议事录, 第 21 条。
- R. 王, 姜立军, 张勇, 杨军. 2015b. SD-PCM: 在写入干扰下构建可靠的超密集相变存储器. 在美国公共科学图书馆学会会报, 19-31。
- R. 王, 姜立军, 张勇, 王立军, 杨军. 2015c. 利用不平衡单元写入来减轻超密集相变存储器中的写入干扰. 发援会议事录, 第 88 条。
- Z. 王, 吉梅内斯, 丛晓军, s. 光宇, 谢. 2014. 基于 STT-RAM 的混合缓存的自适应放置和迁移策略. 《HPCA 论文集》, 第 13-24 页。
- Z. 王, 顾, 姚, 邵. 2015. 实时嵌入式系统中基于 NVM 的暂存存储器上数据变量的耐久性感知分配, 集成电路和系统的计算机辅助设计的 IEEE 交易, 34, 10 (2015), 1600-1612。
- 魏, 陈, 陈. 2015. 使用字节可寻址非易失性存储器加速文件系统元数据访问, 存储上的 ACM 事务 11, 3 (2015), 第 12 条。

39:34 J. Boukhobza 等人。

- W. 魏, 姜德伟, 麦凯思, 熊俊杰, 陈明敏. 2015. 利用程序语义将数据放入混合存储器. 在《契约学报》上. 163-173.
- R. 威廉. 2008. 我们如何找到丢失的忆阻器, IEEE 光谱, 45 (2008), 28-35.
- R. 冬天. 2008. 为什么数据库增长如此之快? 检索自 2016 年 11 月 <http://www.b-eye-network.com/channels/1138/view/7188>
- W. 黄. 2016. FRAM 为 16 位微控制器提供统一内存. 电子设计.
- J. 吴等, 2015. 使用新型传感方案的超高密度相变存储器的大于 2 位/单元的多层存储. 在超大规模集成电路技术学报, T94-T95.
- 页 (page 的缩写) 吴、李德全、陈泽楷、维特尔、米塔尔. 2016. 明确管理的非易失性存储器中的算法导向数据放置. 《HPDC 论文集》, 第 141-152 页.
- X. 吴, 李军, 张立军, 斯佩埃特, 拉杰曼尼, 谢. 2009. 采用不同内存技术的混合缓存架构. 《ISCA 论文集》, 第 34-45 页.
- F. 夏, 姜德生, 熊军, 孙. 2015. 相变存储系统综述, 计算机科学与技术杂志 30, 1 (2015), 121-144.
- C. 徐, 董 x; 谢. 2011. 基于忆阻器的 RRAM 交叉点结构的设计含义. 日期记录.
- C. 徐, 牛, 穆利马诺哈, 朱皮, 谢. 2013. 理解多级单元 ReRAM 存储器设计中的权衡. 发援会议事录, 第 108 条.
- C. 徐, 牛德生, 穆拉利马努哈, 巴拉苏布拉蒙尼安, 张, 于, 谢, 2015. 克服交叉电阻存储器架构的挑战. 《HPCA 论文集》. 476-488.
- C. 徐, 牛德生, 郑玉生, 谢玉生. 2014. 可靠性感知的交叉点电阻存储器设计在...里
《GLSVLSI 学报》, 第 145-150 页.
- W. 徐, 刘军, 张. 2009. 降低相变存储器写能量的数据处理技术. 《国际船舶和港口设施保安规则汇编》, 第 237-242 页.
- C. 薛军, 张勇, 陈勇, 孙刚, 杨俊杰, 李海峰. 2011. 新兴的非易失性记忆: 机遇与挑战. 《法典汇编》第 325-334 页, ISSS.
- B. 杨振宁, 李振英, 金振实, 赵振华, 李振英, 俞炳国. 2007. 一种使用数据比较写入方案的低功耗相变随机存取存储器. 在国际标准学会会报, 3014-3017.
- J. 杨, 威廉姆斯. 2013. 计算系统中的忆阻设备: 承诺与挑战
J. 紧急情况. 技术. 电脑. 系统. 第九条、第二条 (2013 年), 第十一条.
- J. 杨振宁, 苗, 皮克特, 奥赫伯格, 斯图尔特, 刘振宁, 威廉姆斯. 2009. 金属氧化物忆阻开关的电铸机理, 纳米技术, 20 215201 (2009).
- J. 杨等. 2010. 粘附层金属的扩散控制纳米尺度忆阻开关, 高级材料, 22, 36 (2010), 4034-4038.
- 南杨海峰, 柳永胜. 2013. 传感器节点中基于 STT-MRAM 混合主存的缓冲管理. 载于《国际刑事法院学报》, 第 286-289 页.
- 南亚兹丹森纳斯、皮巴斯提、法泽利、帕图格. 2014. 为高耐久性和低功耗编码最后一级 STT-RAM 高速缓存, 计算机体系结构快报, 13, 2 (2014), 73, 76.
- J. 尹先生、南先生、尹俊成先生、金先生、金先生、闵先生、赵先生. 2008. 变色龙: 高性能闪存/FRAM 混合固态硬盘架构, 计算机架构快报, 7, 1 (2008), 17-20.
- K. Yoongu, R. Daly, J. Kim, C. Fallin, L. Ji Hye, L. Donghyuk, C. Wilkerson, K. Lai, O. Mutlu. 2014. 翻转内存中的位而不访问它们: 对动态随机存取存储器干扰错误的实验研究. 《ISCA 论文集》, 第 361-72 页.
- 南俞, 陈炳炎, 新兴存储技术: 近期趋势与展望. 2016. IEEE 固态电路杂志, 8, 2 (2016), 43-56.
- J. 岳, 朱, 利用相变存储器的不对称性加速写操作. 2013a. 《HPCA 论文集》, 第 282-293 页.
- J. 岳, 朱允炆. 2013b. 利用存储体内部的子阵列提高相变存储器性能. 在日期的程序, 386-391.
- J. 云, 李少生, 刘少生. 2012. 基于布隆过滤器的相变存储器动态磨损均衡. 在...里
日期: 1513-1518.
- 张立昌, 布塔巴. 2010. 云计算: 最新技术和研究挑战, 互联网服务和应用杂志, 1, 1 (2010), 7-18.
- 页 (page 的缩写) 周, 赵, 杨, 张. 2009a. 一种使用相变存储技术的耐用且节能的主存储器. 《ISCA 论文集》, 第 14-23 页.
- 页 (page 的缩写) 周, 赵, 杨, 张. 2009b. 使用提前写终止降低 STT-RAM 的能耗. 《ICCAD 论文集》, 第 264-268 页.

2007 年 2 月收到; 2009 年 3 月修订; 2009 年 6 月接受