基于相变存储器的存储技术研究综述

冒 伟 刘景宁 童 薇 冯 丹 李 铮 周 文 张双武
(华中科技大学武汉光电国家实验室/计算机科学与技术学院 武汉 430074)

摘 要 以数据为中心的大数据技术给计算机存储系统带来了机遇和挑战.传统的基于动态随机存储器(DRAM)器件的内存面临工艺尺寸缩小至2X nm及以下所带来的系统稳定性、数据可靠性等问题;相变存储器(PCM)具有非易失性、存储密度高、功耗低、抗辐射干扰等优点,且读写性能接近DRAM,是未来最有可能取代DRAM的非易失存储器,它为存储系统的研究和设计提供了新的解决方案.文中在归纳相变存储器器件发展和研究现状的基础上,对相变存储器在系统级的应用方式和面临的问题进行了比较和分析,研究了基于相变存储器的内存技术和外存技术,分析了当前在PCM的寿命、写性能、延迟、功耗等方面所提出的解决方案,指出了现有方案的优势和面临的缺陷,并探讨了未来的研究方向,为该领域在今后的发展提供了一定的参考.

关键词 相变存储器;非易失存储器;存储技术;计算机体系结构
中图法分类号 TP303; TP333 **DOI号** 10.3724/SP.J.1016.2015.00944

A Review of Storage Technology Research Based on Phase Change Memory

MAO Wei LIU Jing-Ning TONG Wei FENG Dan LI Zheng ZHOU Wen ZHANG Shuang-Wu
(*Wuhan National Laboratory for Optoelectronics / School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan* 430074)

Abstract Data-centric big data technology brings opportunities and challenges for computer storage system. Traditional memory based on Dynamic Random Access Memory (DRAM) faces problems of system stability, data reliability caused by shrinking craft size to 2X nm and below. Phase change memory (PCM) has advantages of high storage density and low power consumption. Furthermore, it is non-volatile, anti-radiation and anti-interference, and its read/write performance is close to DRAM. All these make PCM the most promising candidate for DRAM. It offers a new solution to research and design of storage system. In this paper, we first summarized development history and related work of PCM, then analyzed application mode and problems of PCM on system level. Finally, with a study of memory/storage technologies based on PCM and existing solutions on lifetime, write performance, latency, energy consumption, we pointed out advantages and disadvantages of current technologies and predicted the directions of further research, and thus could be reference work for the development of the field in future.

Keywords phase change memory; non-volatile memory; storage technology; computer architecture

收稿日期:2014-05-14;最终修改稿收到日期:2014-11-26.本课题得到国家"九七三"重点基础研究发展规划项目基金(2011CB302301)、国家自然科学基金(61303046,61173043)、国家杰出青年科学基金(61025008)和中央高校基本科研业务费(HUST:2013TS042)资助.
冒 伟,男,1989年生,硕士研究生,中国计算机学会(CCF)会员,主要研究方向为计算机存储系统、新型存储器件、固态存储. E-mail: morewell@hust.edu.cn. **刘景宁**,女,1957年生,博士,教授,中国计算机学会(CCF)会员,主要研究领域为计算机系统结构、计算机存储系统及高速通道接口技术. **童 薇**(通信作者),女,1977年生,博士,讲师,中国计算机学会(CCF)会员,主要研究方向为海量存储系统、固态存储、I/O虚拟化. E-mail: tongwei@hust.edu.cn. **冯 丹**,女,1970年生,博士,教授,中国计算机学会(CCF)会员,主要研究领域为信息存储系统、网络存储、固态存储、性能评价. **李 铮**,男,1992年生,博士研究生,中国计算机学会(CCF)会员,主要研究方向为新型存储器件、固态存储. **周 文**,男,1985年生,博士研究生,中国计算机学会(CCF)会员,主要研究方向为可重构计算和新型存储器件. **张双武**,男,1990年生,硕士研究生,中国计算机学会(CCF)会员,主要研究方向为计算机存储系统、固态存储、新型存储器件.

1 引　言

信息技术的高速发展将人类社会带入数字时代，人们创造、捕获和复制的信息无处不在，构成规模巨大且不断扩张的“数字宇宙”，其规模在 2012 年为 2.8ZB，预计 2020 年将达到 40ZB. 计算机领域已经进入一个以数据为中心的大数据时代，数据的快速增长和以数据为中心的发展趋势给现有的计算机存储系统带来了机遇和挑战.

大数据所催生的内存计算和拥有越来越多核心的处理器对内存的速度、容量、功耗和可靠性提出了极高的需求，在过去的十年里，内存技术已由 DDR2 (Double Data Rate 2)发展至现在的 DDR4(Double Data Rate 4). 然而内存技术的发展已遇到瓶颈：现有的 DRAM(Dynamic Random Access Memory，动态随机存储器)尺寸已经到达其制造工艺的极限，同时数据刷新所产生的功耗问题随着其容量扩大日益严重，随着内存技术继续发展，工艺尺寸进一步降低，电子的微观特性将越来越明显，加上器件本身的物理特性制约等因素，传统的 DRAM 介质在系统稳定性、数据可靠性等问题上将面临困境.

新型的非易失存储器(Non-Volatile Memory，NVM)的出现，为扩展计算机内存提供了新的途径，同时推动了计算机系统结构的改变. 现有的 NVM 有 PCM(Phase Change Memory，相变存储器)，STT-RAM(Spin Transfer Torque Random Access Memory，自旋转移矩磁随机存储器)，MRAM(Magnetic Random Access Memory，磁性随机存储器)，FeRAM(Ferroelectric Random Access Memory，铁电随机存储器)，RRAM(Resistive Random Access Memory，阻变随机存储器)等，表 1 从产品容量、工艺尺寸、读写性能、寿命、功耗及当前技术瓶颈等多个方面展示、对比了各存储技术的特点.

表 1　存储技术参数(典型值)对比

参数	现有芯片容量级别	理论工艺制程级别	特征尺寸[1]	读操作时间	写操作时间	寿命(耐久性)	数据保持力	写操作功耗[2]	空闲功耗	非易失性质	读过程破坏性	当前主要技术瓶颈
DRAM	~16 Gb	~20 nm	6~10 F^2	<10 ns	<10 ns	>10^{15}	刷新	~0.1 nJ/b	高	易失	破坏性	需刷新，易失，作为内存工艺制程有限
NAND	~1 Tb	~16 nm	4~11 F^2	10~50 μs	0.1~1 ms	10^4~10^6	10 年	0.1~1 nJ/b	低	非易失	非破坏	寿命/性能有限，存储密度较低
STT-RAM	~64 Mb	~32 nm	16~60 F^2	2~20 ns	5~35 ns	10^{12}~10^{15}	>10 年	1.6~5 nJ/b	低	非易失	非破坏	容量小，写功耗较大，稳定性差
RRAM	~1 TB	~11 nm	4~14 F^2	10~50 ns	10~50 ns	10^8~10^{10}	10 年	~0.1 nJ/b	低	非易失	非破坏	材料级存储机理尚不明确
FeRAM	~64 MB	~65 nm	15~34 F^2	20~80 ns	5~10 ns	10^{12}~10^{14}	10 年	<1nJ/b	低	非易失	破坏性	容量小，具有读破坏性，存储密度低
PCM	~8 Gb	~5 nm	4~8 F^2	10~100 ns	20~120 ns	10^8~10^{12}	>10 年	<1 nJ/b	低	非易失	非破坏	容量较小，材料可操作温度范围狭窄

从表 1 中可以看出，与 DRAM 相比 PCM 具有非易失性，可以长期保留数据，工艺制程低，存储规模可扩展性强，不需要刷新，静态功耗低，这些使其在未来内存环境中的应用拥有一定的优势；同时也有不足，如 PCM 具有相对较大的延迟，写速率无法和 DRAM 比拟，寿命也相对有限. 与 NAND Flash 相比，PCM 写性能优秀、寿命长、具有位修改特性，这使其在系统中可以保持高速运行状态而不发生掉速现象，同时可降低存储控制器及缓存开销，相对于 NAND Flash 而言更加适合外存的应用环境；当然就目前而言，PCM 的产品容量和存储密度还没有 NAND Flash 大，还无法完全替代 NAND Flash 在外存系统中的地位. 与其他新型 NVM 介质相比，PCM 在持久性、功耗或者容量等方面更为优秀，研究相对更加成熟，产业化程度更高，作为各类新型非易失存储器的竞争者具有明显的优势，在 45 nm 工艺制程下已经有 Gb 级别产品问世，且部分产品已经成熟应用于智能电表、移动终端等设备中.

相变存储器是一种由硫族化合物材料构成的新型非易失存储器，它利用材料可逆转的物理状态变化来存储信息，具有非易失性、工艺尺寸小、存储密度高、循环寿命长、读写速度快、功耗低、抗辐射干扰等优点. 相变存储器介质材料在一定条件下会在非晶体状态和晶体状态之间发生转变，在此过程中的非晶体状态和晶体状态呈现出不同的电阻特性和光学特性，因此，可以利用“0”和“1”分别表示非晶态和

晶态来存储数据[1]. PCM 写“1”是一个中温结晶的过程,即对硫族化合物施加一个时间较长、强度中等的电脉冲进行加热,使其温度上升到结晶温度以上、溶化温度以下,从而导致结晶,这一过程也被称之为 SET 过程;PCM 写“0”是一个高温淬火的过程,即使用一个强度很高但作用时间很短的电脉冲,使得硫族化合物材料在温度上升到熔点之上后迅速经历一个淬火(热量快速释放,降温速率大于 10^9 K/s)过程,材料将由熔融状态进入非晶态,这个过程被称之为 RESET 过程[3]. 综上可以看出 PCM 写“1”和写“0”是不对称的,写“1”变化慢,需要的电流小,写“0”变化快,需要的电流大,如图 1 所示(图中坐标为参考值).

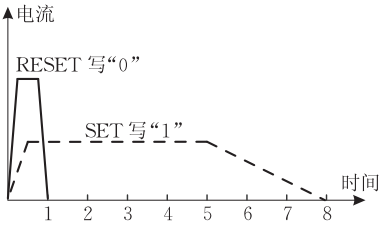

图 1 PCM 写“0”与写“1”的不对称性

由于 PCM 特性优良,是未来最有可能取代 DRAM 的非易失存储器之一,目前国内外已有许多厂商和科研机构进行相关研究,这些研究可分为以下几类:(1)研究 PCM 作为计算机系统内存:相对于 DRAM 而言,PCM 写延迟约高 10 倍,但它具有非易失性,存储密度和功耗也具有一定优势,这使得它可以在需求大容量、低功耗和非易失内存等特性的应用场景中替代 DRAM[4-9];(2)研究 PCM 作为计算机系统外存:PCM 在速度、寿命和功耗上均超过了 Flash 技术,可以作为外存使用;同时 PCM 工艺尺寸较低,这些使得它能够满足海量存储系统中高速率、低功耗、大容量的需求[10-11];(3)研究 PCM 在新环境下的存储应用,如利用 PCM 的非易失性和良好的读写性能而将其应用到检查点(Checkpoint)信息[12]、数据库应用环境中的数据恢复[13-14]等;(4)研究 PCM 应用于特殊环境的嵌入式系统,如智能电表、音频存储[15]等,且由于其优良的抗辐射性能而在航天航空电子领域亦有应用[11];(5)研究 PCM 在移动智能终端领域的应用:美光公司生产的 45 nm 制程 LPDDR2(Low Power DDR2)接口的 PCM 芯片已经应用在手机上,2013 年诺基亚将美光 1 Gb+512 Mb 的 PCM 芯片用于 Asha 系列手机中;三星 GT-E2550 GSM 手机也已经部署了自己的 PCM 芯片.

2 PCM 在计算机系统应用中的问题和挑战

随着相变存储器技术的发展,针对相变存储器器件本身的研究和其在计算机存储系统中的应用研究已经在产业界与学术界陆续展开,从而解决 PCM 所面临的问题和挑战.

2.1 PCM 器件级国内外研究现状

针对 PCM 器件本身的研究,国内外具有一定的差距,国外发展迅速,国内起步较晚;同时立足于产业界、学术界的研究也一直在跟进.

(1)国外研究

20 世纪末,半导体工艺技术发展到纳米量级,给相变存储器的技术特点和优势提供了发挥的空间,国外很多半导体行业公司以及研究机构都相继投入了大量的人力、物力和财力,加入了相变存储器的研究阵营[16],其技术日趋成熟.

2000 年 2 月,Intel(英特尔)和 Ovonyx 发布合作和许可协议,开启了 PCM 产品的研发;同年 12 月,ST(意法半导体)与 Ovonyx 开始合作. 2007 年 ST 和 Intel 宣布成立 Numonyx(恒忆,现已被美光收购),专门致力于 PCM 的研发,PCM 进入了快速发展期. 2007 年 Hynix(海力士,现更名为 SK Hynix)成功开发出 40 nm 工艺 1 Gb 容量的 PCM 芯片,但由于成本和市场接受度等众多原因,该芯片未能进入量产. 2009 年 12 月,恒忆宣布量产 1 Gb 的相变存储器产品;2010 年恒忆发布全新系列相变存储器产品,该系列产品具有更高的写性能、写寿命和设计简易性,适用于无线通信设备、消费电子和其他嵌入式应用设备. 2011 年 6 月 IBM 攻克了相变存储的一大难题:多位封装,采用 90 nm CMOS 工艺制造,实现了每单元存储多位数据,写入和检索数据速度比 NAND Flash 存储器快 100 倍,并可持续使用一千万次. 2012 年 6 月,IBM 和 SK Hynix 宣布联合打造新一代相变存储器.

在市场上,Samsung(三星)与 Micron(美光)是目前在 PCM 技术方面较为领先的两家公司,其中三星开发出的 65 nm 制程、512 Mb 容量的 PCM 芯片已投入量产并应用在三星的手机存储卡中;同时三星已经推出了 20 nm 制程、8 Gb 容量的相变内存颗粒. 美光则是通过收购 Numonyx 而得到了这家

公司开发出的 90 nm 制程、128 Mb 容量的 PCM 技术，并成功研制了面向移动设备领域的 45 nm 制程、1 Gb 容量的 LPDDR2 接口的 PCM 芯片产品，于 2012 年 7 月宣布量产该系列芯片.

如今，Samsung、Micron、Intel、IBM 等多家知名半导体公司投巨资来推进 PCM 的产业化发展，同时斯坦福大学、卡内基梅隆大学、加州大学圣地亚哥分校、惠普实验室、微软研究院等学术研究机构也针对 PCM 从器件到应用展开了广泛研究，推动了 PCM 芯片在存储系统的应用，PCM 的发展速度前所未有.

(2) 国内研究

与国外相比，国内因工艺条件的限制对 PCM 的研究相对少一些. 国内目前对 PCM 技术的研究机构主要有中国科学院上海微系统与信息技术研究所、华中科技大学、中芯国际、北京时代全芯科技等.

中国科学院上海微系统与信息技术研究所对相变材料体系有比较系统的研究，包括相变材料、相变机理、高速 CMOS 衬底材料、纳米加工工艺集成、PCM 芯片设计与制作技术等，其与中芯国际联合搭建了 130 nm 和 45 nm 的 PCM 芯片工艺平台[1]，并已经着手开展 40 nm 工艺上的进程，同时也和美国 SST 公司联合研发 PCM 技术. 2011 年 4 月，中国科学院上海微系统与信息技术研究所成功开发出中国第一款具有自主知识产权的相变存储器芯片，这款 PCM 芯片的存储容量为 8 Mb，具有读、写、擦功能.

华中科技大学自 2007 年开始研究高密度低功耗电阻式相变存储器、相变存储器功能芯片、相变存储器芯片的关键材料以及相关专用测试设备等，已经自主研制出具有简单读、擦、写功能的相变存储器功能芯片.

2011 年 9 月，北京时代全芯科技公司在与美国全芯科技公司及其合作方 IBM 团队的共同努力下，设计完成第一批基于相变存储器的产品芯片，成为我国第一家生产高密度相变存储器芯片的公司. 该公司已经成功设计了 256 Mb 的 LPDDR2 接口 PCM 芯片和 32Mb 的 SPI(Serial Peripheral Interface)接口芯片，并设计了一个 16 Mb 的嵌入式相变存储器宏模块.

2013 年 1 月，北京时代全芯科技公司投资 47.5 亿美元在宁波开建我国首个拥有自主知识产权的存储芯片产业化应用项目，2014 年 3 月宁波 PCM 芯片建设项目奠基，6 月与 IBM 签订技术合作协议，双方将在 PCM 芯片生产领域展开合作，目标指向 PCM 系列产品的产业化.

在国内，随着相变存储器器件的研究逐步展开的同时，该领域发明专利申请的数量也保持了逐年增长的态势，根据文献[17]的数据统计，截至 2012 年 11 月 28 日，我国共有 556 件发明专利申请，其中申请人为国内的发明专利有 396 件，专利数目在不断增多，关注点也在逐渐扩大，可见，国内对 PCM 领域的研究已经愈发关注.

综上，在国内对于相变存储器的研究已经开始，中芯国际、全芯科技等国外研究机构与国内研究机构和部分高校一同合作，在 PCM 的介质材料、存储机理和器件原型等方面展开了深入研究，取得了令人瞩目的进展.

2.2　PCM 自身的缺陷及其相关研究

PCM 具有的优势为其在存储领域的应用奠定了基础，但其自身也存在一些缺陷，特别是写性能、写功耗和写寿命问题，成为了它应用于计算机系统的难题，这促使研究人员提出各种方案和途径以弥补 PCM 的自身缺陷.

根据目前已有的研究[4-9]，PCM 的读延迟与 DRAM 在同一数量级，但写延迟约为 DRAM 的 10 倍；同时由于 PCM 写操作需要密集的电流注入，会带来相对 DRAM 较大的写功耗；且由于电流注入时的热应力等问题，PCM 的写入次数有限，其存储单元只能承受上百万次的写操作. 针对上述问题，学术界已展开很多研究工作，主要包括通过改变 PCM 写操作的流程、减少写操作次数和写数据量、磨损均衡等策略来降低写功耗，提升其性能和寿命.

(1) 改变写操作的执行过程

Qureshi 等人[18]针对 PCM 的读写请求延迟不均衡的特点，提出采用写取消及写暂停策略来提升 PCM 的读性能. 当一个读请求到达 PCM 并等待处理时，如果此时正在响应的写请求还没有开始，就取消该写请求，先执行读请求；如果写请求执行到一半，就在某一个写操作的迭代结束后暂停写，先执行读请求. 这种策略利用 PCM 读写请求时间不均衡的特点，提高了系统的读性能. 写暂停的方法虽然某种程度上提高了读性能，但是受限于请求间的数据相关性，会导致一定的读写数据错误，同时会导致写操作的延迟更长.

针对 MLC(Multi-Level Cell，多层单元)型 PCM 的 Program/Verification(简称 P/V 操作)两阶段反复迭代写操作过程，Jiang 等人[19]提出了一种写截断(Write Truncation)策略. 由于将数据写入到

PCM 的不同单元所需要的 P/V 操作的迭代次数是不同的,为了保证所有数据能正确写入,当前的写策略是使用最大 P/V 操作迭代次数来完成写操作.写截断策略则通过在 P/V 操作进行过程中实时检测存储单元阻值,并与目标状态对应的阻值进行比对,动态判断并识别每次写操作需要的迭代次数,在尽量不影响数据正确性的前提下,将写过程的最后几个迭代操作截断,从而减少 PCM 的写迭代次数;同时采用容错能力更强的 ECC(Error Correcting Code)校验来保证数据的可靠性.该策略减少了写延迟,并在一定程度上降低了 PCM 的写功耗,但是不得不面对因动态识别的准确性而带来的数据可靠性的问题,同时 ECC 也具有一定的开销.

Qureshi 等人[20]随后提出一种将 PCM 应用于内存的新策略 PreSET 来提升 PCM 的写性能.该策略利用了 PCM 写"0"(RESET)与写"1"(SET)的不对称特性,即在 PCM 中写"0"的时间很短,而写"1"的时间几乎是写"0"时间的 8 倍.作者提出当 PCM 内存某一个 line 变脏后,则对该 PCM line 进行 SET 操作(即该 line 此时全部置"1"),当后续对该 line 再进行写操作时只需要执行写"0"的部分,减少了延迟,从而提升写操作的性能.PreSET 技术比之前提出的写取消技术延迟要小,效果更好,但是它自身有一个时间开销问题,即在进行 PreSET 操作时,系统必须要分配一个较长的时间窗口来完成 SET 操作.

Yue 等人[21]提出的两阶段写策略同样利用了 PCM 写"0"和写"1"位之间的速度和功耗差异,该策略将写过程分为两个阶段:写"0"阶段和写"1"阶段.前者利用写 0 更快的特性用较短时间完成所有写"0"操作,而后者在不违反功耗限制的前提下并行执行多个单元的写"1"操作,从而提升总体性能.与此同时,作者提出了一种新的编码方案,统计目标写单元中写入 bit 位"1"和"0"的数目,当"1"位比"0"位多的情况下,进行位翻转,即写"0"而非写"1",并用一个标记位记录位反转,以此通过更快的写"0"操作来进一步加快写"1"阶段的速度.该方法虽然对写性能有所提高,但是带来了功耗上的问题和对 PCM 寿命的影响.随后该作者又在文献[22]中提出新的机制即 PASAK 策略,通过开发 bank 内部子阵列级别的并行性,在不违反功耗限制的前提下,当对 PCM 的一个 bank 进行写操作的同时可以并行地进行多个读操作;该机制利用了写"0"和写"1"的电流不对称特性,提出一个电流预分配机制,即根据在一个数据块中写"0"和写"1"的数目获取精确的电流需求,并以此根据电流分配的盈余来服务在该 bank 中没有子阵列冲突下的其他读请求,这可以在没有额外功耗开销的前提下提升 PCM 的并行性.当然该策略并没有考虑如何降低 bank 冲突率和当 bank 冲突发生时如何最大限度地降低性能损耗,且当大量 bank 冲突发生时该策略的适用性也有待考验.

(2)减少写次数或写入数据量

Joo 等人[23]提出了把 PCM 作为 cache 的设计方案,该方案采用"写前读"策略,即在写之前进行读操作,以防止出现冗余的数据更改;在此基础上作者部署了数据位翻转策略,即在写之前计算当前值和写入值的海明距离,并据此确定是否进行数据位翻转以进一步减少位写操作;同时采用磨损均衡策略,利用 Bit-Line 偏移,当某一 line 的写次数到达阈值之后,就更改当前写的偏移值,把写的数据移向其他 line.该方案既发挥了 PCM 高密度、非易失性的优势,又在性能开销不大的情况下,节省了系统功耗,提升了系统的寿命[23].但是在上述机制中,不得不考虑如何降低各个策略本身的开销对性能带来的影响,如 Bit-Line 偏移所需要的标记位、海明距离的计算等额外开销及其移位后对读写性能的影响.

针对 PCM 写操作的去冗余方法,Cho 等人[24]在现有研究的基础上进行了改进,实现了每次写操作实际写入的数据量不超过要写入数据量的 1/2.具体过程为在每次写操作时,比较要写入的数据与要写入单元上的原始数据的差异,如果两者差异数据量小于原始数据的 1/2,直接将差异数据写入,如果大于 1/2,则将要写入的数据全部取反,同时在写操作进行的 cache line 上增加一个 flip 标记位以表示数据位被取反,从而在后续的读操作中能够有所识别.这种方法大大减少了写次数,减少了延迟,降低了功耗,延长了寿命.然而,该策略引入了额外的读数据、比较数据和数据取反的开销,增加了读写硬件的复杂度和读写延迟.

Lee 等人[4]提出了缓冲区重组(Buffer Reorganization)和部分写策略.对于前者,多个缓冲区行利用局部性原理来合并写操作,从而隐藏了 PCM 的延迟,降低了功耗;对于后者,跟踪数据的修改,写入数据时和原始数据进行比较,只将有修改的 cache line 或者 word 写入到 PCM 阵列,由此大大减少写操作次数和写入数据量,提高 PCM 的寿命.上述方案能够隐藏 PCM 的写延迟,降低功耗并能够提升寿命,但方案中合并写、数据比较等机制本身的开销

无法忽略.

(3)通过磨损均衡提升寿命

Zhou 等人[5-6]提出了 3 种提高 PCM 寿命的机制：① 消除冗余位写入. 在 PCM 单元的写操作路径中加上一个 XNOR 逻辑门，通过此判断要写数据的冗余位信息，然后只写入有修改的位，这样减少了写数目，降低了写功耗，延长了 PCM 的使用寿命；② 行移位. 当采用消除冗余位写入后，增加了位更改的局部性，因此通过增加一个偏移器，在一个行中应用一个移位机制，使得每次写入的部分依次偏移定值以避免对某些数据位频繁写入；③ 段交换. 在一个更大的粒度——段上进行移位，在写入过程中，定期互换内存段的高部与低部，以进一步达到磨损均衡，提高 PCM 寿命. 上述 3 种机制综合提升了 PCM 的寿命，但是都带来了各自的开销，特别是行移位和段交换需要在硬件电路上增加移位寄存器和偏移器.

Yun 等人[25]针对 PCM 本身提出一种动态磨损均衡机制，进一步提高 PCM 的寿命. 作者提出基于布隆过滤器(Bloom Filter)的磨损均衡策略，即采用多重布隆过滤器进行冷热数据的识别，方案中维持一个冷数据表和一个热数据表；当一个请求到达时，首先在冷热数据表中寻找，如果命中则进行冷热数据的交换，否则就直接访问请求本身的地址，并更新过滤器计数器值. 同时作者还提出了一个动态更新热数据阈值的思想，实时更新冷热数据的判断依据. 为了避免因冷热数据判定错误而带来额外的开销，该策略为冷热数据识别维持了一个三级队列结构 L3-L1，每一个队列都采用 FIFO(First In First Out)策略，且热数据的“热”性质依次增加，新判定的热数据首先存放在 L3 中，随着其访问频率的递增，会从 L3 升级到 L2，直至 L1. 该磨损均衡策略，提高了 PCM 的寿命. 另外，该机制实际上是借鉴了 NAND Flash 的磨损均衡思想，但策略中的三级队列结构需要占用存储空间，同时冷热数据识别、迁移等机制的进行会影响系统的读写性能，带来一定的延迟.

Ferreira 提出 3 个策略以提升 PCM 存储系统的寿命：① 写回最小化的缓存替换策略. 在选择替换块时优先选择一个没有修改过的块进行替换，以减小因替换操作带来的写回操作发生的几率，同时尽量让频繁修改的条目处在 cache 中以增加可合并写的几率；② 避免不必要的写机制. 采用页面分区机制，即将页面分为若干子页面，只有当子页面有脏位时才将该子页面写回，同时联合 RWR(Read-Write-Read)机制，即在要写回到 PCM 中时读取原始页面的数据并进行比对，不一致时再真正写回，以避免不必要的写发生；③ 磨损均衡算法. 通过映射表和页面写操作计数器，在写回操作时修改映射关系以将写操作频繁的页面转移到不频繁的页面，以此提高 PCM 内存系统的寿命[9]. 上述方案中子页面的粒度选取将直接影响到系统的性能和空间开销的大小，同时选择替换块、分割子页面和统计写操作频率亦会对系统性能造成影响.

(4)降低 PCM 写功耗

Xu 等人在文献[26]中提出通过优化数据处理流程来降低 PCM 的写功耗. 由于 PCM 的读功耗远小于写功耗，且对一个 PCM 存储单元写不同的值所需的功耗具有显著差异，该方案为每一个存储在 PCM 单元中的 word 额外设置一个 XOR(异或逻辑)word，在进行读写操作时可以利用该 XOR word 进行相应的异或操作，一旦欲写入的 word 和与新的 XOR word 异或的结果等于读操作的结果时，则放弃这次写. 以此为基准，当需要进行写操作时先读取写地址内容，并通过 selective-XOR 寻找一个能使得本次写操作数据量最优的新 XOR word，从而获得最低的写能耗. 该方案的缺陷是：① 带来了额外的 XOR 存储开销；② 额外的读操作会有一定的能耗开销；③ 寻找最优的 XOR word 会带来开销.

Hay 等人在文献[27]中提出一种“功耗令牌”(power token)技术来减少 PCM 的写功耗. 该技术根据 PCM 芯片电源引脚所能提供的最大功耗和每个 bit 位写所需的功耗计算出一个 PCM 芯片所能支持的最大写操作数目，从而以该数目控制“功耗令牌”的数量，每次写操作进行时需要先读原始内容，通过比对来决定令牌的发放，当一个 bit 位拥有“功耗令牌”时才能够被写，从而避免很多无用的写操作发生. 该方案实现了在不超过功率预算的前提下降低写操作次数，从而减少功耗. 当然，方案的实现必须得到内存控制器的支持，同时令牌本身带来的开销也不容忽视.

Bock 等人在文献[28]中提出通过避免无用的写回操作来降低 PCM 的功耗. 在该方案中，当一个写回操作的数据不再被系统所使用时，则定义此时向低级别 cache 的写回操作是无效的. 例如，当一个属于无效内存区域的脏块从 cache 中被驱逐时就会产生一个无效写回. 一旦某一个写回被判定为无效写回，就直接放弃该写回操作，从而避免了对 PCM

物理芯片的无效写操作.该策略可以在降低功耗的同时提升 PCM 的寿命,并对性能没有显著的影响,不过具有无效写回操作的判定开销.

除此之外,前述(1)~(3)中很多研究在提升性能或寿命的同时,也潜在地降低了 PCM 的写功耗,如缓冲区重组策略[4]、冗余位消除策略[5-6]、RWR 机制[9]、写截断策略[19]、写前读策略[23]和写操作去冗余策略[24]等.综上而言,优化写操作,减少写操作次数、写入数据量或写操作本身所需的迭代数目,为同时优化 PCM 的功耗和寿命提供了一条有价值的研究路线.

2.3 PCM 在系统中的应用层次及引发的变革

当前的计算机系统架构,采用传统的 DRAM 作为内存,Disk 作为外存.随着 NVM 的出现,传统架构正发生着变化,它已经开始部分或者全部替代内存或外存:在作外存时,NVM 可以完全替代 Disk 或与之构成混合外存;在作内存时,NVM 可以和 DRAM 联合使用构成混合内存,同时也可以考虑完全替代 DRAM 来作内存使用.图 2 展示了 NVM 的出现对计算机系统结构产生的影响.

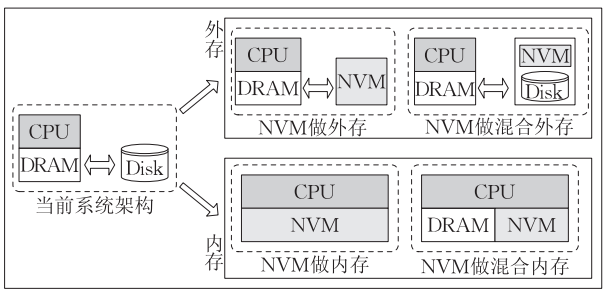

图 2 NVM 对计算机系统结构的影响

随着器件研究的推进,PCM 在系统中的应用逐渐起步,已经给现有的存储系统带来了新的机遇和变革.PCM 在计算机存储系统领域的研究,主要方向有两个:一是作为外存设备使用,二是作为内存介质使用.在内存领域的应用,又集中在两个方面,一个是直接采用 PCM 取代 DRAM 作为内存(Main Memory),另一个是采用 PCM 和 DRAM 作为混合内存(Hybrid Memory).由于目前 PCM 仍存在一些缺陷,用作外存的可行性相对较高.

Ranganathan 和 Chang 在文献[29]中提出,当前计算机存储系统架构正在经历从使用机械硬盘到使用固态硬盘作为非易失存储设备的演变,未来将由存储级内存(Storage Class Memory,SCM)实现统一内外存系统.作者分析了当前计算机应用环境的变化:一方面,收集和处理大规模数据的需求增加,实时商务分析等诸多基于数据中心的应用不断地出现在生活中;另一方面,数据的产生速度已经超过了存储介质的密度、价格等方面的发展,处理数据的能力明显滞后于采集和存储数据的能力.同时论述了在面对上述需求和矛盾时,当前系统架构和软硬件方面的设计方案和设计思想,包括硬件架构、软件系统、上层算法、以数据为中心的数据计算等,认为在未来传统的内存和存储体系结构之间的区别将会变得含糊,数据计算将会在以数据为中心的数据存储和通信中变得普遍,且传统的服务类进程将拥有更多专业的计算模式;在将来会有新的软件栈诞生,它会避免现有解决方案的缺陷,拥有可字节寻址的持久存储特点,也会有新的算法来匹配这种新的软件硬件的架构设计等.因此该领域不得不重新考虑利用技术革新如采用非易失存储器、光通信、多核、异构计算等技术为数据中心系统提供新的解决方案,以最大限度地适应技术的发展,适配于新的应用场景.

在传统的块 I/O 路径上,操作系统几乎通过中断完成所有的异步 I/O.Yang 等人[30]的研究针对 PCM 读取速度快的优点,提出将目前 I/O 路径中的异步 I/O 访问模式改变为同步 I/O 访问模式,即采用轮询方式取代中断,与 PCM 设备通信.作者通过实验证明同步完成模式是安全可行的,并认为未来将新一代 NVM 介质引入同步完成模型,将具有显著的性能优势,这为今后计算机系统的设计提供了一种新的思路.但是该思想没有考虑 PCM 的写性能问题,也不得不面对同步模式下不可忽略的 I/O 开销以及在内核中因应用程序的复杂性而来的技术瓶颈.

Coburn 等人[31]提出了一个轻量级的面向持久对象的系统——NV-Heaps,该系统基于 PCM 实现,允许程序开发者将系统重要的数据结构信息存储在非易失存储器中,并将其物理地址直接映射到应用程序的地址空间,利用 PCM 可以本地覆盖写的特性实现就地更新.NV-Heaps 通过三种机制实现了当系统发生掉电等意外时能够恢复到与之前一致的状态.首先采用有选择的进行系统恢复的思想,即在持久堆(persistent heap)失败时恢复,而在易失的堆对象、栈对象以及线程上下文发生意外时不进行恢复;其次采用 flush-on-commit 更新策略和事务日志策略以确保掉电后系统的更新不会丢失;最后使用软件事务性内存(Software Transactional Memory,STM)来隔离线程和持久堆之间的一致

性. NV-Heaps 提供了一个轻量级、高性能、面向持久对象的系统，但是其无法确保所有的应用程序能够得到恢复，同时 flush-on-commit 会带来很高的运行时开销，而 STM 应用价值也具有局限性，并不是所有的应用程序都会使用到.

Moneta-D 是一个新颖的存储体系结构[32]，其采用实际的 PCM 作为存储设备. 该体系结构为每个访问提供一个私有的、虚拟化的接口，并由硬件完成文件系统保护检查，使得应用程序可以在没有操作系统的干预下访问文件数据，从而消除操作系统和文件系统的开销；同时该结构仅仅需要文件系统做很小的改变就能够支持新的访问模型，而不需要修改现有的应用程序接口. 虽然 Moneta-D 降低了执行文件系统权限检查的开销，但还是带来了因权限管理、访问等其他操作所导致的系统额外延迟.

针对目前数据库数据和键值(key-value)存放在内存中需考虑的因意外掉电或操作失败所面临的数据恢复问题，Narayanan 等人[13]提出一种 WSP (Whole-System Persistence)方法，利用 PCM 实现非易失内存，这样因断电或者操作失败而导致的数据恢复就能在本地快速进行，该方案不同于 NV-Heaps 的 flush-on-commit，而是采用了 flush-on-fail 的思想，其具体实现包括一个保护和恢复程序，设备重启程序，并使用可编程单片机部署了一个功耗监控器，在硬件实现方案中添加了一个 Power-GEM 模块，即一个采用超大电容的供电子系统，当系统掉电时，该模块将服务于保护和恢复程序保存和恢复应用程序和操作系统的状态. 该方案利用 PCM 提供了一个具有非易失性的内存原型，但其本质上只是在系统掉电或操作失败时利用 PCM 的非易失性，并没有将 PCM 的其他优势如功耗、密度、性能等充分开发并应用于内存系统.

3　PCM 内存技术研究综述

采用 PCM 部分或者完全取代现有的 DRAM 内存需探索新的内存管理方法，研究新的内存控制器硬件和新的内存管理软件. 这些研究假定 PCM 和 DRAM 一样，提供 DDR 的接口，CPU 直接通过 DDR 内存控制器访问 PCM.

3.1　基于 PCM 的直接内存技术研究

Qureshi 等人[7]针对 MLC 结构的 PCM 在容量倍增时所面对的性能降低问题，提出一个 MMS (Morphable Memory System)思想，MMS 能够在内存应用中有效地整合 MLC 型的 PCM 设备，其通过检测负载对存储资源需求的变化来分配内存容量：当某一阶段负载对内存容量需求不高时，一些 MLC 单元将按照 SLC(Single-Level Cell，单层单元)单元使用，从而降低延迟，提高读写速度；当工作负载对容量要求很高时，MMS 将上述 SLC 存储单元恢复到高密度的 MLC 模式使用，从而满足负载对容量的需求. 该方案能够在 PCM 容量与性能之间寻求一种平衡，但其需要比较复杂的内存硬件控制器和管理软件支撑，特别是当负载需求变化时，MLC 和 SLC 之间必须进行数据的迁移和映射信息等方面的改变.

在内存应用环境中，当 PCM 页面发生错误后，传统策略是直接丢弃该页面. 针对此问题 Chen 等人[8]提出一个 PCM 内存系统——rPRAM(redundancy PRAM)，以通过重新利用含有错误比特的 PCM 页面来增加 PCM 的页面利用率，提升 PCM 的寿命. 在该方案中，当多个页面发生错误的 bit 位位置不同时，可以进行错误兼容，即这几个错误页面可以与一个 DRAM 共同构成一个 RAID4，从而继续存储数据. 方案中设定当一个页面发生了大于 160 bit 位的错误时，就放弃该页面. 该方案最大的挑战和局限性在于随着错误的逐渐增加，寻找能够进行错误兼容的页面会越来越困难，从而导致页面的重新利用率越来越差，且开销也会逐渐增大.

PCM 是非易失存储器件，当其用作内存时，在掉电后仍将保留数据，这种非易失性会使得系统具有快速启动和休眠恢复的能力，但也带来了数据安全性的问题. 对此 Kong 等人[33]提出了基于 PCM 内存系统的硬件加密方案——计数器加密模式，该方案的思想是为每个 cache line 设计一个计数器，同时还设计多个 block 级别的计数器，对每一个数据块的加密操作是联合 cache line 级别和 block 级别的计数器共同完成的，当某 cache line 写回或者其他操作发生时，必须通过该 cache line 的计数器和拥有该 cache line 的所有 block 计数器共同控制才能完成，从而确保数据操作的安全性. 在该加密方案的基础之上，利用上述计数器统计发生在 PCM 存储单元的写操作次数，从而据此进行磨损均衡策略的实施，以尽可能地降低加密机制对 PCM 寿命的影响. 当然这种加密算法在提高安全的同时，带来了较大的开销，且加密算法的复杂度也不容忽视.

Chhabra 等人[34]提出了一种针对非易失内存

的增量式数据加密方法——i-NVMM. 在内存环境中,一个应用程序的工作数据集合往往小于常驻内存的数据集合,因此 i-NVMM 的思想即为通过一个预测机制,利用局部性原理来判断数据页对处理器而言是否还会有用,从而确定该数据被加密的时刻,即方案采用逐步加密策略,先仅对内存中暂时不会使用的数据进行加密,对于正在使用的数据则暂时不进行加密,并对其进行监视,直到符合预测条件时再快速完成加密,以此来降低加密过程对运行时性能的影响. 该方案需要精确地预测处理器将会使用的数据集,因此预测的精确度和预测带来的开销对该方案的效果有很大影响,同时加密会带来较大的延迟.

3.2 基于 PCM 的混合内存技术研究

混合内存系统设计思想的出现,为 PCM 在内存领域提供了一个新的应用途径.

当前绝大多数基于 PCM 的内存研究都是采用软件模拟器的方式进行,对此 Kwon 等人[35]研究了把 PCM 作为真正的内存的可行性以及所面临的问题. 作者首先采用 Samsung 公司 1Gb 的 PCM 芯片进行了一系列性能测试,以此为依据探索了在真正的原型系统中 PCM 作为内存的关键技术以及需要体现的优势,主要研究了 PCM 写延迟(存储单元写延迟+R 漂移延迟)与写数据大小之间的关系,量化了在基于 PCM 的内存子系统中 PCM 延迟对性能的影响,作者通过实验结果得出结论——为了服务一个读请求而暂停一个正在进行的写操作,至少需要 25 μs 的时间开销,并证明 PCM 的写功耗限制了内部写操作的数据宽度,如果想增大数据带宽,就一定会增加写延迟. 最后作者通过混合内存模拟器的方案展示了一个 PCM 和 DRAM 的混合内存系统性能要优于一个纯粹的 PCM 内存系统;同时混合内存系统中 DRAM 的容量大小会对系统性能产生很大影响.

Qureshi 等人[36]探讨了一个由 PCM 存储器和小容量的 DRAM 缓冲区构成的混合内存系统架构及该架构中两种介质的容量配比问题. 在该架构中操作系统使用一个页表管理 PCM 介质,同时采用 Lazy-Write Organization、Line-Level Writes、Fine-Grained Wear-Leveling 等基于该页表的策略来减少 PCM 的写操作,提高 PCM 使用寿命. 其中 Lazy-Write Organization 的思想为当发生缺页时,从磁盘中读取数据并直接写到 DRAM 中,而当页面被逐出 DRAM 且该页被修改过时才写入 PCM 中;Line-Level Writes 则指在 DRAM tag 中为每个 cache line 设置一个 dirty 标记位,以确保只向 PCM 写回那些修改过的行;Fine-Grained Wear-Leveling 则利用逻辑扇区到物理扇区之间的细粒度映射进行磨损均衡. 实验证明在该架构中当 DRAM 缓冲区容量占 PCM 存储设备容量的 3%左右时,几乎可以隐藏 DRAM 和 PCM 之间延迟的差距. 当然,上述机制的缺陷也很明显,特别是细粒度映射表的巨大空间开销以及脏位的判断、比较操作带来的性能延迟等.

Dhiman 等人[37]提出一个与上述实施方案不同的混合内存系统——PDRAM,在 PDRAM 系统的开发过程中,为了保障该系统的可靠性,作者提出一个具有成本效益的记账(book keeping)硬件技术,用于存储 PCM 在页级别的写操作频率;同时在软件方面,系统内存控制器为 PCM 维护了一个写访问次数的 map 表(以 page 为粒度),通过利用由硬件提供的写频率信息,实时监控页面的写操作次数,当其到达某一既定阈值时,则发起中断来执行磨损均衡策略. 该方案中硬件可以在很小的开销下维持和追踪页层的访问,而软件的管理策略可以提高 PCM 寿命. 该方案中利用页面写操作次数来发起中断进行磨损均衡的方式,是以牺牲性能为代价来提高 PCM 寿命的.

Zhang 等人[38]给出了另一种基于 PCM 的混合系统思想,将 DRAM 和 PCM 放在同一个平面地址空间内,其基本思想是将很少修改的页放入 PCM 而把修改频繁的页面放在 DRAM. 在最初所有页面存储在 PCM 中,DRAM 则作为操作系统的一个写入分区,在运行过程中,通过 MQ(Multi-Queue)算法把修改频繁的页迁移到 DRAM 中. 方案中 MQ 算法的实现依靠 16 个 LRU 队列,每一个页为一个队列元素,对页的写访问进行计数,当计数达到一定阈值时,则认为该页属于频繁修改的页,从而进行迁移操作;同时把那些之前在 DRAM 中但具有较少写入的页面迁移回到 PCM 中. 该方案通过页面布局来充分发挥 PCM 和 DRAM 的读写优势,但 MQ 的队列开销很大,页面的迁移操作也将影响到系统的性能.

Ramos 等人[39]描述了一个基于 PCM 的混合存储系统的页面管理策略,设计方案包含了一个复杂的内存控制器来部署一个称之为"基于等级的页面放置 RaPP(Rank-based Page Placement)"机制,其主要思想为:将频繁写入的页放入到 DRAM 中,

将含有非关键数据的页和写频率低的页放在 PCM 中；同时根据页面的写操作频率以及写密集度对页面进行动态排序，并将排在队列最前的页面（Top-Ranked Pages）从 PCM 中迁移到 DRAM 中. 方案采用一个改进的 MQ 算法进行元素的升降级：当队列中的引用计数器达到迁移阈值时，RaPP 就将存储在 PCM 队列中的页面迁移到 DRAM 中. 虽然通过上述策略能够发挥混合存储介质的优势，但仍要面临与文献[38]同样的问题.

Mladenov[40] 提出了一种基于大容量 PCM 阵列和小容量 DRAM 缓冲区的混合内存系统，该系统由内存管理器和内存控制器组成，内存管理器通过 Internal Memory bus 与内存控制器互连. CPU 请求通过 Memory bus 到达内存管理器中的请求控制器，然后进入先进先出的读写队列，当请求开始执行时首先查找 DRAM buffer，对于读请求如果命中则直接从 DRAM buffer 中取出数据，否则将读请求发送至 PCM Request Processing 模块，并从 PCM 阵列中读取出数据返回至 DRAM buffer 中；对于写请求，若 DRAM buffer 有空间则直接写至 DRAM 中，若被填满，则先将 DRAM 中最早处理过的请求的数据写回 PCM 中，再将数据写入 DRAM. 系统会定期地检查 DRAM，当内存空闲时则使用基于 LRU 的替换算法，将 DRAM 中的一部分数据写回 PCM 中，为今后的写请求预留空间. 该混合存储系统中 DRAM 缓冲区利用程序的空间局部性响应 CPU 的部分请求，可以有效弥补 PCM 速度的不足. 除此以外，该混合内存系统还在 PCM 阵列中部署了“start-gap”磨损均衡策略，以此提升 PCM 的寿命.

Baek 等人[41] 则将双阶段压缩策略引入 PCM 和 DRAM 的混合系统. 该策略分为两个阶段：第 1 阶段通过一个简单的、低延迟的、字（word）粒度的压缩算法减少 PCM 的访问次数，从而增加了有效存储容量；第 2 阶段采用 bit 粒度的压缩算法，进一步降低 PCM 访问次数，增加 PCM 的寿命. 同时该策略利用压缩后剩余的内存空间部署一个低开销磨损均衡技术，提升了 PCM 寿命. 双阶段压缩算法本身的实现复杂性不容忽视，且由此带来的延迟等一系列问题会对性能有所影响.

Park 等人[42] 提出 3 个策略来降低混合内存系统的功耗：(1) 运行时自适应的 DRAM 功耗递减策略. 在 DRAM 的每一个 row 内维持一个时间变量，当该 row 写入数据后该时间变量值会定期衰减，当减少到 0 后就将该 row 中的数据驱逐，且该 row 不再刷新，以此减少 DRAM 的刷新功耗；(2) 绕开 DRAM 策略. 在该混合内存系统中 DRAM 充当 PCM 的 cache，当第一次对混合存储系统进行读时则绕开 DRAM，从 PCM 中读出数据，当再次访问该数据时，则标记该数据为热数据，此时才将该数据从 PCM 拷贝到 DRAM 中，通过减少因热数据而引起的 DRAM 刷新操作来降低功耗；(3) 保持脏数据策略. 作者认为脏数据往往在未来有更多的被写机会，因此提出让脏数据在 DRAM 中停留更长的时间，降低 DRAM 写回的开销，同时采用写合并策略，减少 PCM 的写次数，进一步降低系统功耗. 此方案从降低混合系统功耗出发，但忽略了策略实现的复杂度.

Hu 等人[43] 提出了一种在嵌入式系统中的混合内存系统，提出软件磨损均衡策略 SWL（Software Wear-Leveling），该策略首先使用数据最优分配算法 ODA（Optimal Data Allocation）为每一个存储区域进行数据分配，并给 PCM 的每个存储区域地址分配计数器变量，以数组方式来记录每个存储区域地址的写操作次数；然后利用该数组，将写请求分配到写操作次数最小的那个区域地址，最后该区域计数器加上此次发生的写操作频数以更新该变量值. 该策略实现了 PCM 中写操作的均匀分布，同时方案中增加了硬件级优化，提升了嵌入式系统中 PCM 的寿命. 然而，该策略中使用 ODA 算法作为预处理会给系统带来延迟，同时大量计数器的查找、更新等操作和方案中的硬件优化都具有一定的开销.

4　PCM 外存技术研究综述

除内存以外，PCM 还应用于外存系统中，将 PCM 与现有的外存系统融合，成为了它在外存中的一大应用热点.

4.1　基于 PCM 的外存技术研究

在 PCM 应用于外存的技术中，大多数研究是针对其与 NAND Flash 的混合外存策略.

Kim 等人[14] 提出采用 PCM 存放页内日志的 IPL-P（In-Page Logging with PCM）方法，它利用 PCM 在读写速度上的优势和字节可寻址的能力，将数据库应用环境中频繁读写的日志文件存放在 PCM 中，以提高基于闪存的数据库系统的性能. 该方法虽然利用了 PCM 相对于闪存的性能优势，却忽略了 PCM 的寿命问题，其本身对 PCM 的寿命是

一种极大的考验.

Sun 等人[44]受到 IPL-P 策略的启发,在 PCM 和 NAND Flash 混合外存架构中,提出用 NAND Flash 来存储数据页而 PCM 存储日志页,利用 PCM 支持本地更新的特性,避免了因日志页频繁更新而造成存储空间的浪费,同时因为 PCM 的性能优于 Flash,从而缓解了日志页的读拥塞问题,且对 NAND Flash 的寿命也有积极意义.但方案完全忽视了 PCM 的寿命问题,并没有采用相应的策略来缓解因方案本身对 PCM 寿命造成的威胁.

Liu 等人[45]提出了用于混合外存系统的块映射算法——WAB_FTL,该方案利用 PCM 存储映射表,通过两种机制尽力对映射表中的位进行保护而不被频繁地修改,从而提升 PCM 的寿命,此两种机制分别为(1) Lazy-Merge. 可延迟映射表条目的更新操作;(2) Cooling-Pool. 可减少映射表条目的更新操作.对于前者,当替换块被擦除后,将旧替换块的所有有效页复制到一个新块中,再把旧替换块中相应一致的映射表记录更新到新替换块中;对于后者,采用缓冲池缓存频繁更新的映射表记录,以减少对 PCM 的擦写操作.方案中 Lazy-Merge 策略具有一定的实现代价,且 Cooling-Pool 需要额外的空间开销.

除了与 NAND Flash 联合做混合外存以外,PCM 替代 NAND Flash 的研究也已展开.

Akel 等人在文献[46]中提出了一个基于 PCM 的真实的高性能固态存储系统 Onyx,该系统部署在 BEE3 FPGA 原型系统之上,采用美光公司的 P8P 系列芯片,存储容量为 10 GB,PCM 阵列通过 PCIe 总线和主系统连接,具体的读写操作控制通过相变存储器 DIMM 控制器完成,同时在该原型系统中真正实现了 start-gap 磨损均衡机制.测试结果表明对于小的写请求(<2 KB)和任意大小的读请求,该系统的性能和基于 Flash 的 SSD 相当甚至更优.但对于更大的写请求,该系统的性能较差.

Moneta 是一个基于模拟 PCM 设备的 PCIe 存储阵列原型[47],它提供了一个精心设计的软件、硬件接口,使得任务分发以及完成操作原子化.该接口由硬件调度优化表及一个存储堆栈组成,硬件调度优化表接收访问请求,根据请求的大小将其分割成子请求,然后通过 DMA(Direct Memory Access)快速向主机传输,当调度程序完成所有传输请求时,发出一个中断并且设置一个标记状态,操作系统接受中断,完成请求.同时,由于主机端发生中断时进程唤醒需要较大的上下文切换时间,因此该方案采用进程不断循环等待的策略而不是简单的挂起,以此来减少开销.通过上述一系列机制,Moneta 极大地提升了小的随机访问请求的性能.但 Moneta 不可避免的问题是在较小请求的情况下,如 4 KB 请求大小,软件开销会极大的限制系统的整体性能,核态与用户态的切换、文件系统的检查等都会造成很大的延迟.

Shao 等人借鉴了 NAND Flash 中闪存转换层(Flash Translation Layer, FTL)的概念,在文献[48]中提出了一个对等的相变存储转换层 PTL 策略.PTL 立足于 PCM 自身的特性诸如支持本地更新、支持以字节为单位的读写操作、写操作之前不需要擦除操作等,包含读写请求管理模块、磨损平衡模块、可靠性模块、地址映射模块,能有效地隐藏 PCM 的物理特性,同时作者提出了一种针对嵌入式系统中 PCM 的简单有效的磨损平衡方法 AWL(Application-specific Wear Leveling),该算法通过识别 PCM 中的热数据区域和冷数据区域,把热数据上的写操作均匀地分配到其他地方,以此达到磨损均衡,提高 PCM 的使用寿命.除此之外,作者还采用了地址映射算法,解决了由于磨损均衡导致的逻辑地址和物理地址不对应的问题.PTL 是一种适用于嵌入式环境中透明的、有效管理 PCM 的系统级框架.PTL 没有很好地管理 PCM 的写操作,也没有采用任何方法来减少 PCM 的功耗,这是它后期要解决的问题.

Lee 等人[12]则立足于 NVM 的优势,提出一个基于 PCM 的文件系统 Shortcut-JFS.由于文件系统运行时需要记录大量的日志信息,并需要快速访问日志信息,作者提出了两个策略来实现系统写操作的优化和检查点(checkpoint)信息的改善:(1)差异更新策略.利用 PCM 按位更改的特性,在文件系统记录日志时,Shortcut-JFS 仅仅更改日志中变更的字节,但当需要更新的数据大小大于块大小的一半时,则更改整块日志数据而不使用差异更新策略;(2)本地更新策略.在文件系统中当要通过检查点对文件系统数据进行还原时,采用本地更新策略,即如果要更新的数据已经凑满一个完整的数据块,则利用一个指针指向该日志数据块而不再写回文件系统,如果不满一个块,则利用 PCM 优于 Flash 的本地更新特性直接写入要更新的部分,而不再需要另开辟空间写整个数据块,这大大减少了写操作次数,同时节省了存储空间开销.该方案中的差异更新思想已不鲜见,且 Shortcut-JFS 忽略了针对 PCM 介质的磨损均衡问题和写性能问题.

4.2　基于 PCM 的 SCM 技术研究

存储级内存 SCM 是一种采用非易失存储介质实现的新型存储技术，支持位访问，性能可以和 SRAM、DRAM 相比，存储密度高，容量大，功耗低，具有非易失性. 与传统的磁盘不同，SCM 可以被内存中的硬件机制所保护，也不需要复杂的调度机制，且速度与 DRAM 相当，不需要 cache，作为内存使用时还可以起到断电保护的作用，成为存储系统发展的一个新路径，采用 SCM 作为统一内外存使用还可能会带来整个计算机结构的变化. 因此有不少学者认为采用 NVM 实现统一内外存将是未来的发展趋势之一.

Lam 在文献[49]中论述了适用于 SCM 技术的各种存储介质，包括 PCM，STT RAM，MRAM，FeRAM 等. 在这些新型存储器中，作者认为 PCM 具有明显的优势，并论述了 SCM 在计算机存储层次结构中的地位：既可以作为内存的下一级外存的 Cache 使用，也可以直接作为外存使用. 当 SCM 作为内存使用时，它可以直接与 CPU 连接，通过内存控制器硬件管理，CPU 可直接通过 Load/Store 命令操作 SCM 中的存储单元，降低访问延迟，扩展内存的容量；当作为外存 I/O 设备使用时，系统可以通过 I/O 控制器，采用文件系统等软件方式来管理 SCM，即使这种方式访问延迟高，但与传统的磁盘相比仍然具有巨大的优势. 除此之外，作者在文中从硬件到软件、从上层到下层依次论述了 SCM 在系统中发挥自身优势所面临的技术挑战，如采用 SCM 作为 I/O 设备时，应该考虑 SCM 提供什么样的上层接口等，启发研究人员思考 SCM 应该作为什么样的设备使用才能最大地发挥它的性能.

Condit 等人在文献[50]中提出了基于可字节寻址的持久 RAM（Byte-addressable, Persistent RAM，简称 BPRAM）的文件系统：BPFS（a File System for BPRAM）. 在整体方案中使用短电路影子分页（Short-Circuit Shadow Paging）技术为 BPFS 的可靠性提供保证，并以此提供比传统文件系统更好的性能. 短电路影子分页技术通过原地更新（In-place update）、原地附加（in-place append）、部分写时复制（partial copy-on-write）3 个机制来支持基于字节寻址的操作，保证了 BPRAM 本地更新的特性，可以使系统在文件系统树的任何位置更新，避免对 root 文件系统的传播拷贝开销，实现小单元的原地更新操作，不用考虑额外的拷贝操作，甚至当进行大的写操作时，更新操作也会被限制在一个小的子树文件系统上，仅对那些没有变化的数据进行拷贝；同时作者提出一个基于 BPRAM 的硬件架构，让 BPRAM 直接与内存总线连接，并与 DRAM 平行，这样将易失和非易失内存的地址空间分开，CPU 可以直接通过 load 和 store 操作来访问 BPRAM，满足 BPFS 要求的原子性及执行的顺序性，为 BPFS 提供可靠性保护.

Wu 等人[51]随后提出了 SCMFS，一个专门针对 SCM 的文件系统. SCMFS 文件系统沿用了常规文件系统的接口，可以支持现有的应用. 它建立在虚拟地址空间上，通过内存管理单元 MMU（Memory Management Unit）完成虚拟地址到物理地址的转换. SCMFS 布局相对于现有的文件系统来说非常简单，主要包含 3 个部分：第 1 部分是超级块（super block），包含了整个文件系统的所有信息，如文件系统的块大小、索引节点 inode 等信息；第 2 部分是 inode 表，包含了每个文件或者目录的基本信息，如文件模式、文件名、文件大小、文件数据的起始地址等信息；第 3 部分则为具体的文件内容. 需要指出的是，SCMFS 的垃圾回收策略在后台进程运行，在垃圾回收时，首先检查空文件（null file）的个数，如果超过阈值就释放部分空文件，如果用户需要更多的空闲空间，就会首先考虑释放“冷”文件，如果在系统中的空文件太少，该机制就会直接创建空文件.

5　PCM 未来系统级研究方向探讨

随着大数据时代的到来以及多核处理器、虚拟化、内存计算等技术的不断发展，应用场景对存储系统的要求越来越高，相变存储器的出现为存储系统的发展提供了新的路径. 在未来，以 PCM 为代表的 NVM，将进一步在计算机存储系统中崭露头角，而针对它们在系统级的应用仍有很多挑战性的问题需要深入研究，这些研究将可能从以下几个方面展开：

（1）基于 PCM 的存储系统的组织结构方法研究：

当前存储系统的组织结构是专为易失、读写差异小、几乎无寿命问题的 DRAM 以及传统的硬盘、固态盘等存储介质而设计的，这种系统组织结构对于新型非易失存储器而言是不适用的，无论是当前的内存管理方法、访问接口设计，还是 I/O 请求调度等都没有充分考虑 PCM 的寿命、性能、读写不均衡等问题，导致了非易失存储器的特性不能够得到充分发挥，同时还可能会将 PCM 的弱点放大，不利于构建面向未来大数据和内存计算环境的高性能、

低功耗、大容量的存储系统.

如上所述,从 PCM、DRAM、NAND Flash 等多种介质的优缺点出发,研究以 PCM 为代表的 NVM 在异构混合存储系统中的组织方法,合理组合多种存储介质,构建多介质的异构混合存储环境,建立可以充分发挥各存储介质特性的体系结构,解决多介质异构混合存储时的系统优化设计问题,实现新型非易失存储器与现有存储技术和系统的完美融合.

(2)基于 PCM 的存储系统的访问方法研究:

传统存储系统中的访问方法是立足于 DRAM、Flash 等设计的,它将不再适用于具有可字节编址和位修改等特性的 PCM. 和 DRAM 不同,PCM 的读写不对称使其难以按流水线方式执行读写混合 I/O 请求,当前 PCM 与现有内存系统在访问特性上有显著差异,与此同时 PCM 支持本地修改等异于 Flash 的特性也使得当前外存领域的访问方式需要优化和改进.

研究基于 PCM 存储系统的多接口适配的访问方法,以匹配新型非易失存储器的特性,从而隐藏多介质在访问粒度、延迟、带宽及寿命等方面的差距,提升存储系统的性能. 未来研究将可能包括:① 研究 PCM 在内存环境中字节粒度寻址的读写访问方法,充分挖掘 PCM 通道间、芯之间以及芯片内部的多层次访问并行性;② 研究在外存环境中块粒度寻址的高效读写访问方法,并遵循业界针对非易失存储器的接口标准(如 NVMe 协议);③ 优化访问路径,减少系统 I/O 调用给性能带来的影响;④ 利用 PCM 的读写特点来优化读写操作和流程,以此减少访问延迟;⑤ 立足 PCM 特性优化系统中的数据结构,减少对 PCM 无用的写操作和写入数据量,以提升系统性能和寿命.

(3)基于 PCM 的存储系统数据可靠性研究:

随着工艺制程的降低,非易失存储器的存储单元不断变小,当 PCM 采用更小制程、提供更高存储密度和更大容量时,其存储单元的错误率随之升高. 同时,PCM 存储单元的可擦写次数有限(10^8 ~ 10^{12}),频繁的擦写会导致芯片单元很快到达寿命极限. 这些将使存储系统面临数据发生错误、损坏以及丢失的风险,对数据可靠性造成了极大的威胁.

未来的研究将立足于 PCM 的特性,通过多种途径来保障数据的可靠性,研究将可能在以下几个方面展开:① 研究降低当前已有的纠错机制(软硬件)所需的开销;② 研究可配置、适应数据集属性的组合校验算法,即区别不同属性的数据集,根据其所需的可靠性需求采用不同纠错能力和开销的校验算法,以平衡其纠错强度和校验开销;③ 研究新的通过减少写操作次数、写入数据量来提升 PCM 的寿命的策略;④ 研究新颖、可用范围广的磨损均衡策略,在现有磨损均衡基础上进行创新、优化,设计出可应用于不同需求环境下的磨损均衡策略,提升 PCM 寿命;⑤ 研究基于 PCM 的坏块复用和数据容错机制,进一步增加 PCM 的使用寿命,提高数据可靠性;⑥ 研究数据一致性的保障和维护,根据存储系统数据一致性需求、访问接口粒度等因素,设计低开销、多路径的数据更新策略和数据一致性维护方法.

(4)基于 PCM 的存储系统数据安全性保障研究:

由于 PCM 具有非易失性,即当系统断电时,PCM 存储的数据并不会消失,从而通过恶意修改数据所导致的执行状态可能是持久的,即使设备断电,系统也会存在冷启动攻击的风险. 因此非易失特性会使系统被入侵和数据被盗窃的风险增大. 所以当采用 PCM 构建内存子系统时,需要考虑数据的安全性保障机制.

对此,未来该领域还需要研究针对操作系统的加密机制,通过加密模块对写入 PCM 的数据进行加密,防止存储数据被窃取或泄密的情况发生;研究利用访问权限控制等策略来保证数据的访问安全性;特别针对 PCM 中的系统关键数据,需采用强度更高的加密、上锁等算法,防止恶意的入侵修改所引起的系统安全问题,保障基于 PCM 的存储系统的数据安全性.

(5)基于 PCM 的存储系统软件优化研究:

由于 PCM 异于传统存储介质的特性,使得 PCM 存储技术不能良好地兼容当前存储系统的内存管理、文件系统等软件架构. 基于 PCM 的存储系统,在软件层仍然需要改进,以进一步优化和提升存储系统的性能.

未来基于 PCM 的存储系统软件优化研究将可能包括:① 结合各存储介质的特性,基于 PCM 存储管理架构,研究冷热数据识别算法和数据热度分级管理等软件策略,降低存储系统中的读写操作开销,实现负载均衡;② 立足于 PCM 在存储系统中的应用场景(如统一内外存),针对 PCM 支持本地修改、位修改和可字节编址等特性,研究适应于 PCM 的文件系统,从而提升文件系统乃至存储系统的性能;③ 研究基于 PCM 的内存分配机制及其优化策略,从操作系统层入手面向文件系统、虚拟内存等进行优化,降低页面分配等多种内存管理开销,充分地利

用 PCM 的非易失性提高系统性能；④研究设计新的软件调度算法，通过调度策略的设计和优化，达到系统性能的提升.

(6)基于 PCM 的真实存储原型系统的研究：

现有的研究还面临着几乎没有可用的基于 PCM 的真实硬件原型平台的尴尬局面，绝大多数研究均是在软件模拟器上进行的，表 2 展示了上述部分研究采用的模拟器平台，当前比较成熟的模拟器有 PCRAMsim[52]、Simics[53]、M5[54] 和 DRAMsim[55] 以及近些年备受学者青睐的全系统模拟器 GEM5①. 由于 PCM 技术研究还处于起步阶段，其应用场景和价值尚未完全开发实现；而且目前市面上的主流存储器仍然不是 PCM，适合于当前存储环境的大容量、高性能的 PCM 物理芯片稀贵，这些都导致当前系统级的研究几乎全都是基于软件模拟器进行的，从而无法获取最真实的实验数据以进行更加专业、深入的研究.

表 2　基于 PCM 的混合存储研究-模拟器一览

文献	模拟器平台	文献	模拟器平台
[6]	CACTI-D	[4]	SESC Simulator
[9]	Simics	[19]	Simics
[23]	PCMsim	[35]	McSim
[21]	1. M5 Simulator 2. DRAMsim	[22]	1. M5 Simulator 2. DRAMsim
[36]	In-house system simulator	[37]	1. M5 Simulator 2. CACTI 4.1
[39]	1. M5 Simulator 2. DRAMsim	[41]	1. Simics 2. GEMS
[43]	1. Simplescalar 2. CACTI /NVsim	[45]	The author developed a simulator themselves

利用相变存储器物理芯片，实现真实的存储硬件原型系统，包括基于 PCM 的内存原型系统和外存原型系统，解决目前相关研究没有原型平台的尴尬局面，通过在平台上获得最真实的数据，展开更有说服力、有数据依据的相关研究，将对当前内/外存储系统架构的研究工作起到积极作用.

综上所述，面对新应用环境的需求，思考如何在存储系统领域的设计中充分利用 PCM 等 NVM 的优势，并通过新的机制和策略来克服它们的劣势，与现有存储系统融合并获得更加优秀的性能，从而最大限度地匹配以数据为中心的发展趋势，将是未来的主要工作之一.

6　结束语

本文总结了基于相变存储器的存储技术研究，分析了当前 PCM 应用于存储系统的相关技术的优势和缺陷，并探讨了该领域未来的研究和发展方向.

基于 PCM 的内存研究，为内存技术的扩展提供了新的思路，基于 PCM 的混合内存研究，使其可以和传统的 DRAM 系统竞争，同时其优于 Flash 和传统硬盘的特性，将进一步加快它在外存领域的应用和发展. 在国际半导体工业协会对新型存储技术的规划中，已将相变存储器列入优先实现产业化的名录. 作为已进入产业化前期的新型存储技术，相变存储器是近几年发展最为迅速、距离产业化最近、商业化前景最为广泛的新型存储介质之一. 面对以 PCM 为代表的 NVM 的快速发展的趋势，其相关研究也已经如火如荼地展开.

随着 Samsung、Micron、IBM 等大公司接连进行实验级别芯片的研究，并开始量产面向移动设备的 PCM 芯片，利用相变存储器改变传统计算机存储系统的时机已经到来！相信对 PCM 的一系列研究，也将成为今后对其他新型 NVM 研究的重要途径和切入点，并以此为契机，加快步伐研究 NVM 对整个计算机系统带来的变革！

参　考　文　献

[1] Song Zhi-Tang. Phase Change Memory. Beijing: Science Press, 2010(in Chinese)
(宋志堂. 相变存储器. 北京：科学出版社，2010)

[2] Eilert S, Leinwander M, et al. Phase change memory: A new memory enables new memory usage models//Proceedings of the IEEE International Memory Workshop 2009 (IMW'09). Monterey, USA, 2009: 1-2

[3] Chen Y C, Rettner C T, Raoux S, et al. Ultra-thin phase-change bridge memory device using GeSb//Proceedings of the International Electron Devices Meeting. San Francisco, USA, 2006: 777-780

[4] Lee B C, Ipek E, Mutlu O, et al. Architecting phase change memory as a scalable dram alternative. ACM SIGARCH Computer Architecture News, 2009, 37(3): 2-13

[5] Lee B C, Zhou P, Yang J, et al. Phase-change technology and the future of main memory. IEEE Micro, 2010, 30(1): 143-143

[6] Zhou P, Zhao B, Yang J, et al. A durable and energy efficient main memory using phase change memory technology// Proceedings of the 36th Annual International Symposium on Computer Architecture. New York, USA, 2009: 14-23

① The Gem5 Simulator [EB/OL]. http://www.gem5.org/Main_Page

[7] Qureshi M K, Franceschini M M, Lastras-Montaño L A, et al. Morphable memory system: A robust architecture for exploiting multi-level phase change memories//Proceedings of the 37th Annual International Symposium on Computer Architecture. New York, USA, 2010: 153-162

[8] Chen J, Winter Z, Venkataramani G, et al. rPRAM: Exploring redundancy techniques to improve lifetime of PCM-based main memory//Proceedings of the Parallel Architectures and Compilation Techniques (PACT) International Conference. Galveston TX, USA, 2011: 201-202

[9] Ferreira A P, Zhou M, Bock S, et al. Increasing pcm main memory lifetime//Proceedings of the Conference on Design, Automation and Test in Europe. Leuven, Belgium, 2010: 914-919

[10] Raoux S, Burr G W, Breitwisch M J, et al. Phase-change random access memory: A scalable technology. IBM Journal of Research and Development, 2008, 52(4.5): 465-479

[11] Liu Jin-Lei, Li Qiong. Application research on new-volatile phase change memory PCM. Journal of Computer Research and Development, 2012, 49(Suppl.): 90-93(in Chinese)
(刘金垒, 李琼. 新型非易失相变存储器 PCM 应用研究. 计算机研究与发展, 2012, 49(增刊): 90-93)

[12] Lee E, Yoo S, Jang J E, et al. Shortcut-JFS: A write efficient journaling file system for phase change memory//Proceedings of the IEEE 28th Symposium on Mass Storage Systems and Technologies (MSST). San Diego, USA, 2012: 1-6

[13] Narayanan D, Hodson O. Whole-system persistence//Proceedings of the 17th International Conference on Architectural Support for Programming Languages and Operating Systems. New York, USA, 2012: 401-410

[14] Kim K, Lee S W, Moon B, et al. IPL-P: In-page logging with PCRAM. Proceedings of the VLDB Endowment, 2011, 4(12): 1363-1366

[15] Xu Lin-Hai, Chen Xiao-Gang, Song Zhi-Tang, et al. Design of audio recording and playing system based on phase change memory. Journal of Functional Materials and Devices, 2012, 18(4): 327-331(in Chinese)
(许林海, 陈小刚, 宋志堂等. 基于相变存储器的音频存储播放系统设计. 功能材料与器件学报, 2012, 18(4): 327-331)

[16] Liu Bo, Song Zhi-Tang, Feng Song-Lin. Current situation and developing trend on phase-change memory in China. Micronanoelectronic Technology, 2007, 44(2): 55-61(in Chinese)
(刘波, 宋志棠, 封松林. 我国相变存储器的研究现状与发展前景. 微纳电子技术, 2007, 44(2): 55-61)

[17] Fan Chong-Fei, Yang Yan, Zhang Si-Mi, et al. Review of patent technology related to phase change memory. Metallic Functional Materials, 2013, 20(3): 54-59(in Chinese)
(范崇飞, 杨燕, 张思秘等. 相变存储器专利技术现状和趋势分析. 金属功能材料, 2013, 20(3): 54-59)

[18] Qureshi M K, Franceschini M M, Lastras-Montaño L A. Improving read performance of phase change memories via write cancellation and write pausing//Proceedings of the IEEE 16th International Symposium on High Performance Computer Architecture (HPCA). Bangalore, India, 2010: 1-11

[19] Jiang L, Zhao B, Zhang Y, et al. Improving write operations in MLC phase change memory//Proceedings of the IEEE 18th International Symposium on High-Performance Computer Architecture. New Orleans, USA, 2012: 1-10

[20] Qureshi M K, Franceschini M M, Jagmohan A, et al. PreSET: Improving performance of phase change memories by exploiting asymmetry in write times//Proceedings of the 39th Annual International Symposium on Computer Architecture (ISCA). Portland, USA, 2012: 380-391

[21] Yue J, Zhu Y. Accelerating write by exploiting PCM asymmetries//Proceedings of the 19th IEEE International Symposium on High Performance Computer Architecture (HPCA'13). Shenzhen, China, 2013: 182-193

[22] Yue J, Zhu Y. Exploiting subarrays inside a bank to improve phase change memory performance//Proceedings of the Conference on Design, Automation and Test in Europe. San Jose, USA, 2013: 386-391

[23] Joo Y, Niu D, Dong X, et al. Energy-and endurance-aware design of phase change memory caches//Proceedings of the Conference on Design, Automation and Test in Europe. Leuven, Belgium, 2010: 136-141

[24] Cho S, Lee H. Flip-N-Write: A simple deterministic technique to improve PRAM write performance, energy and endurance//Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture. New York, USA, 2009: 347-357

[25] Yun J, Lee S, Yoo S. Bloom filter-based dynamic wear leveling for phase-change RAM//Proceedings of the Conference on Design, Automation and Test in Europe. Dresden, Germany, 2012: 1513-1518

[26] Xu W, Liu J, Zhang T. Data manipulation techniques to reduce phase change memory write energy//Proceedings of the 14th ACM/IEEE International Symposium on Low Power Electronics and Design. New York, USA, 2009: 237-242

[27] Hay A, Strauss K, Sherwood T, et al. Preventing PCM banks from seizing too much power//Proceedings of the IEEE/ACM International Symposium on Microarchitecture. Porto Alegre, Brazil, 2011: 186-195

[28] Bock S, Childers B, Melhem R, et al. Analyzing the impact of useless write-backs on the endurance and energy consumption of PCM main memory//Proceedings of the IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS). TX, USA, 2011: 56-65

[29] Ranganathan P, Chang J. (Re) Designing data-centric data centers. IEEE Micro, 2012, 32(1): 66-70

[30] Yang J, Minturn D B, Hady F. When poll is better than interrupt//Proceedings of the 10th USENIX Conference on File and Storage Technologies. Berkeley, USA, 2012: 3-3

[31] Coburn J, Caulfield A M, Akel A, et al. NV-Heaps: Making persistent objects fast and safe with next-generation, non-volatile memories//Proceedings of the 16th International Conference on Architectural Support for Programming Languages and Operating Systems. New York, USA, 2011: 105-118

[32] Caulfield A M, Mollov T I, Eisner L A, et al. Providing safe, user space access to fast, solid state disks//Proceedings of the 17th International Conference on Architectural Support for Programming Languages and Operating Systems. New York, USA, 2012: 387-400

[33] Kong J, Zhou H. Improving privacy and lifetime of PCM-based main memory//Proceedings of the IEEE/IFIP International Conference on Dependable Systems and Networks (DSN). Chicago, USA, 2010: 333-342

[34] Chhabra S, Solihin Y. i-NVMM: A secure non-volatile main memory system with incremental encryption//Proceedings of the 38th Annual International Symposium on Computer Architecture (ISCA). San Jose, USA, 2011: 177-188

[35] Kwon S, Kim D, Kim Y, et al. A case study on the application of real phase-change RAM to main memory subsystem//Proceedings of the Conference on Design, Automation and Test in Europe. San Jose, USA, 2012: 264-267

[36] Qureshi M K, Srinivasan V, Rivers J A. Scalable high performance main memory system using phase-change memory technology. ACM SIGARCH Computer Architecture News, 2009, 37(3): 24-33

[37] Dhiman G, Ayoub R, Rosing T. PDRAM: A hybrid PRAM and DRAM main memory system//Proceedings of the 47th Design Automation Conference, 46th ACM/IEEE. San Francisco, USA, 2009: 664-669

[38] Zhang W, Li T. Exploring phase change memory and 3D die-stacking for power/thermal friendly, fast and durable memory architectures//Proceedings of the 18th International Conference on Parallel Architectures and Compilation Techniques. Raleigh, USA, 2009: 101-112

[39] Ramos L E, Gorbatov E, Bianchini R. Page placement in hybrid memory systems//Proceedings of the International Conference on Supercomputing. New York, USA, 2011: 85-95

[40] Mladenov R. An efficient non-volatile main memory using phase change memory//Proceedings of the 13th International Conference on Computer Systems and Technologies. New York, USA, 2012: 45-51

[41] Baek S, Lee H G, Nicopoulos C, et al. A dual-phase compression mechanism for hybrid DRAM/PCM main memory architectures//Proceedings of the Great Lakes Symposium on VLSI. New York, USA, 2012: 345-350

[42] Park H, Yoo S, Lee S. Power management of hybrid DRAM/PRAM-based main memory//Proceedings of the 48th Design Automation Conference. New York, USA, 2011: 59-64

[43] Hu J, Zhuge Q, Xue C J, et al. Software enabled wear-leveling for hybrid PCM main memory on embedded systems//Proceedings of the Design, Automation & Test in Europe Conference & Exhibition (DATE). Grenoble, France, 2013: 599-602

[44] Sun G, Joo Y, Chen Y, et al. A hybrid solid-state storage architecture for the performance, energy consumption, and lifetime improvement//Proceedings of the IEEE 16th International Symposium on High Performance Computer Architecture (HPCA). Bangalore, India, 2010: 1-12

[45] Liu D, Wang T, Wang Y, et al. A block-level flash memory management scheme for reducing write activities in PCM-based embedded systems//Proceedings of the Conference on Design, Automation and Test in Europe. San Jose, USA, 2012: 1447-1450

[46] Akel A, Caulfield A M, Mollov T I, et al. Onyx: A protoype phase change memory storage array//Proceedings of the 3rd USENIX Conference on Hot Topics in Storage and File Systems. Berkeley, USA, 2011: 2-2

[47] Caulfield A M, De A, Coburn J, et al. Moneta: A high-performance storage array architecture for next-generation, non-volatile memories//Proceedings of the 43rd Annual IEEE/ACM International Symposium on Microarchitecture. Washington, USA, 2010: 385-395

[48] Shao Z, Chang N, Dutt N. PTL: PCM translation layer//Proceedings of the IEEE Computer Society Annual Symposium on VLSI (ISVLSI). Amherst, USA, 2012: 380-385

[49] Lam C H. Storage class memory//Proceedings of the 10th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT). Shanghai, China, 2010: 1080-1083

[50] Condit J, Nightingale E B, Frost C, et al. Better I/O through byte-addressable, persistent memory//Proceedings of the ACM SIGOPS 22nd Symposium on Operating Systems Principles. New York, USA, 2009: 133-146

[51] Wu X, Reddy A L. SCMFS: A file system for storage class memory//Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis. New York, USA, 2011: 39

[52] Dong X, Jouppi N P, Xie Y. PCRAMsim: System-level performance, energy, and area modeling for phase-change RAM//Proceedings of the 2009 International Conference on Computer-Aided Design. New York, USA, 2009: 269-275

[53] Magnusson P S, Christensson M, Eskilson J, et al. Simics: A full system simulation platform. Computer, 2002, 35(2): 50-58

[54] Binkert N L, Dreslinski R G, Hsu L R, et al. The M5 simulator: Modeling networked systems. IEEE Micro, 2006, 26(4): 52-60

[55] Wang D, Ganesh B, Tuaycharoen N, et al. DRAMsim: A memory system simulator. ACM SIGARCH Computer Architecture News, 2005, 33(4): 100-107

MAO Wei, born in 1989, M. S. candidate. His research interests include computer storage system, novel non-volatile memory devices, solid state storage.

LIU Jing-Ning, born in 1957, Ph. D., professor. Her research interests are computer system structure, computer storage system, high-speed interface and channel technology.

TONG Wei, born in 1977, Ph. D., lecturer. Her research interests include massive networked storage system, solid state storage and input/output virtualization.

FENG Dan, born in 1970, Ph. D., professor. Her research interests include information storage system, network storage, solid state storage, performance evaluation.

LI Zheng, born in 1992, Ph. D. candidate. His current research interests include novel non-volatile memory devices, solid state storage.

ZHOU Wen, born in 1985, Ph. D. candidate. His research interests include reconfigurable computing and novel non-volatile memory devices.

ZHANG Shuang-Wu, born in 1990, M. S. candidate. His research interests include computer storage system, solid state storage, novel non-volatile memory devices.

Background

Storage technology research based on Phase-Change Memory (PCM) is a relatively new direction in the field of computer storage, which includes utilization of PCM technology in system-level applications, related memory and storage technologies, and integration strategies of PCM with existing storage systems. Numerous studies focused on above those have been carried out both at home and abroad.

In this paper, we summarized the research on storage technologies based on phase-change memory, analyzed the advantages and weaknesses of current technologies, and forecasted the future directions on research and development of those fields.

Our research team focuses on PCM-based systems, and we have carried out series of research. We developed a PCM simulation with simple read/write functions to simulate working status of real PCM chips and established a hardware prototype platform based on Micron P8P chips which could read and write correctly. Currently we are doing some research on hybrid storage and extended memory based on this hardware platform. The former can use PCM to improve performance and extend lifetime of Flash. The latter can settle the scalability problem of DRAM, improve main memory performance and reduce energy consumption significantly; Meanwhile, we conducted some research on chip-level and channel-level reading/writing parallelism of PCM on the prototype platform, and progress had been made. In future, our team will concentrate on the prototype platform, and we will do further research on impact of NVM, represented by PCM, on computer architecture, and other work, such as lifetime, writing performance and power consumption of PCM.

Our work is supported by the National Basic Research Program (973 Program) of China (No. 2011CB302301), the National Natural Science Foundation of China (Nos. 61303046 and 61173043), the National Science Fund for Distinguished Young Scholars (No. 61025008) and the Fundamental Research Funds for the Central Universities (HUST: 2013TS042).