

一种混合内存系统访存热点控制方法

王 强, 陈 岚, 郝晓冉

(中国科学院 微电子研究所, 北京 100029)

摘 要: 相变内存(Phase-Change Memory, PCM)被认为是最有可能取代 DRAM 的非挥发存储器件, 它与 DRAM 读取性能相当且无需动态刷新, 但存在着读写非对称和寿命受限的问题, 因此在设计新型计算机系统存储结构时被考虑与 DRAM 共同构建内存子系统. 针对 PCM-DRAM 平面混合内存架构中由于访存局部性所引起的热点区域集中读写访问所带来的不利影响, 提出了一种基于访存热点控制的有效策略, 来降低内存系统功耗和增加 PCM 使用寿命.

关键词: 相变存储器; 混合内存系统; 有效管理; 热点控制

中图分类号: TP333.1

文献标识码: A

文章编号: 1000-7180(2014)01-0001-04

Efficient Management with Hotspots Control for Hybrid Memory System

WANG Qiang, CHEN Lan, HAO Xiao-ran

(Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100029, China)

Abstract: As a popular Non-Volatile Memory technology, Phase-Change Memory (PCM) is considering as the most potential candidate for DRAM due to its comparable read latency with DRAM, high capacity and low idle power. But hybrid memory architecture which consists of both PCM and DRAM is suggested due to asymmetric read-write and limited endurance of PCM device. In this paper, we propose an efficient management with hotspots controlling to reduce power consumption and improve PCM lifetime for PCM-DRAM flat hybrid main memory system to avoid the dense read or write access to memory hotspots caused by locality.

Key words: phase-change memory; hybrid main memory; efficient management; hotspots control

1 引言

计算机内存系统所消耗的能量不断增长, 甚至在整个系统中已经高达 40%^[1]. 随着新型非挥发存储技术的快速发展, 相变内存 PCM^[2] 逐渐具备了替代 DRAM 内存器件的潜力. 相变存储技术的发展, 得益于 1990 年代对于相变材料的研究成果. 相变材料可以用作双稳态的存储材料, 其存储单元设计通常采用 1T/1R 的结构, 利用同 DRAM 类似的 CMOS 外围控制电路完成读写操作及与系统的连接.

PCM 作为一种非挥发存储器件具有高存储密度和低功耗的特性, 与 DRAM 有相近的存取速度

和相同的存取方式, 也可以按位寻址, 但保存数据无需刷新且尺寸收缩性比 DRAM 好. 利用 PCM 取代 DRAM 构成未来计算机体系中的高性能低功耗内存系统, 已经成为目前研究的热点和重点. 但是 PCM 也有自身缺陷存储单元的读写特性具有不对称性, 数据写入速度慢、写入功耗高, 且存储单元写入寿命通常只有 $10^7 \sim 10^8$ 次. 为解决上述问题, 研究者提出用 PCM 部分取代 DRAM 来共同构建混合内存系统的方法, 弥补 PCM 的不足的同时能充分发挥 PCM 低功耗的特性. 如图 1 与传统内存结构的对比, PCM 和 DRAM 被置于统一的地址空间内.

为了合理有效地构建 PCM-DRAM 混合内存系统, 本文提出一种基于预测和地址重映射的控制

收稿日期: 2013-04-16; 修回日期: 2013-06-22

基金项目: 中国科学院战略性先导科技专项(XDA06020401)

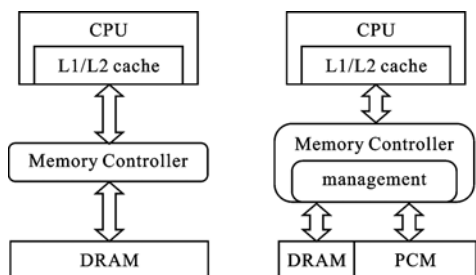


图1 传统内存结构与平面混合内存结构对比

机制. 针对程序访存局部性所引起的热点区域集中读写访问, 利用指数平滑预测的方法找出 DRAM 中读取热点页帧, 基于地址映射和页帧交换机制将这些 DRAM 访存热点与 PCM 存储区域中的写入热点页帧交换存储位置, 实现对整个内存系统的有效管理. 这里的物理页帧(frame)是对应物理内存空间中的存储结构.

2 平面混合内存热点控制

2.1 访存热点定义

基于 PCM 和 DRAM 的平面混合内存系统利用两种不同特性的物理内存器件, PCM 的数据写入延时和功耗远高于 DRAM 且寿命有限, 而 DRAM 虽然读取速度比 PCM 快, 但是数据读取及待机功耗更大. 由于程序访存行为具有局部性的特点, 如图 2 所示, 既可能出现读取操作集中的物理页帧也可能出现写入操作集中的物理页帧. 我们把 PCM 中写入操作集中的物理页帧和 DRAM 中读取操作集中的物理页帧定义为访存热点(Hotspots), 这些热点直接影响 PCM-DRAM 混合内存系统性能和功耗, 位于内存地址空间中的每一个物理页面都有可能成为这样的热点. 由此可见, 建立和管理上述新型内存系统的原则就在于, DRAM 中的读访问热点页帧需要与 PCM 区域中写访问热点页帧互换存储位置, 以降低读取和待机功耗并减少 PCM 损耗. 找出混合内存系统中的访存热点是有效控制管理的关键. 内存空间的控制尺度与体系结构所支持的分页机制保持一致(如 x86 的 4 kB 页、ALPHA 的 8 kB 页).

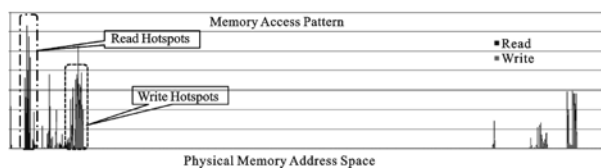


图2 访存分布及读写热点区域示意图

PCM 的读写具有非对称性, 写操作的延迟和功

耗大于 DRAM 且使用寿命受限, 因此内存控制机制为 PCM 每个物理页帧建立一个写计数表, 记录每个页面的写入次数. 将程序对内存的访问序列按控制周期划分, 系统同时记录当前周期内被访问过的 PCM 中写入次数最多的物理页帧作为 DRAM 热点的替换帧, 即针对每次 PCM 访问如果发现该处写入次数高于替换帧则更新 PCM 替换帧号记录.

对于 DRAM 存储区域, 需要描述物理页帧的访存特征: 对每一个物理页帧每个周期内的读写特征用 f_n 来描述其读写密集情况. 计算公式(1)如下:

$$f_n = 1 \times F_r + (-1) \times F_w \quad (1)$$

式中, f_n 表示某个物理页面在一段内存访问序列中程序访存的读写密集程度, F_r 和 F_w 分别表示当前一段序列中内存读和写所占比例. 如果一个物理页面的内存读取所占比重越高, f_n 越接近于 1; 而如果一个物理页面的内存写入比重越高, f_n 越接近 -1. 随着程序的执行, 每一个周期内读写特征可以在该周期结束后用访存特征 f_n 明确表示, 如果定义该周期之前的页面整体读写特征 I_n , 利用简单指数平滑^[3]的方法可以判断出到下一周期物理页面的读写特征 I_{n+1} , 两者之间的关系如式(2)所示:

$$I_{n+1} = \alpha \times I_n + (1 - \alpha) \times f_n \quad (2)$$

当前周期内 I_n 最接近 1 的即列为 DRAM 访存热点区域. 其中, 系数 α 控制了早期预测值和当前记录值的权重关系, 越小则越侧重于考虑当前周期内访存特征值的影响, 越大则越侧重于考虑预测历史的影响. 由于程序对物理页面多数情况下出现阶段性的集中访问, 所以 α 应该选取较小的数值来更多地考虑访存特征的实时性变化, 实验表明 0.4 较为合适.

2.2 地址映射控制机制

为了借助硬件来实现对混合内存的控制从而提高效率, 系统采用地址重映射来实现从物理地址到实际内存访问地址转换的方式. 内存控制机制维护一个地址重映射表, 如图 3 所示, 以从总线传入的 CPU 访存请求所指向的物理地址页帧号(frame number)为查询入口, 通过查找重映射表来获得该物理地址所对应的内部实际访存地址页帧号, 再结合页内偏移量得出实际访存地址并送入内存控制器完成访存. 如上文所述, 8 kB 的页面用末 13 位地址来表示页内偏移量(offset), 剩余位表示页帧号. 为了保证数据访问的一致性, 物理地址到实际访存地址的转换必须保证是一一映射的关系. 另一方面, 根

据实际访问位置和读写请求类型记录当前物理页帧的读写情况以提供热点控制机制做出判断. 对一个控制周期中所产生的 PCM 和 DRAM 访存热点, 控制器利用两个同物理页帧同样大小的 transfer buffer 来完成内存中两帧的交换, 受限把 PCM 和 DRAM 各自的访存热点物理页帧拷贝到 buffer 中, 并根据所记录的两帧号拷回各自位置, 同时更新地址映射关系.

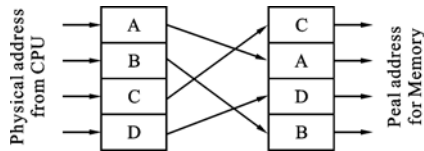


图 3 物理地址映射机制

3 实验分析

3.1 实验方法

实验采用基于 Gem5^[4] 体系结构仿真系统来进行, 并在内存系统中构建了按地址空间分离的 PCM-DRAM 平面混合结构, 利用基于 trace 的方法分析 10^7 次程序访存行为, 仿真平台整体框架如图 4 所示. 为减少不相关部件对内存系统评估过程的影响, 系统采用 2 GHz In-order CPU, 32 kB / 64 kB 的指令/数据 cache. 系统内存总量为 128 MB, 因为 PCM 的存储密度预计将会达到 DRAM 的 4 倍, 所以对于混合内存系统将其其中 1/4 容量的 PCM 用 DRAM 取代, 测试程序选自 meidabench2^[5] 中的 12 个程序并运行在 Linux 操作系统下. 本文所对比的内存系统结构包括以下四种 (1) PCM-only 内存: 仅有 128 MB PCM, (2) DRAM-only, 仅有 128 MB DRAM, (3) PCM + DRAM, PCM 32 MB 和 DRAM 96 MB 混合内存系统和 (4) PCM+DRAM w HC, 基于热点控制的混合内存系统. 本文综合了文献[6-7]里所提供的参数来对整个内存系统的性能和功耗进行评估. 具体参数如表 1 所示, 其中功耗参数都是按 64 B 的 cache line size 给出的.

表 1 PCM 和 DRAM 特性参数

Parameter	PCM	DRAM
Latency		
Read	28 ns	15 ns
Write	150 ns	22 ns
Row buffer hit	15 ns	15 ns
Power		
Read	0.05 nJ/bit	0.1 nJ/bit
Write	0.5 nJ/bit	0.1 nJ/bit
Idle	0.005 W	1.3 W/GB

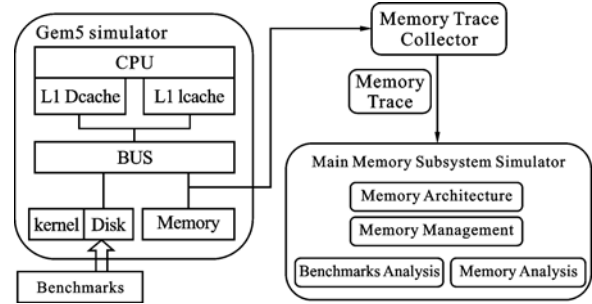


图 4 仿真平台整体框架

3.2 结果分析

所有的仿真结果(图 5)都根据 DRAM-only 内存系统做了归一化处理. 综合 12 个测试程序结果可以看出, 完全使用 PCM 代替 DRAM 的系统中, 由于 PCM 访问延迟比 DRAM 高, 特别是内存写入延

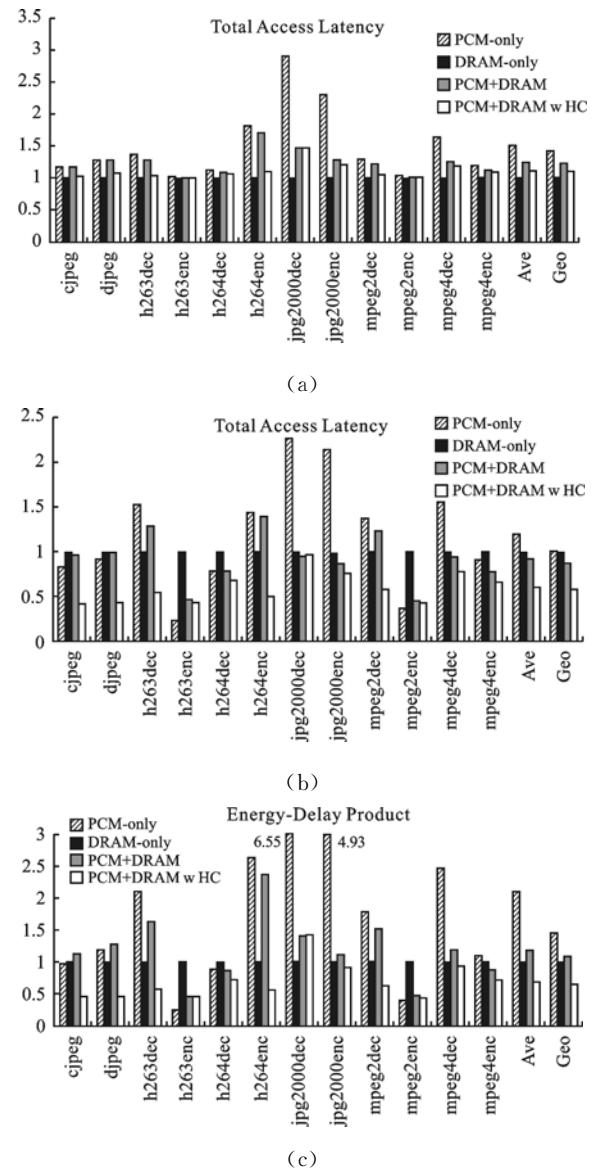


图 5 归一化的内存系统访问延时、功耗及功耗延时积比较

迟高出 10 倍以上,因此给系统延迟带来了 44% 的增长,这与程序访存的读写比例有直接关系,对于读写操作相当的程序如 h264enc、jpg2000dec 和 jpg2000enc,延迟增长十分明显,说明过多的内存写操作是导致系统性能下降的主要因素。

将位于地址空间内的一小部分 PCM 换成 DRAM 后,系统延迟的增长降至 23%,特别是在 PCM-only 中性能恶化比较严重的三个程序都有改善,同时功耗也比 DRAM-only 系统下降了 11%,说明在这一部分地址空间内发生的访问发挥了 DRAM 读写速度快和写入功耗低的优势。但是 cjpeg 和 djpeg 的功耗反而比 PCM-only 系统分别高出了 12% 和 7%,主要是因为程序对 DRAM 区域的访问不密集,无法利用 DRAM 提升数据写入性能,却引入了 DRAM 功耗方面的缺陷。

在加入对混合内存系统的热点控制之后,利用平均仅 10% 的性能开销换取了 40% 的功耗节约,相对于 DRAM-only 内存系统功耗延时积的收益平均可达 36%,其中 cjpeg、djpeg、h263dec、h264enc 和 mpeg2dec 功耗优化可达 50% 左右。但 jpg2000dec 的功耗延时积却高于 DRAM 系统,主要因为读写分布对 DRAM 区域占用率较高,我们考虑通过适当增加 DRAM 容量来进行改善,这将是后续工作的一个内容。

新型混合内存系统架构对 PCM 器件寿命的提升也有很大帮助。考虑访存中写操作的分布不均衡性,我们利用文献[8]给出的 PCM 损耗模型并结合对 PCM 存储区域写入操作的最大值来对结果进行一定程度的更正,如式(3)所示:

$$\text{Lifetime} = \frac{E \times \text{Capacity}}{\text{Bandwidth}} \times \left(1 - \frac{W_{\max}}{E}\right) \times 2^{-25} \quad (3)$$

在加入少量 DRAM 之后分担了部分原有 PCM 的损耗,其寿命相比 PCM 存储系统平均提升 5 倍。而加入热点控制之后,由于有效控制和转移了 PCM 地址空间内的热点区,PCM 存储模块的平均寿命提升了 16 倍。

4 结束语

相变内存被用于未来计算机内存子系统的开发,但它面临被密集写入和过度损耗的问题。本文基于相变内存 PCM 建立的 PCM-DRAM 平面混合内存系统,提出了一种针对 PCM 和 DRAM 地址空间

内所出现的访存热点控制管理方法。实验表明,该机制能有效降低 PCM-DRAM 混合内存系统功耗并提高 PCM 器件的使用寿命。

参考文献:

- [1] Lefurgy C, Rajamani K, Rawson F, et al. Energy management for commercial servers[J]. Computer, 2003, 36(12): 39-48.
- [2] Burr G W, Breitwisch M J, Franceschini M, et al. Phase change memory technology[J]. Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures, 2010, 28(2): 223-262.
- [3] Hwang C H, Wu A C H. A predictive system shut-down method for energy saving of event-driven computation[C]//Proceedings of the 1997 IEEE/ACM international conference on Computer-aided design. San Jose, IEEE Computer Society, 1997: 28-32.
- [4] Binkert N, Beckmann B, Black G, et al. The gem5 simulator[J]. ACM SIGARCH Computer Architecture News, 2011, 39(2): 1-7.
- [5] Fritts J E, Steiling F W, Tucek J A. MediaBench II video: expediting the next generation of video systems research[C]//Electronic Imaging 2005, International Society for Optics and Photonics. Brazil, 2005: 79-93.
- [6] Dhiman G, Ayoub R, Rosing T. PDRAM: a hybrid PRAM and DRAM main memory system[C]//Design Automation Conference, DAC'09. 46th ACM/IEEE. San Francisco, IEEE, 2009: 664-669.
- [7] Seok H, Park Y, Park K H. Migration based page caching algorithm for a hybrid main memory of DRAM and PRAM [C]//Proceedings of the 2011 ACM Symposium on Applied Computing. Taichung, Taiwan, ACM, 2011: 595-599.
- [8] Qureshi M K, Srinivasan V, Rivers J A. Scalable high performance main memory system using phase-change memory technology [J]. ACM SIGARCH Computer Architecture News, ACM, 2009, 37(3): 24-33.

作者简介:

王 强 男,(1985—),博士研究生.研究方向为非挥发计算技术与计算机系统结构。

陈 岚 女,(1968—),博士,研究员,博士生导师.研究方向为超深亚微米芯片设计方法学、计算机体系结构及物联网关键技术。

郝晓冉 女,(1981—),博士,助理研究员.研究方向为多媒体通信技术及物联网智能终端体系结构。