doi:10.3969/j.issn.1673-9833.2021.03.011

基于频率时效感知的混合内存写冷热页面调度

刘　兵[1, 2]，汪令辉[3]，张　涛[2]，陈友良[4]

（1. 中国科技大学 计算机科学与技术学院，安徽 合肥 230027；2. 铜陵职业技术学院 信息工程系，安徽 铜陵 244061；
3. 铜陵有色金属集团公司，安徽 铜陵 244000；4. 中国安全生产科学研究院，北京 100012）

摘　要：为了提高相变存储器和DRAM所组成混合内存中PCM的耐久性，减少页面在PCM上的写操作，通过一个新公式结合写访问频率的统计和最近写访问间隔对页面写冷热的影响，计算频率时效的写权值，并根据最近写频率和计算的频率时效权值来进行混合内存页面冷热页的调度。通过仿真分析比较，这种方法可以减少 PCM 页面的写次数。

关键词：非易失存储器，相变存储；混合内存；页面写预测

中图分类号：TP333　　**文献标志码**：A　　**文章编号**：1673-9833(2021)03-0074-06

引文格式：刘　兵，汪令辉，张　涛，等 . 基于频率时效感知的混合内存写冷热页面调度 [J]. 湖南工业大学学报，2021，35(3)：74-79.

A Hybrid Memory Writing Cold and Hot Page Scheduling Based on Frequency Time Interval

LIU Bing[1, 2]，WANG Linghui[3]，ZHANG Tao[2]，CHEN Youliang[4]

（1. School of Computer Science and Technology，University of Science and Technology of China，Heifei 230027，China；
2. Department of Information Engineering，Tongling Vocational and Technical College，Tongling Anhui 244061，China；
3. Tongling Nonferrous Metals Group Co., Ltd.，Tongling Anhui 244000，China；
4. China Academy of Safety and Science & Technology，Beijing 100012，China）

Abstract：In view of an improvement of the durability of PCM in the hybrid memory composed of PCM and DRAM, as well as a reduction of the page writing operation on PCM, a new formula is used to calculate the write weight value of frequency aging with the statistics of the write access frequency and the influence of the latest write access interval on the page writing heat and cold combined together, followed by a calculation of the hot and cold pages of the hybrid memory according to the latest write frequency and the calculated frequency aging weight. Based on the simulation analysis and comparison, it is verified that this proposed method helps reduce the number of PCM page writing.

Keywords：non-volatile memory；phase change memory；hybrid main memory；page write prediction

0　引言

随着大数据技术、人工智能和工业互联网等技术的发展，需处理的数据量越来越多，对内存的节能、存储密度、随机写、高并发性随机读和实时处理分析等都提出了更高的要求。当前主要的内存技

收稿日期：2020-11-11

基金项目：安徽省高校自然科学研究基金资助重点项目（KJ2020A0971）；安徽省高校质量工程基金资助项目（2018mooc222，2018mooc230）

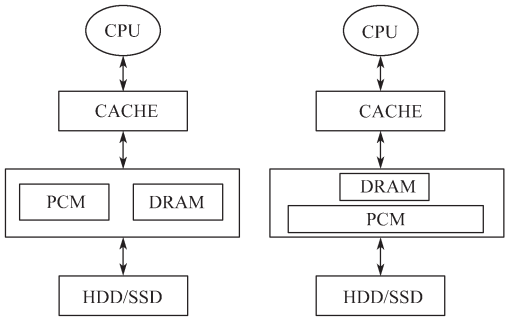
作者简介：刘　兵（1974-），安徽肥东县人，男，铜陵职业技术学院高级工程师，副教授，主要从事数据库技术和大数据方面的教学与研究，E-Mail：yg_liu@163.com

术是 DRAM（dynamic random access memory），要通过不断地刷电来保持数据，能源的消耗比较大。另外，DRAM 的存储集成也已经接近极限。非易失性存储（non-volatile memory，NVM）技术为解决这一问题提供了一种新方法，其中以相变存储器（phase change memory，PCM）[1] 性能最为突出，其作为近些年存储技术发展的热点技术，有着广泛的应用前景。PCM 相对于 DRAM 的优点是存储密度较大、功耗低；缺点是写入的速度比 DRAM 慢、写的次数有限。因此，减少 PCM 的写操作，提高其写耐久性，是许多研究者探讨的问题。

1　相关研究

1.1　混合内存架构

针对 PCM 和 DRAM 的特点，目前的研究集中在将 PCM 和 DRAM 二者的优点结合。在混合内存 [2–5] 结构上的使用，分为同级混合内存和层次混合内存，如图 1 所示。



a）同级混合　　b）层次混合

图 1　同级混合和层次混合内存结构框图

Fig. 1　Peer and hierarchical hybrid memory structure

1）同级混合内存 [6]（图 1a）利用 PCM 字节寻址的特点，内存由 PCM 和 DRAM 两部分构成，并当作一个整体统一编址。访问时，根据页面特点，将页面分别放入 PCM 或 DRAM。2）层次混合内存 [7]（图 1b）。其将 DRAM 作为 PCM 的缓存，先访问 DRAM，如 DRAM 没有命中，再访问 PCM，通过 DRAM 的写无限性缓冲 PCM 的写有限性。

1.2　已有混合内存冷热页的判定

由于 PCM 的写耐久性有一定的次数限制，读写不均衡，写时间较长，所以在混合内存缓冲区管理调度策略的设计中，通常要达到 2 个目标：1）减少 PCM 的写次数，延长 PCM 的使用年限；2）提高访问时缓冲区页面的命中率，从而减少页面调度时资源的消耗及同时产生的 PCM 写操作。

Seok H. 等 [8–9] 提出以“最近最少使用”（least recently used，LRU）算法为基础的 LRU-WPAM（LRU with prediction and migration）算法，增加了一个页面的读写预测，根据判断缓冲区页面是否命中。当未命中时，用最近最少使用页面置换；命中时，根据读写请求修改页面权值，再判断权值是否达到阀值。判断页面是读倾向高的页面（“读热页”），还是写倾向高的页面（“写热页”），如果页面达到阀值，将“读热页”移动移动进 PCM，将“写热页”移动进 DRAM。Lee S. 等 [10] 提出 CLOCK-DWF 算法，将 DRAM 和 PCM 各组成一个环状队列，当空间充足时，把读请求页面存入混合内存的 PCM 中，写请求页面放入混合内存的 DRAM 中。当 DRAM 空间不足时，进行冷热页的调度，将写冷页调度进入 PCM，PCM 的空间不足时，使用 CLOCK 算法调度页面。类似的还有 Chen K. M. 等 [11] 提出的 MHR-LRU（maintain-hit-ratio LRU），刘兵等 [12] 提出的 FWLRU（favors write LRU）策略等。

以上算法都涉及页面冷热页的判定。LRU-WPAM 中给每个页面设置权值，当页面是“读请求”时，权值增加，当是“写请求”时，权值减少，通过权值和阀值的比较判定页面的读写热页类型。CLOCK-DWF 通过每个页面写次数来判断“写热页”和“写冷页”。其它几种算法也都通过次数来判断页面的冷热。

2　频率时效感知页面划分

在 PCM 的读写操作中，根据 PCM 的特性，读操作和 DRAM 中的操作区别不大，写操作的使用对于 PCM 的扬长避短有决定性的作用。如果能准确及时地预测出“写冷页”和“写热页”，既可利用 PCM 的低能耗、存储密度大的特点，又可避免写操作有限的缺点，从而提高 PCM 的写耐久性，同时提高页面的命中率。但已有的冷热页面预测或者判定方法，忽略了如下几个方面的问题：

1）页面访问有局部性

存储系统负载访问有局部性 [13] 的特点，即写操作聚集在若干页面上。在某一时间段内，若干页面访问次数很多，比较密集，其它页面没有访问或者零星访问。

2）“写热页”和“写冷页”和页面调用的时段有关页面调用的阶段性

某些页面写入后，可能很长时间不再调用，也可能阶段性爆发，并且在较近时段发生过写操作页面为“写热页”的概率比较大，即局部爆发和爆发的间隔时段有关。

针对上述问题，本文提出根据先前访问的频率距离现在访问的间隔、当前局部爆发访问的特点，将页面的局部写频率和上次的高频访问和最近高频访问的时间间隔来计算权值，并根据权值对页面“写”冷热进行划分，即频率时效写页面划分。

2.1 模型定义

页面写访问的局部爆发性、访问频率和最近写访问间隔对页面写的冷热有着直接影响。根据这一特点，本文通过写访问频率和最近写访问间隔、频率时效（frequency time interval，FTI）进行计算，预测页面的冷热度。首先引入如下几个概念。

局部写访问统计器（local write access statistics，LWAS）。如图 2 所示，该统计器为长度为 20 的队列，按照写访问的时间顺序，记录最近发生的 20 次写请求访问，并不重复统计最近每个页面的访问次数，计算得到局部写访问频率的值为 P_n。

图 2 局部写访问统计器

Fig. 2 Local write access statistics

频率时效页面写冷热权值计算公式为

$$W_2=\frac{W_1}{W_{\text{Dist}}}+CP_n。\tag{1}$$

式中：W_1 为当前页面上一次写访问时的权值；W_2 为出现最近页面写请求时计算的权值；W_{Dist} 为上一次最近的权值除以这个页面的最近写距离。C 取值时，先假定为 0.4~0.6 的区间，然后经过实验数据测定，取 0.5 比较合适。当页面没有出现过，W_1 没有值时，取默认值 0.45，W_{Dist} 取默认值 1。

高频访问页容器（high frequency access page container，HFAC）。该容器为一链表，由局部写访问统计器中 $P_n \geqslant 2$ 的页面按照时间次序组成，每个节点由页面序号和权值构成，权值根据式（1）得到。当 LWAS 中出现大于两次的页面时，将页面放入 HFAC，如果 HFAC 中出现过这个页面，在记录值后，将其从前面链表中删除。

混合内存 CLOCK 链表（hybrid memory CLOCK）。DRAM 和 PCM 混合内存页面整体链表，按 CLOCK 算法处理。

CLOCK-DRAM。将 DRAM 中页面按 CLOCK 算法组织并处理页面。

CLOCK-PCM。将 PCM 中页面按 CLOCK 算法组织并处理页面。

2.2 冷热页调度

DRAM 和 PCM 按照 4:1 的比例进行配置，DRAM 存储的页面个数为 *DSize*，PCM 存储的页面个数为 *PSize*。

频率时效页面写冷热度权值计算过程如下：

当出现页面写访问请示时，将页面序号放入局部写访问统计器头部；

统计局部写访问统计器，如果出现 $P_n \geqslant 2$ 的页面，读取高频访问页容器中各项，寻找是否存在此页面；

如果高频访问页容器有该页面，根据式（1）计算页面权值，放入高频访问页容器头部，删除先前的节点；如果没有该页面，计算权值放入高频访问页容器。

在混合内存中，由于 PCM 的写次数限定性，写页面存放以 DRAM 为优先，以频率时效的 CLOCK 算法（frequency time interval CLOCK，FTI-CLOCK）来实现页面的调度。出现写请求时，进行局部写访问频率统计并按式（1）计算权值，将权值插入 HFAC。如果出现 $P_n \geqslant 2$ 写请求时，调度原则如下：

如果页面在 CLOCK-DRAM 中，执行操作，如果页面在 CLOCK-PCM 中或者未命中，查找 DRAM 中是否存在空闲空间；

如果存在空闲空间，将页面调入 DRAM，如果没有空闲空间，比较 CLOCK-DRAM 和 HFAC，查找 CLOCK-DRAM 不在 HFAC 中的页面，如果存在，按照 CLOCK 算法将页面转换进 PCM 或者淘汰，如果没有，浏览 HFAC；

将 HFAC 中权值最小，且在 DRAM 中的页面，与页面置换。

2.3 算法过程

频率时效页面写冷热权值计算过程如算法 1 所示，其中输入为 W_1、W_{Dist} 和 P_n，输出为计算的权值 W_2。

算法 1 频率时效页面写冷热权值计算

```
if W1 is not null then
   W2=W1/WDist+Pn/2.0;
else
   W2=0.45+Pn/2.0;
endif;
return W2;
end.
```

频率时效的 CLOCK（FTI-CLOCK）调度过程如算法 2 所示，在 $P_n \geqslant 2$ 的页面写请求时，执行算法进行页面调度。

算法 2 FTI-CLOCK 页面调度

```
if page is in CLOCK-DRAM then
 page.write;
```

```
else
 if DRAM has free space then
  Scheduling page enters DRAM;
  CLOCK-DRAM.add;
 else
  foreach page in CLOCK-DRAM
   if page is not in HFAC then
    if page in CLOCK-PCM then
        Page exchange of CLOCK-PCM and CLOCK-DRAM;
     else
        CLOCK-DRAM eliminates pages and adds Miss pages;
      Hybrid Memory CLOCK alter;
     endif;
    else
     Page exchange of Minimum weight page in HFAC and in CLOCK-DRAM;
    endif;
   endif;
  endif;
 end.
```

3　实验仿真及分析

3.1　实验方法

为了模仿混合内存环境，通过在 ubuntu 18.04 系统上架设仿真模拟器 GEM5[14] + NVMain[15] 来实现 DRAM 和 PCM 混合内存实验环境。GEM5 是 GEMS 和 M5 结合的全系统模拟器，它有 ISA 和多种 CPU 模型，本实验用它来模仿整个系统，NVMain 是循环级的内存模拟器，本实验用它来模仿 PCM，从而实现 DRAM+PCM 的实验环境。实验时采用系统级仿真模式 SE，每个页面设为 4 kB 大小，延迟数据：PCM 参照 F. Bedeschi 等的研究 [16]，DRAM 参照 Micron 的测试 [17]。

具体实验数据集 [16] 测试参数见表 1。

表 1　实验数据集

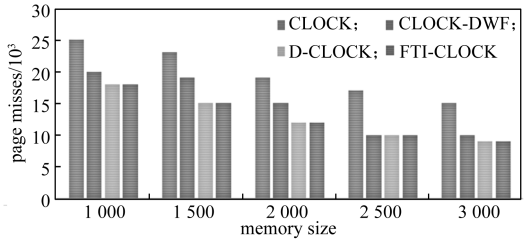
Table 1　Experimental data set

Type	Total Reference	Read/Write Ratio	Locality	Total Request
Trace9151	10 000	90%/10%	80%/20%	50 000
Trace8987	10 000	90%/10%	50%/50%	50 000
Trace3377	10 000	50%/50%	80%/20%	50 000
Trace1899	10 000	10%/90%	80%/20%	50 000
OLTP	10 000	75%/25%		472 523

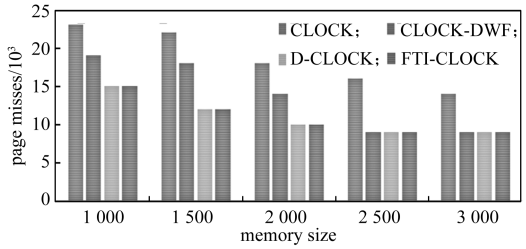
本实验数据集由两部分构成：真实数据和合成数据。真实数据采集于安徽省芜湖市某天猫网站某段时间的交易记录，数据集经过去噪处理，有 356 733 次读和 115 790 次写；合成数据通过开源软件 DiskSim 获得，通过它对磁盘的模拟读写操作来获取比例不同的局部性读写操作数据集，表 1 中的数据集中 Locality（局部性），如“80%/20%”，表示在 20% 的局部空间上发生的 80% 的读写操作。

3.2　存储空间变化的 PCM 写次数

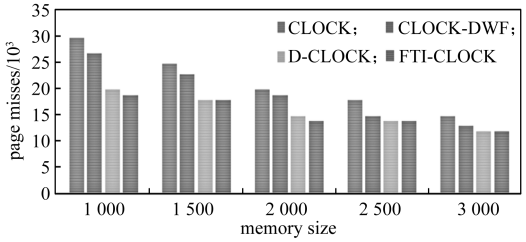
将数据集 Trace9151、Trace8987、Trace3377、Trace1899 和 OLTP 在频率时效下的 FTI-CLOCK 页面调度和 CLOCK、CLOCK-DWF 和 D-CLOCK 的页面调度进行比较。图 3 给出了 5 组数据集在 4 种不同页面调度下的 PCM 写次数统计，本次实验中内存页面逐渐增大，DRAM 和 PCM 按照 4:1 统一修改的比例进行配置。



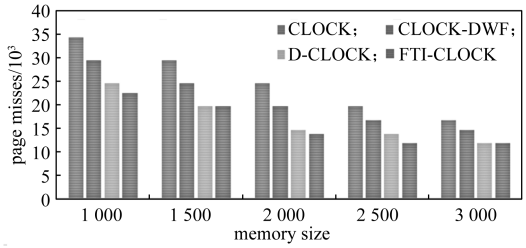
a）Trace9151



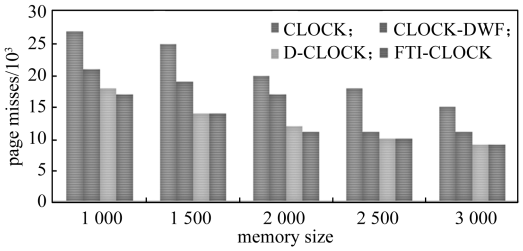
b）Trace8987



c）Trace3377



d）Trace1899



e）OLTP

图 3 存储空间变化的不同调度 PCM 写次数

Fig. 3 Different scheduling PCM write times with storage space changing

通过数据集在 4 种调度策略下的 PCM 写次数的数据显示，如图 3 中 a~e 图所示：

1）随着混合内存空间容量的增大，各数据集在 4 种调度策略的写次数都下降。实验结果显示当存储空间增大时，可以显著减少 PCM 的写次数；

2）合成数据集中读写的比例，对 PCM 写的次数影响较大，实验结果表明当写比例增大时，PCM 写的次数明显增大；

3）数据的局部操作性对 PCM 的写次数有影响，但不是很大；

4）实验结果显示，频率时效的 FTI-CLOCK 调度算法，可以有效减少 PCM 的写次数。

3.3 固定存储空间的 PCM 写次数

当存储空间固定为 2 GB，DRAM:PCM 为 4:1，实验数据集在 FTI-CLOCK 页面调度、CLOCK、CLOCK-DWF 和 D-CLOCK 情况下，PCM 写次数的实验结果如图 4 所示。

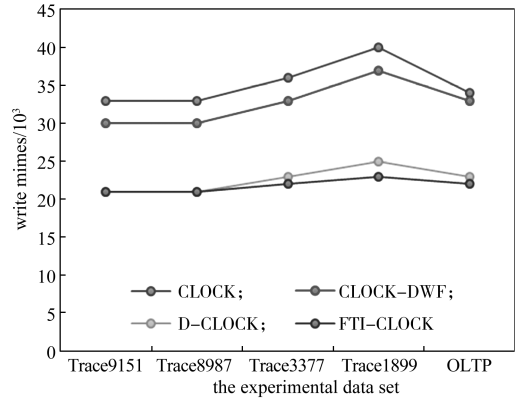


图 4 存储空间固定的不同调度 PCM 写次数

Fig. 4 Different scheduling PCM write times with fixed storage space

通过分析实验得出的数据可知，当存储空间、比例一定时：CLOCK 算法没有对混合存储空间进行区分，进行无区别的读写操作，PCM 的写次数较多；CLOCK-DWF 操作，仅根据页面的读写请求，就对页面的冷热进行划分，并将读页面置换进 PCM，页面划分较为简单，造成写 PCM 写次数还是比较高；D-CLOCK 根据当前页面的写次数和平均写次数比较来划分页面的冷热，降低了 PCM 写次数，但没有考虑页面写的局部爆发和时间间隔；FTI-CLOCK 考虑了页面的局部爆发写特点，并将局部写频率和写的时间间隔相结合，在 4 个算法的写操作中，写次数最低。实验证明，频率时效的 FTI-CLOCK 调度能够有效减少 PCM 写次数，明显地优化 PCM 写，提高 PCM 的使用时长。

4 结论

作为新一代存储材料，PCM 有着许多优点，有着较高的存储密度，并且低能耗，已经进入工程应用阶段，但如何解决 PCM 的写耐久性是一个急需解决的问题，多年来，许多研究人员给出了多种解决方案。本文通过分析内存页面写的局部性和时效性，提出了新公式将二者结合在一起，通过计算权值的形式来区分页面的冷热。

1）通过局部访问统计器对局部密集写访问的频率进行了统计；

2）将最近时间的高频写请求，上次密集写和这次访问的时间间隔统一到一个计算公式中，并根据频率和时效计算权值；

3）在考虑局部密集写访问和频率时效权值的基础上，实现写页面的调度，实验结果表明，该方法可以有效降低 PCM 的写次数；

4）本文只是从比较小的数据出发来实现频率时效的写冷热页面调度，但对大数据环境下，如何通过局部写访问和时效性来进行页面的调度，是下一步研究的方向。

参考文献：

[1] 张鸿斌，范 捷，舒继武，等 . 基于相变存储器的存储系统与技术综述 [J]. 计算机研究与发展，2014，51(8)：1647−1662.
ZHANG Hongbin，FAN Jie，SHU Jiwu，et al. Summary of Storage System and Technology Based on Phase Change Memory[J]. Journal of Computer Research and Development，2014，51(8)：1647−1662.

[2] OUKID I，LASPERAS J，NICA A，et al. FPTree：a Hybrid SCM-DRAM Persistent and Concurrent B-Tree for Storage Class Memory[C]//Proceedings of the 2016 International Conference on Management of Data. New York：ACM，2016：371−386.

[3] LEE S，LIM K ，SONG H，et al. WORT：Write

Optimal Radix Tree for Persistent Memory Storage Systems[C]// Proceedings of the 15th USENIX Conference on File and Storage Technologies(FAST 2017). Berkeley：USENIX Association，2017：257−270.

[4] WANG C D，WEI Q S，WU L K，et al. Persisting RB-Tree into NVM in a Consistency Perspective[J]. ACM Transactions on Storage，2018，14(1)：1−27.

[5] HWANG D，KIM W，WON Y，et al. Endurable Transient Inconsistency in Byte-Addressable Persistent B+-Tree[C]//Proceedings of the 16th USENIX Conference on File and Storage Technologies(FAST 2018). Berkeley：USENIX Association，2018：187−200.

[6] ZHANG T，XING J，ZHU J，et al. Exploiting Page Write Pattern for Power Mangement of Hybrid DRAM/PRAM Memory System[J]. Journal of Information Science and Engineening，2015，31(5)：1633−1646.

[7] BORTOLOTTI D，BARTOLINI A，WEIS C，et al. Hybrid Memory Architecture for Voltage Scaling in Ultra-Low Power Multi-Core Biomedical Processors[C]// Design，Automation & Test in Europe Conference & Exhibition (DATE). Dresden：IEEE，2014：1−6.

[8] SEOK H，PARK Y，PARK K W，et al. Efficient Page Caching Algorithm with Prediction and Migration for a Hybrid Main Memory[J]. ACM SIGAPP Applied Computing Review，2011，11(4)：38−48.

[9] SEOK H，PARK Y，PARK K H. Migration Based Page Caching Algorithm for a Hybrid Main Memory of DRAM and PRAM[C]//SAC’11：Proceedings of the 2011 ACM Symposium on Applied Computing. Taichung：Association for Computing Machinery，2011：595−599.

[10] LEE S，BAHN H，NOH S H. CLOCK-DWF：a Write-History-Aware Page Replacement Algorithm for Hybrid PCM and DRAM Memory Architectures[J]. IEEE Transactions on Computers，2013，63(9)：2187−2200.

[11] CHEN K M，JIN P Q，YUE L H. A Novel Page Replacement Algorithm for the Hybrid Memory Architecture Involving PCM and DRAM[C]// IFIP International Conference on Network and Parallel Computing. Ilan：Springer，2014：108−119.

[12] 刘　兵，汪令辉，张　锐，等 . 改进的偏向写调度的混合内存缓冲区调度策略 [J]. 湖南工业大学学报，2020，34(4)：48−53.
LIU Bing，WANG Linghui，ZHANG Rui，et al. An Improved Hybrid Memory Buffer Scheduling Strategy with Write Preference[J]. Journal of Hunan University of Technology，2020，34(4)：48−53.

[13] BHADKAMKAR M，GUERRA J，USECHE L，et al. BORG：Block-Re ORGanization for Self-Optimizing Storage Systems[C]// Proceedings of the 7th USENIX Conference on Fileand Storage Technologies (FAST’09). San Francisco：USENIX Associa-Tion，2009：183−196.

[14] BINKERT N，BECKMANN B，BLACK G，et al. The Gem5 Simulator[J]. ACM SIGARCH Computer Architecture News，2011，39(2)：1−7.

[15] POREMBA M，ZHANG T，XIE Y. NVMain 2.0：a User-Friendly Memory Simulator to Model (Non-)Volatile Memory Systems[J]. IEEE Computer Architecture Letters，2015，14(2)：140−143.

[16] BEDESCHI F，RESTA C，KHOURI O，et al. An 8Mb Demonstrator for High-Density 1.8V Phase-Change Memories[C]//2004 Symposium on VLSI Circuits. Honolulu：IEEE，2004：442−445.

[17] Micron Technology. Micron 8GB：X4, X8, X16 DDR3l SDram Description[EB/OL]. [2019−11−08]. https:// www.micron.com/ products/Dram/ddr3-sdram.

（责任编辑：申　剑）