PCRAM-aware cluster allocation algorithm for hybrid main memory hierarchy

Shunfen Li, Xiaogang Chen, Mi Zhou, Gezi Li, Yiyun Zhang, and Zhitang Song

State Key Laboratory of Functional Materials for Informatics and Nanotechnology Laboratory, Shanghai Institute of Micro-system and Information Technology, Chinese Academy of Sciences, 865 Changning Road, Shanghai 200050, China

Abstract: Phase Change Random Access Memory (PCRAM) has several characteristics that make it a promising candidate for main memory. In this paper, the challenges involved in incorporating PCRAM into the main memory hierarchy of flash-based systems are explored, and a hybrid memory hierarchy consisting of PCRAM and DRAM for SSD is proposed. To reduce the write traffic of Flash, a buffer management scheme for the hybrid memory is implemented. The proposed scheme above is implemented on basis of Disksim platform. The simulation results indicate that the proposed buffer algorithm can decrease the number of random write and erase operations of the SSD and improve the average system response rate obviously.

Keywords: PCRAM, hybrid main memory hierarchy, buffer management, cluster

Classification: Storage technology

References

[1] Y. Xie: IEEE Des. Test Comput. 28 (2011) 44. DOI:10.1109/MDT.2011.20
[2] B. C. Lee, E. Ipek, O. Mutlu and D. Burger: Proc ISCA ’09 (2009) 2.
[3] R. Freitas and W. Wilcke: IBM J. Res. Develop. 52 (2008) 439. DOI:10.1147/rd.524.0439
[4] B. Lee, P. Zhou, J. Yang, Y. Zhang, B. Zhao, E. Ipek, O. Mutlu and D. Burger: IEEE Micro 30 (2010) 130. DOI:10.1109/MM.2010.24
[5] G. Sun, Y. Joo, Y. Chen, D. Niu, Y. Xie, Y. Chen and H. Li: High Performance Computer Architecture (2010) 1. DOI:10.1109/HPCA.2010.5416650
[6] G. Dhiman, R. Ayoub and T. Rosing: Proc. of Design Automation Conference (2009) 464.
[7] S. Eilert, M. Leinwander and G. Crisenza: IEEE International Memory Workshop (2009) 168. DOI:10.1109/IMW.2009.590604
[8] K. Sarpatwari, A. Yanamandra, V. Narayanan, Y. Xie, M. J. Irwin and O. A. Karim: Proc. GLSVLSI ’08 (2008) 395.
[9] G. Ganger, B. Worthington and Y. Patt: The DiskSim Simulation Environment Version 3.0 Reference Manual: Technical Report CMU-CS-03-102 5.
1 Introduction

Traditional memory hierarchy design consists of embedded memory for on-chip caches, commodity DRAM for main memory, and NAND flash memory for storage. As the increasing size of the memory system, the traditional DRAM-based main memory systems are starting to hit the cost and power limit [1, 2, 3]. In addition, flash memory has characteristics of not-in-place update, random write, and asymmetric I/O costs among read, write, and erase operations [4, 5, 6]. Hence, more effort should be taken to overcome the drawbacks to improve access performance of flash-based systems. Emerging memory technologies such as PCRAM, combines the density of DRAM, and the nonvolatility of flash memory, and so become attractive as alternatives for the future memory hierarchies [7, 8].

In this paper, to improve the SSD performance in terms of capacity, power, and cost, a novel energy efficient hybrid main memory hierarchy based on PCRAM and DRAM for the SSD is proposed. To overcome above drawbacks of flash-based systems, a buffer management scheme which is called P-ACAA (PCRAM-Aware Cluster Allocation Algorithm) for hybrid main memory and flash-based system is implemented. The above scheme is implemented on basis of Disksim platform. To study the best performance of P-ACAA policy in different application, simulations are performance with different types of traces. The simulation results show that in the types of Write-intensive application, the P-ACAA policy performs much better. To further prove excellent performs of the above policy in the types of Write-intensive application, simulation experiments are presented with the same types of traces using different count. The simulation results indicate that P-ACAA can decrease the random write and erase operation numbers of the SSD, and improve the average system response rate obviously.

2 Hybrid main memory hierarchy for SSD and the P-ACAA policy

The paper explores the challenges involved in incorporating PCRAM into the main memory hierarchy for the SSD, and a hybrid hierarchy modeling for memory is built with PCRAM and DRAM, as Fig. 1(a) shows. The new main memory hierarchy for the SSD is accessed by using the SSD controller. To get the best capacity and latency, PCRAM is used as composition of main memory capacity. In the hybrid main memory organization, the larger PCRAM storage has the capacity to hold most of the pages needed during program execution, thereby reducing the Mass Storage accesses due to paging. When performing the read operation, the SSD controller can directly read data from PCRAM. In the new hybrid hierarchy model, the PCRAM is managed by the Operating System using a Page Table, which is used in a manner similar to current DRAM main memory systems.
An efficient buffer replacement algorithm, which is called P-ACAA, is presented for the system with DRAM/PCRAM hybrid main memory and flash memory storage devices. The purpose of P-ACAA is to improve the overall I/O performance by reducing the random write counts of flash incurred in the write process. PCRAM acts as a buffer for main memory between the Flash memory and the DRAM. The pages that belong to the same block of Flash memory are grouped together forming a cluster node maintained in the PCRAM, and the clusters are maintained in LRU order. The cluster is moved to the beginning of the list when a page in the cluster is updated, or a new page is added to the cluster.

When P-ACAA receives a write request, it searches the cluster node list first. On a hit, the data is overwritten in the cluster. Otherwise, a new page is allocated from PCRAM and attached to the page list of the corresponding cluster node, and the requested sector is written to the page. If the cluster node does not exist, a new cluster node allocated from PCRAM should be added at the head of the cluster node list. Since all write requests are buffered, the actual data write on flash memory is occurred only when a block is replaced. When P-ACAA receives a read request, it searches the cluster node list first. On a hit, P-ACAA returns the data from the PCRAM to CPU. If P-ACAA can’t find the sector in the cluster node, it reads the data from flash storage to CPU. When the page number of cluster node arrives a certain number, the cluster node is written back to the Flash memory, and all the pages belonging to this cluster node are evicted from the PCRAM.

To realize the P-ACAA algorithm, a data structure is contrived as shown in Fig. 1(b). Horizontally, the cluster node list is a linked list of blocks sorted by the used frequency. Vertically, the page node list is a list of pages belonging to the corresponding block. A cluster node contains a block number, a node counter, a cluster node list pointer, and a page list pointer. The block number identifies a unique block in flash memory. The cluster node list pointer is for the horizontal linked list, in which all the cluster nodes are sorted by recency used for the block-level LRU policy. The node counter denotes the number of pages allocated in this cluster node. The page list pointer is point to the list of page nodes that belong to the cluster. Page node list has the page data associated with the page number. The overhead of data structure of each cluster node is a disadvantage of P-ACAA algorithm compare to other method. The data structure of cluster nodes will consume the limited DRAM resource, and the management of data structure will
increase the system overhead. Even if there is some overhead of the scheme, the P-ACAA policy performance is better than other policy in the type of Write-intensive application which can decrease the random write and erase operations of SSD. Using the overhead to get better performance is acceptable.

3 Simulation and evaluation

The performance of P-ACAA policy is evaluated on the basis of DsikSim simulate platform. The simulator knows the parameters related to current technologies as exactly as possible [9]. The traces are extracted from disk access logs of real user activities on FAT32 file system. The page-level mapping scheme and a 1-GB NAND Flash memory with 128 KB pages in each block are assumed for simulation.

To assess the performance of P-ACAA, LRU and CLOCK-DWF [10] policy, the write throughput and the required number of block erase operations are simulated with the buffer size from 1 MB to 6 MB. The LRU and CLOCK-DWF policy only uses DRAM as buffer while P-ACAA policy uses 1 MB DRAM and different sizes (1 MB to 5 MB) of PCRAM as buffer. To study the efficiency of different buffer management schemes, the simulation experiments are performed with a set of real traces. Fig. 2 shows the simulation result of the write numbers and erase counts based on the same set of real traces with the different size of buffer.

It can be seen from Fig. 2 that the performance of P-ACAA policy is best in terms of reducing the number of write and erase operations from the point that using the buffer size of 4 MB. As the advantage is not obvious with the buffer size of 4 MB, the buffer size of 5 MB in the later simulation is used to compare the performance.

The paper researches the random access performance of PCRAM as main memory. In the paper, the DRAM is used to maintain the key information of the scheme, while the PCRAM is used to maintain the pages which are belonged to the clusters. The specification of DRAM, FLASH and PCRAM configuration in the simulation are shown in Table I. As the write and read latency is larger than the DRAM, so the 1M PCRAM performance is no better than 1M DRAM. The paper is mainly research the performance of PCRAM as main memory, and using PCRAM to reduce the capacity of DRAM and improve the system performance of SSD. And the paper focuses on exploiting P-ACAA policy performance under the Write-intensive application which can decrease the random write and erase operations of SSD. Using the overhead to get better performance is acceptable.
intensive application. With the capacity of PCRAM increasing, the number of the pages in the cluster that maintained in PCRAM is more, and the buffer performance of the PCRAM is better in terms of reducing the number of write and erase operations, as shown in Fig. 2.

Table 1. (a) The specification of DRAM configuration. (b) The specification of FLASH configuration. (c) The specification of PCRAM configuration.

| Parameters       | Value          | Parameters       | Value          | Parameters       | Value          |
|------------------|----------------|------------------|----------------|------------------|----------------|
| Read latency     | 50 ns          | Read latency     | 20 μs          | Data Writing width| 8 bits        |
| Write latency    | 30 ns          | Write latency    | 100 μs         | Data reading width| 8 bits        |
| Write Bandwidth  | 1 GB/s         | Block erase latency | 1.5 ms        | Write latency    | 1 μs          |
|                  |                | Write Bandwidth  | 80 MB/s/die    | Read latency     | 60 ns         |
|                  |                |                  |                | Write Bandwidth  | 80 MB/s/die    |

Fig. 3. (a) and (b) is the comparison of write and erase operation number under the same count of different type of traces. (c) and (d) is the comparison of the write and erase operations number in the case of Write-intensive application with different count of traces.

To study the best performance of P-ACAA policy in different application, simulation experiments are performance with different types of traces that are classified into four types according to the ration of write operations. The types are expressed as 20/100, 50/100, 80/100 and 100/100 meaning that the proportions of the write operations in the trace are 20%, 50%, 80% and 100% respectively. With these four classifications, one million traces are generated which have read/write operations to simulation the performance of different buffer policy. In the case of 100/100 trace, as shown in Fig. 3(a) and (b), the proposed P-ACAA policy performs much better than other schemes in terms of reducing both the write and erase operation numbers of Flash. Thus, P-ACAA policy performs best in the types of Write-intensive application. To further prove excellent performs of the P-ACAA policy in Write-intensive application, simulation experiments are presented with the same types of traces, the count of which are 600000, 800000 and 1000000.
respectively. As can be seen from Fig. 3(c) and (d), the simulation results indicate that the P-ACAA policy with different number of traces performs best in terms of reducing the write and erase operation numbers of Flash. Compared to the other schemes, the P-ACAA policy reduces the number of writes and erases. This performance gains can be explained by the fact that evicting data first from the buffer helps retaining those data that are frequently accessed.

Another important advantage of the P-ACAA policy is that the policy can improve the average system response rate obviously, as shown in Fig. 4. The P-ACAA policy completes simulation work faster than other policies. The probability is that the cumulative time distribution probability of the different policy completes all write requests. Since all write requests are buffered, the actual data write on flash memory is occurred only when a block is replaced. The above policy reduces the random write of flash memory to improve the average response rate of the SSD.

![Fig. 4. The comparison of P-ACAA with LRU, CLOCK-DWF and NO-CACHE in terms of average response rate of system in the case of Write-intensive application using the count of trace is 10000.](image)

4 Conclusion

As a promising non-volatile memory technology, PCRAM is expected to play an important role in the memory hierarchy in the near future. This paper focuses on exploiting PCRAM as main memory for flash-based systems. The challenges involved in incorporating PCRAM into the main memory hierarchy of flash-based SSD are explored, and a hybrid main memory hierarchy consisting of PCRAM storage coupled with a small DRAM buffer is proposed. The hierarchy has both the latency benefits of DRAM and the capacity benefits of PCRAM, which can improve the main memory performance of SSD. A buffer management scheme, which is called P-ACAA for hybrid main memory hierarchy, is implemented improving the overall I/O performance of SSD. The above scheme is implemented on basis of Disksim platform. The simulation results indicate that the P-ACAA policy performs best in the types of Write-intensive application which can both decrease the number of random write and erase operations of SSD and improve the average system response rate obviously.

Acknowledgments

Supported by National Key Basic Research Program of China (2011CB932804, 2013CBA01900, 2011CBA00607, 2010CB934300).