Middle-East Journal of Scientific Research 17 (6): 827-831, 2013
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2013.17.06.11997

Proposed Hybrid Memory Using Dram and Pcm to Attain Better Performance

M. Prabhu, S. Rajarajan and K.S. Suresh

School Of Computing,
SASTRA University, Thanjavur, Tamilnadu, India

Abstract: Computer system architecture is a combination of processor, different levels of memory and peripheral devices. The memory in a computer determines the stability and efficiency. The different layers of the memory hierarchy can be distinguished by the response time and potentiality to withstand the data. The upper layer of the hierarchy is the main memory, which is expected to be fast, non volatile and should able to withstand higher number of write cycles. DRAMS are being used currently, which fails in the case of non volatility. Hence the research on non-volatile memories is at full swing. The introduction of PCM (phase change memory) is considered to be a major achievement in non volatile memory. But the major drawback with PCM is the lesser life time or ability to withstand minimal number of write cycles, longer write time and high power utilization during writes. This paper proposes an architectural framework to exploit the best out of both memories, in which PCM will share the load of the DRAM. This proposed architecture will deliver the expected main memory layer with higher speed, non volatile nature and higher lifetime.

Key words: PCM · Hybrid main memory · PCM endurance · Non volatile main memory · Wear levelling

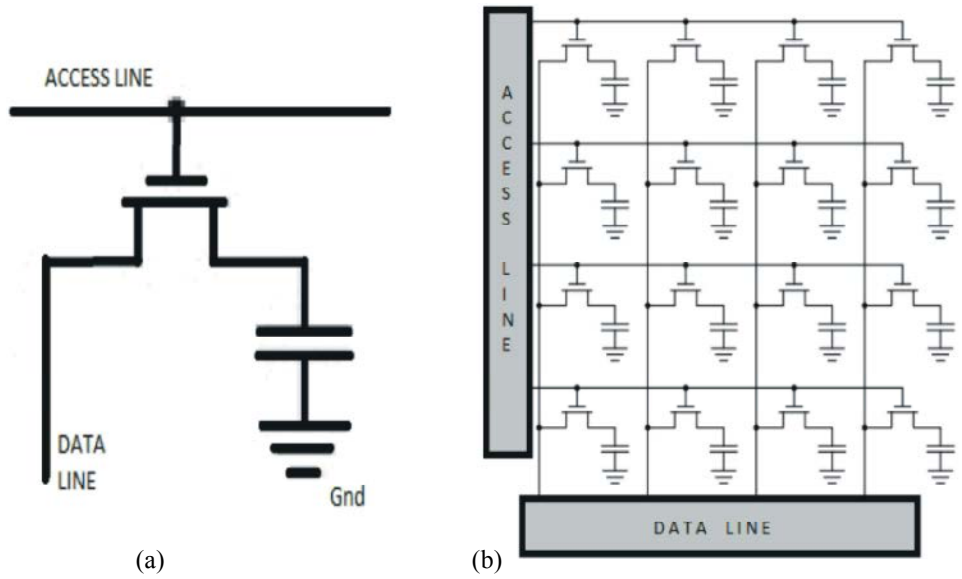
INTRODUCTION

The role of the main memory is played by the DRAM for almost 30 years. Before proceeding with the proposed architecture, we should discuss about the existing memory architecture with its DRAM main memory and the new type of memory which is to be integrated with existing. In the critical memory hierarchy of computer system, each upper layer can be considered as the buffer to the lower ones. In the hierarchical order main memory is considered to be very important. This is due to the reason that it is the memory which makes the contact with the processor in all of the execution. Any alterations in the existing order may lead to a greater impact and overheads. An ideal main memory should be faster, cheaper, persistent and bigger. It should possess higher endurance. There are two options to meet the requirements of an ideal memory. One is to make the DRAM as the ideal one and the other is to implement a new memory hierarchy using the available non volatile memories. The former failed due to volatile nature of DRAM. SRAM and flash memory which fall into the latter failed due to their low speed and block erase nature. Most of the NVM technology memories are at prototype level. Phase Change Memory is one among them with some promising qualities. Despite of its advantages, it also has limitation that includes lengthy writing time and lower endurance. This paper proposes an architecture combining PCM as part of the existing main memory and a strategy to solve the major overheads.

This paper is organized as follows. Section 2 covers the characteristics of both the memories and their comparison. Section 3 covers the research work done on both DRAM and PCM. Section 4 explains the proposed architecture and the strategy adopted. Section 5 and 6 covers the outcome of the proposed framework and future paths which lead to the betterment of this architecture.

DRAM and PCM: The technology which very much fit into the description today for the main memory is Dynamic Random Access Memory (DRAM). A DRAM memory cell consists of a transistor which provides access to states and a capacitor to hold the state as shown in Figure 1(a). The charged and discharged capacitor determines the value in the memory cell. The data line and access line of the memory cell is used to read and write data to it. To read the data in the cell,

Corresponding Author: M. Prabhu, School Of Computing, SASTRA University, Thanjavur, Tamilnadu, India.



(a) (b)

Fig. 1:

the access line is raised which causes current to flow in data line. To write data line is set and access line is raised for the time long enough to charge the capacitor. The arrangement of memory cells into rows and columns are shown in Figure 1(b). The major advantage of the DRAM memory cell is the simplicity to access, higher density of cells and faster read, write operations without any erase mechanism. Despite of its advantages, DRAM possesses lot of disadvantages. The size of the memory cell determines the density of DRAM. Therefore to get DRAM of higher density, the capacitor size should be minimal, which result in frequent current leakage and increased memory refresh. A memory read operation does also require memory refresh. Frequency in refreshing makes a DRAM consumes more energy. The DRAM also suffers from "soft errors" and random bit corruption caused by cosmic radiation.

PCM on the other hand is a non volatile memory which can compete with DRAM of today in most of the aspects. The technique to store the data in PCM is totally different from the traditional approach.PCM is based on the Chalcogenide alloy (Ge2Sb2Te5) which can exhibit two physical states, Crystalline and amorphous. The crystalline state of the PCM conducts electricity with lower resistance and the amorphous state does not. The variation in the resistance between the two states makes the PCM to use as memory device. The crystalline state which conducts electricity is treated as logic 1 state and amorphous as logic 0. A PCM write operation unlike

Table 1: Comparison of PCM and DRAM

ATTRIBUTES	PCM	DRAM
Non-Volatile	Yes	No
Cell area	$6F^2$	$4F^2$
Granularity	Small/Byte	Small/Byte
Read Time	<10ns	12ns
Write/Erase Time	<10ns	100ns
Retention Time	64ms	>10y
Endurance	10^6	Unlimited
Write Energy (J/bit)	4E-15	6E-12
Standby Power (W/GB)	1E-1	1E-3

a DRAM requires application of heat to the memory cell at the level of 600°C. The cooling time determines the physical state or data retained in the memory cell. When quenched in ~10ns, it creates amorphous or logic 0 memory cell and if it is in ~100ns, it makes a crystalline state. The attributes of the PCM are similar in most of the way with the DRAM that includes bit alterability and granularity. The major advantage of a PCM over the DRAM is the non volatile nature. The comparison of PCM with the DRAM is shown in the table 1 [1]. The major drawbacks with the PCM are its endurance and write bandwidth.

Although PCM try to compete with DRAM in the read latency, the write operation makes it a slower device. The slower erase mechanism of PCM reduces its write bandwidth. The endurance of PCM is 10^6, which is very low for a main memory alternative. The major advantages of PCM over the DRAM are its non volatile nature and retention time.

Works Done So Far to Transform Dram and Pcm as Optimal Memory: The comparison of PCM with the main memory clearly states the advantages and limitations of both. Most of the research done so far covers only the endurance issue of the PCM, which is supposed to solve either by making modifications to the hardware or by adopting suitable software strategy. In [2], hybrid architecture with proposed CLOCK-DWF algorithm which predicts the future writes and thereby it makes most of the writes to DRAM. Another approach to normalize the effect of long write latency and increase endurance of PCM is the use of DRAM cache for PCM main memory [3]. But there is a problem with architecture implementing DRAM cache for PCM main memory. Those architectures normally suffer from long latency of PCM writes especially in case of larger programs. Most of the approaches to increase endurance involved Wear-levelling and Write-Reducing. In [4] implementation of Bloom filters to manage the write count in PCM is proposed. Write-Activity-Aware Page Table Management for wear levelling of PCM is adopted in [5]. FREE-PAGE-ALLOCATION() and WORN-OUT-AVOIDANCE() are the algorithms proposed to achieve wear levelling in PCM, by returning a younger page for the page write request and if the older page need frequent update, the contents of the page are moved to the younger page [6]. Cache address remapping and randomized address remapping are also implemented in DRAM caches and PCM to minimize the number of PCM writes and wear-levelling [7].

A. Motivation: The hybrid architecture proposed so far covers PCM as main memory with DRAM cache or combined PCM-DRAM main memory approach focused on wear levelling and minimizing writes. Minimizing the write hits helps the betterment of the architecture in several ways. In a PCM write, erasing the memory cell consumes much of the time and energy. So any approach made to minimize the write in a PCM increases its endurance, speed and minimize the power consumption. Most of the hybrid memory architecture approaches focus on minimizing writes in PCM and thus extending its endurance. [8] None of the past architectures and approaches focus on the maximizing read operation in PCM. The proposed architecture focus on minimizing the PCM write but not the read, so that non-volatility of PCM can be utilized to a greater extend with higher endurance and power consumption. Moreover, the focus towards keeping most required data in PCM will make efficient utilization of PCM and DRAM in the proposed architecture.

Proposed Architecture: The kernel space is mainly occupied by interrupt handling routines, process scheduling information, memory management routines and hardware drivers. The routines and drivers specified will not undergo much of the modification and are used at most during boot up in a computer. This quality of the kernel space is highly suitable for adopting it to the proposed architecture. The proposed architecture covers the idea for minimizing the writes in the PCM based on the category of page written to it. [9] It uses both PCM and DRAM as the main memory. In main memory, data are arranged in the form of pages. The idea is to keep these constantly used kernel pages in the PCM part of the main memory. The space of the DRAM main memory is occupied with the pages of user applications which undergo frequent update. The proposed architecture with PCM-DRAM main memory is shown in the figure 2.

The approach stated will result in a part main memory with kernel information which is non volatile. The user application pages are supported by the DRAM as usual. The figure 3 shows the detailed architecture of memory controller which support PCM and DRAM. The page walk mechanism determines the availability of a page in PCM. Upon unavailability, the requested kernel page is obtained through the DRAM from the disk.

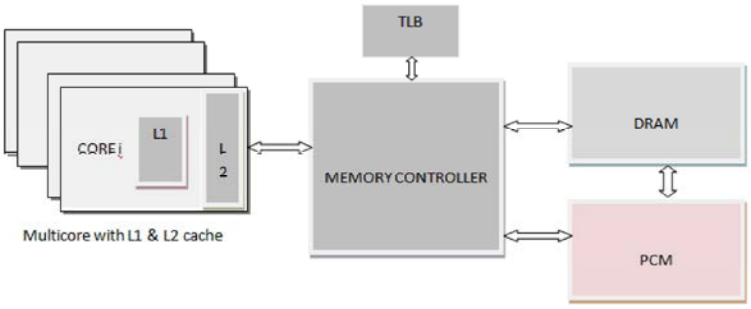


Fig. 2: Proposed hybrid architecture with memory controller.

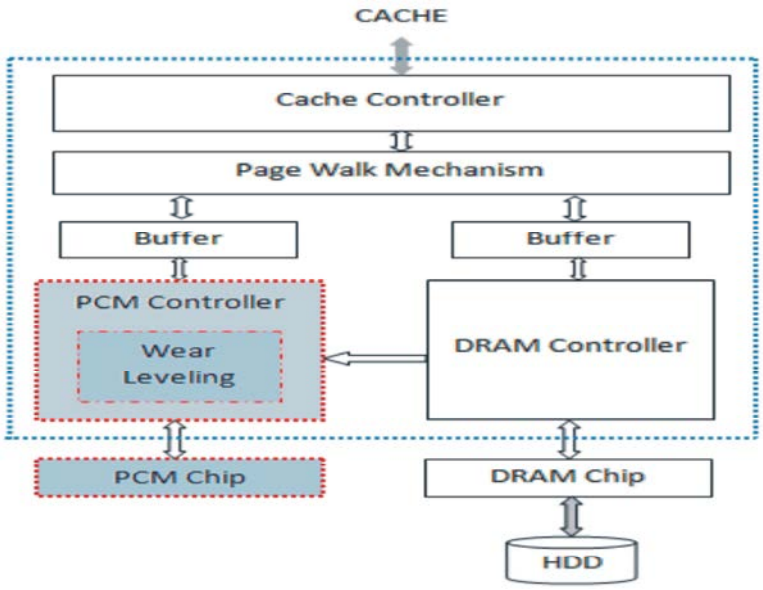


Fig. 3: Memory Controller supporting DRAM and PCM

Now, the problem with this approach is uneven wear out. The static kernel pages in the PCM occupy the fixed memory cells. As the result of this, the pages which undergo change will wear our earlier compared to the static page.

Wear Levelling: This is solved by including the wear levelling technique discussed in [6]. Two approaches has been proposed which include bucket wear levelling and age based wear levelling. Since there is less latency in Bucket wear levelling it is most suited for the proposed architecture. Bucket based wear levelling includes two algorithms, FREE-PAGE-ALLOCATION () and WORN-OUT-AVOIDANCE (). Two buckets are maintained, one with free pages and other with pages currently being used. Both buckets hold younger pages. This is tracked by page write count of each page in accordance with threshold count R. Each time, upon request, FREE-PAGE-ALLOCATION () algorithm returns a free page from free list bucket, if available. In contrast, the last younger page from in-use list bucket is chosen. The cold data of younger page is moved to the older free frame in the older free frame list. The reference count of young page is increased by one.

WORN-OUT-AVOIDANCE () algorithm on the other hand levels the wear out in the PCM. When a page q needs update, the following conditions may occur. When q is not in oldest in-use page list, the update is done and writes count of the page c[q] is incremented by one. When q has been written already for R times, then the page is moved to next bucket with older in-use pages and c[q] is reset as zero. When q is already in bucket with older pages, based on $0 < c[q] < R/2$ two cases were delineated. If c[q]=0, then the pages has been moved from previous level bucket due to excessive writes and if c[q]!=0, then it suggest that the page has been moved by the FREE-PAGE-ALLOCATION() algorithm. A old paged page when fetched by FREE-PAGE-ALLOCATION () algorithm must have cold data. [10] Hence q undergo instant update when it solve $0 < c[q] < R/2$. On the other hand when $c[q] = 0$ or $c[q] >= R/2$ young page is chosen to avoid q attaining wear out level.

Advantages in Proposed Architecture: In the architecture that has DRAM main memory, execution of an application requires code movement from secondary memory to main memory. This results in higher response time for an application. The proposed architecture saves the kernel pages in PCM main memory. Since the PCM is loaded mostly with kernel pages, the number of writes made to the PCM memory cell is minimized when compared in having user application pages. Due to the persistence of kernel pages in PCM, architecture that support short boot time and response time can be obtained. The strategy of adopting lower number writes also results in higher endurance and lower power consumption. The inclusion of wear levelling technique increases the endurance of PCM further.

CONCLUSION

Implementation of PCM-DRAM architecture can solve the problem of volatile main memory to a greater extend. But the sophistication of PCM can be enjoyed by solving the functional overheads in implementing the PCM to existing DRAM main memory architecture. This paper covers a proposed architecture which focuses towards the pages kept in DRAM and PCM. As the time and power consumption of PCM is high during write operation, the idea to keep kernel pages in PCM will reduce the both compared with keeping the user application pages in it. The proposed architecture will result to system with minimal boot time. Future work will cover the approaches to include user application pages in PCM thereby increasing the endurance further.

REFERENCE

1. Sang Bum Kim and H. Chung, 2012. Lam, Transition of Memory Technologies, VLSI Technology, Systems and Applications VLSI-TSA, 2012 International Symposium, 23-25 April 2012, pp: 1-3.
2. Soyoon Lee, Hyokyung Bahn and H. Sam Noh, 2011. Characterizing Memory Write References for Efficient Management of Hybrid PCM and DRAM Memory, 19th Annual IEEE International Symposium on Modelling, Analysis and Simulation of Computer and Telecommunication Systems, pp: 168-175.
3. Meza, J. Jichuan Chang, Han Bin Yoon, O. Mutlu and P. Ranganathan, 2012. Enabling Efficient and Scalable Hybrid Memories Using Fine-Granularity DRAM Cache Management, IEEE Computer Architecture Letters, 11(2): 61-64.
4. Joosung Yun, Sunggu Lee and Sungjoo Yoo, 2012. Bloom Filter-based Dynamic Wear Leveling for Phase-Change RAM, Design, Automation & Test in Europe Conference & Exhibition, 2012, 12-16 March 2012, pp: 1513-1518.
5. Tianzheng Wang, Duo Liu, Zili Shao and Chengmo Yang, 2012. Write-activity-aware page table management for PCM-based embedded systems, Design Automation Conference (ASP-DAC), 2012 17th Asia and South Pacific, pp: 317-322.
6. Chi-Hao Chen, Pi-Cheng Hsiu, Tei-Wei Kuo, Chia-Lin Yang and C.Y.M. Wang, 2012. Age-based PCM wear leveling with nearly zero search cost, Design Automation Conference DAC, 2012 49th ACM/EDAC/IEEE, pp: 453-458.
7. Gang Wu, Huxing Zhang, YaoZu Dong and Jingtong Hu, 2012. CAR: Securing PCM Main Memory System with Cache Address Remapping, Parallel and Distributed Systems (ICPADS), 2012 IEEE 18th International Conference, pp: 628-635.
8. Kabiru Jinjiri Ringim, 2013. Understanding of Account Holder in Conventional Bank Toward Islamic Banking Products, Middle-East Journal of Scientific Research, 15(2): 176-183.
9. Muhammad Azam, Sallahuddin Hassan and Khairuzzaman, 2013. Corruption, Workers Remittances, Fdi and Economic Growth in Five South and South East Asian Countries. A Panel Data Approach Middle-East Journal of Scientific Research, 15(2): 184-190.
10. Sibghatullah Nasir, 2013. Microfinance in India Contemporary Issues and Challenges. Middle-East Journal of Scientific Research, 15(2): 191-199.