Energy Efficient Buffer Cache Replacement

Jianhui Yue
Univ. of Maine, Orono
jyue@eece.maine.edu

Yifeng Zhu
Univ. of Maine, Orono
zhu@eece.maine.edu

Zhao Cai
Univ. of Maine, Orono
zcai@eece.maine.edu

Abstract

Power consumption is an increasingly impressing concern for data servers as it directly affects running costs and system reliability. Prior studies have shown most memory space on data servers are used for buffer caching and thus cache replacement becomes critical. This paper investigates the tradeoff between these two interacting factors and proposes three energy-aware buffer cache replacement algorithm. On a cache miss for a new block b in a file f, it evicts an victim block from the most recently accessed memory chip. Simulation results based real-world TPC-R I/O trace show that our algorithm can save up to 12.2% energy with marginal degradation in hit rates.

1. Introduction

In order to bridge the ever-widening gap between disk and processor speeds, high-end storage servers are often equipped with large capacity memory. For example, the IBM Bluegene at LLNL has 32 Tera-bytes [7] and up to 2 Tera-byte can be installed on a single server [6]. Many previous studies [4,6] have shown that memory is one of major sources of power consumption. As the memory capacity continues to increase rapidly to alleviate the I/O bottleneck, memory energy efficiency becomes a pressing concern.

Buffer cache replacement schemes play an important role in conserving memory energy, since buffer cache is frequently more than 77% of total available memory on desktop computers and much more on storage servers [5]. Specifically, memory energy is impacted from two aspects: (1) Replacement algorithms with high hit rates help reduce the total amount of memory traffic and the overall running time; (2) Data layout generated by replacement algorithms determines the access sequence and utilization of memory chips, and hence influences the opportunities for each chip to enter lower power modes and also to deploy "free ride" DAM overlapping. Conventional replacement algorithms only aim to optimize cache hit rates and do not consider the power status of memory chips. Accordingly energy saved due to higher cache hit rates may not make up for extra energy cost incurred by naughty data layouts across memory chips, resulting in inferior energy efficiency eventually. This motivates us to study new cache replacement algorithms that optimize the tradeoff between cache hit rates and energy-efficiency.

In this paper, we propose an energy-aware buffer cache replacement algorithm that achieve energy saving at little sacrifice to cache hit rates. The key idea of our algorithms is that we consider the power status of memory chips when making replacement decisions. Instead of evicting out the block that is most likely to be access in the farthest future, we discard the one that is suboptimal in terms of hit rates but potentially has large energy salvage.

The rest of this paper is organized as follows. Section 2 briefly describes the background of power-aware memory chips and DMA overlapping. Section 3 presents our energy-efficient buffer cache management algorithm. Section 4 gives our simulation results. Section 5 concludes the paper.

2. Background

In the RDRAM technology [1], each memory chip can be independently set to one of four states: active, nap, standby and powerdown, in decreasing order of power consumption. A chip must be in the active state to perform reading or writing. In the other three states, the chip powers off different components to conserve energy and thus can not access data before switching back to the active state. The transition from a lower power state to a higher one requires some synchronization time delay.

On a storage server, recent DMA controllers, such as Intel's chipset E8870 and E7500 [2], allow multiple DMA transfers on different buses to access the same memory module simultaneously in a time multiplexing fashion. Typically, the peak transfer rate of a memory chip can be a multiple factor of the bandwidth of the PCI bus. Multiplexing various slow disk and network I/Os to the same memory chip can reduce the waste of active memory cycles and hence save memory energy. [6]

Table 1. TPC_R Simulation Results

Cache Size(MB)	64	160	256	352	448
HitRate Reduction(%)	0.1	0.37	0.11	0.29	0.48
Energy Reduction (%)	12.2	2.21	3.6	2.13	5.96

3. Energy Efficient Buffer Cache Replacement Algorithms

Our previous study [8] shows that, among eight conventional cache replacement algorithms studied, 2Q [3] has the best memory energy efficiency in most cases. Our experiments have shown that 2Q has a stronger capability of clustering hot blocks into a small set of memory chips than the other algorithms, which significantly increases the energy saving opportunities through DAM overlapping and power state transition. However, all these algorithm are essentially not energy aware since their goal is only to optimize cache hit rates and do not consider the power status of memory chips during cache replacement. In this paper, we take 2Q as an example algorithm to investigate how to judiciously take advantage of the memory technology to save energy.

When a data miss occurs, this algorithm chooses a victim block from the most recently used (MRU) chip. It trades cache hit rates for potential energy saving. Specifically, instead of replacing the block that is the least likely to be accessed in the future, this algorithm replaces one that is not likely to be accessed very soon and resides in a chip that is potentially in an active state. This algorithm predicts the most recently used memory chip is still in the active state and thus can serve current request without powering up overhead. More importantly, by concentrating memory accesses on the last accessed chip, it creates more opportunities for DMA overlapping on this memory chip and also provides more chances for other chips to enter power saving states. We name this algorithm as *chip_2Q*.

4. PERFORMANCE AND ENERGY Evaluation

We have developed a detailed trace-driven simulator that faithfully emulates network DMA and disk DMA operations and accurately records energy consumed by each memory chip to evaluate our energy aware replacement algorithm.Under the TPC-R workload, we conclude that the proposed algorithm can achieve significant energy savings with very marginal scarification to the performance in terms of hit rates. As shown in Table. 1, the maximum energy saving over *2Q* can be as much as 12.2%. However, When compare with the average hit rate of *2Q* across all experiment cache sizes, the average hit rate of *chip_2Q* is only degraded by 0.27% (see Tab. 1).

5. Conclusion

In this paper, we have proposed a generic power aware strategy. On a cache miss for a block b of a file f, it chooses a victim block from the most recently accessed memory chip achieving tradeoff between cache performance and memory energy efficiently. We use one real-world I/O server traces, TPC-R, to examine the performance characteristics and energy implications of our strategity. Experimental results show that our power aware strategies can save up to 12.2% more energy than 2Q with only a marginal scarification of cache hit rates up to 0.48%.

Acknowledgements

References

[1] R. Inc. Rambus memory chips. http://www.rambus.com.

[2] Intel. Server and workstation chipsets. http://www.intel.com/products/server/chipsets/.

[3] T. Johnson and D. Shasha. 2Q: A low overhead high performance buffer management replacement algorithm. In *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB)*, pages 439–450, San Francisco, CA, USA, 1994. Morgan Kaufmann Publishers Inc.

[4] A. R. Lebeck, X. Fan, H. Zeng, and C. Ellis. Power aware page allocation. In *ASPLOS-IX: Proceedings of the ninth international conference on Architectural support for programming languages and operating systems*, pages 105–116, New York, NY, USA, 2000. ACM Press.

[5] M. Lee, E. Seo, J. Lee, , and J. Kim. Pabc: Power-aware buffer cache management for low power consumption. *IEEE Transactions on Computers*, 56(4), 2007.

[6] V. Pandey, W. Jiang, Y. Zhou, and R. Bianchini. DMA-aware memory energy management for data servers. In *The Proceedings of the 10th International Symposium on High-Performance Computer Architecture (HPCA'06)*, 2006.

[7] M. E. Tolentino, J. Turner, and K. W. Cameron. An implementation of page allocation shaping for energy efficiency. In *Proceedings of 3rd Workshop on High-Performance, Power-Aware Computing*, April 2007.

[8] J. Yue, Y. Zhu, and Z. Cai. Evaluating memory energy efficiency in parallel i/o workloads. In *Proceedings of 2007 IEEE International Conference on Cluster Computing (CLUSTER)*, pages 21 – 30, Austin, TX, USA, Sept. 2007.