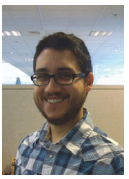
Row Buffer Locality Aware Caching Policies for Hybrid Memories


HanBin Yoon


Justin Meza


Rachata Ausavarungnirun


Rachael A. Harding


Onur Mutlu

Phase change memory (PCM) is a promising technology that can offer higher capacity than DRAM. Unfortunately, PCM's access latency and energy are higher than DRAM's and its endurance is lower. Many DRAM-PCM hybrid memory systems use DRAM as a cache to PCM, to achieve the low access latency and energy, and high endurance of DRAM, while taking advantage of PCM's large capacity. A key question is what data to cache in DRAM to best exploit the advantages of each technology while avoiding its disadvantages as much as possible.

We propose a new caching policy that improves hybrid memory performance and energy efficiency. Our observation is that both DRAM and PCM banks employ row buffers that act as a cache for the most recently accessed memory row. Accesses that are row buffer hits incur similar latencies (and energy consumption) in DRAM and PCM, whereas accesses that are row buffer misses incur longer latencies (and higher energy consumption) in PCM. To exploit this, we devise a policy that avoids accessing in PCM data that frequently causes row buffer misses because such accesses are costly in terms of both latency and energy. Our policy tracks the row buffer miss counts of recently used rows in PCM, and caches in DRAM the rows that are predicted to incur frequent row buffer misses. Our proposed caching policy also takes into account the high write latencies of PCM, in addition to row buffer locality.

Compared to a conventional DRAM-PCM hybrid memory system (FREQ), our row buffer locality-aware caching policy (RBLA and RBLA-Dyn) improves system performance by 14% and energy efficiency by 10% on data-intensive server and cloud-type workloads. The proposed policy achieves 31% performance gain over an all-PCM memory system, and comes within 29% of the performance of an all- DRAM memory system (not taking PCM's capacity benefit into account) on evaluated workloads.

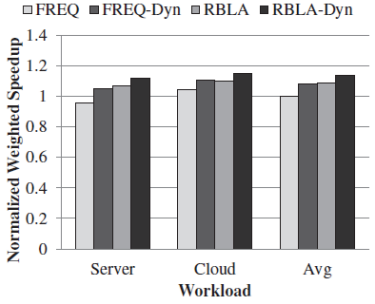


Fig. 1: Performance of RBLA and RBLA-Dyn.

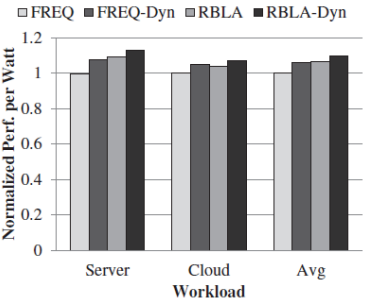


Fig 2: Energy Efficiency of RBLA and RBLA-Dyn.