PS-Cache: An Energy-Efficient Cache Design for Chip Multiprocessors

Joan J. Valls*, Alberto Ros†, Julio Sahuquillo*, María E. Gómez*

*Department of Computer Engineering
Universitat Politècnica de València (Spain)
joavalmo@fiv.upv.es, {jsahuqui,megomez}@disca.upv.es

†Dept. de Ingeniería y Tecnología de Computadores
Universidad de Murcia (Spain)
aros@ditec.um.es

1 Introduction

As silicon resources become increasingly abundant, core counts grow rapidly in successive chip-multiprocessors (CMP) generations. Parallel workloads represent an important segment for current and future CMPs mainly when many-core processors are considered. Unlike multiprogrammed workloads, the accessed blocks in these workloads can be classified in two categories: private, accessed only by one core, and shared, accessed by several cores. This paper takes advantage of this classification to access only a subset of the ways on each L1 cache access, thus reducing dynamic power consumption.

Recent research works have also concentrated on proposals that take advantage of the private-shared block access behavior to enhance the performance or energy consumption. For example, in [2] authors partition the last level cache (LLC) into private and shared ways by adding complex logic. As it is done in this work, the number of ways that are accessed on a cache access is reduced; hence also the dynamic energy consumption.

2 PS-Cache

The main goal of our approach is to take advantage of the Private-Shared classification to design a more power-efficient cache architecture that reduces the number of ways looked up on each cache access. Instead of accessing all the ways in a set, only a subset of them are searched.

Each cache line has attached a PS (Private-Shared) bit which indicates the type of each block (according to the page table and the TLBs, similar to [1]). The PS bit provided by the TLB is compared with the PS bits of all the ways in the set. A simple logic is included to select the wordline (WL) of those ways whose PS bit matches the value of that obtained from the TLB. This means that in the tag array only a subset of the tags are read and then compared with the tag of the physical address. Consequently only a subset of data blocks are read. This allows the proposal to reduce significant dynamic power consumption across the memory accesses since, in general, only a small fraction of ways is required to be searched in many memory operations.

In addition, the devised mechanism takes advantage of the invalid bit to reduce the static power consumption. The power of all ways in an invalid state is turned off so they are discarded from the process of searching the block.

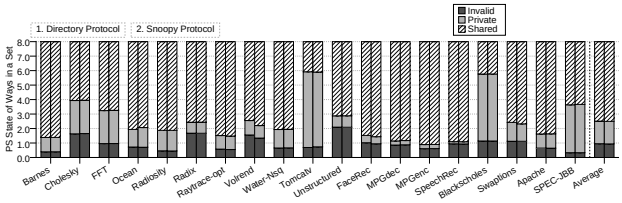

Figure 1: Average number of ways in a set of each type.

Benefits of the proposal mainly depend on the average number of ways that are looked up by the cache accesses. This numbervaries depending on which cache is being accessed to (L1 or L2), on the applications behavior and on the type of block we are looking for.

Figure 1 shows the average number of blocks of each type in the 8-way set associative L1 cache. Results are shown for both snoopy and directory-based protocols. As observed, on average, there are less than two private blocks in a set, whose access may result in important energy savings. Nevertheless this number depends greatly on the application.

Compared to conventional directory or snoopy protocols where all ways are searched, the PS cache can reduce the dynamic power consumption by as much as 22.46% and by 22.43% for directory and snoopy protocols respectively in the L1.

Finally, notice that performance is not damaged since only those ways that can contain a block are accesed.

3 Acknowledgments

This work was supported by the Spanish Ministerio de Economía y Competitividad (MINECO) and by FEDER funds under Grant TIN2012-38341-C04-01/03.

References

[1] B. Cuesta, A. Ros, M. E. Gómez, A. Robles, and J. Duato. Increasing the effectiveness of directory caches by deactivating coherence for private memory blocks. In *38th Int'l Symp. on Computer Architecture (ISCA)*, pages 93–103, June 2011.

[2] K. T. Sundararajan, V. Porpodas, T. M. Jones, N. P. Topham, and B. Franke. Cooperative partitioning: Energy-efficient cache partitioning for high-performance cmps. In *HPCA 2012*, pages 311–322.