ILLINOIS INSTITUTE OF TECHNOLOGY

Scalable Computing Software Laboratory Technical Report

Department of Computer Science

Illinois Institute of Technology

Optimizing Memory Concurrency at Each Memory Layer in a Multi-Tasking Environment

Yu-Hang Liu
Department of Computer Science
Illinois Institute of Technology
yuhang.liu@ iit.edu

Xian-He Sun
Department of Computer Science
Illinois Institute of Technology
sun@iit.edu

Oct 01, 2015

Technical Report No. IIT/CS-SCS2015-10

http://www.cs.iit.edu
10 West 31st Street, Chicago, IL 60616

ABSTRACT

As applications become more and more data intensive, memory system performance becomes even more vital to the performance of future many-core processors. It is a challenge to smartly allocate and thus efficiently utilize memory systems to improve system performance. This is especially true for utilizing memory concurrency where the required concurrency varies dynamically with data access patterns and different tasks may have different data access patterns. In this study, we propose a system solution to determine and allocate memory concurrency, in terms of memory banks, to utilize memory performance. We first introduce a recursive formula to reveal the impact of memory concurrency at each memory layer toward the final system performance. We, then, formalized the concurrency issue and transformed it into an optimization problem. Finally, we propose an analytical methodology, Smart-C, to determine the optimal concurrency at each layer of a memory hierarchy automatically and systemically. Smart-C can determine, and allocate, the appropriate concurrency for each individual task in a multi-tasking environment for best overall performance. Cycle-accurate simulations show that our adaptive design can smartly manage memory hardware concurrency. Compared to the conventional equally allocation approach, our method can reduce data stall time by up to 5.2-folds, and can improve performance (IPC) by up to 14.11%, and 7.96% on average, with only a small amount of hardware overhead.

Categories and Subject Descriptors

D.3.3 [**Computer Systems Organization**]: Performance of Systems

General Terms

Performance, Algorithms, Theory, Measurement, Verification

Keywords

Many-core processors; high-end computing systems; data access patterns; data stall time; memory concurrency; dynamic partitioning