Abstract

This paper proposes a novel efficient Non-Uniform Cache Architecture (NUCA) scheme for the Last-Level Cache (LLC) to reduce the average on-chip access latency and improve core isolation in Chip Multiprocessors (CMP). The architecture proposed is expected to improve upon the various NUCA schemes proposed so far such as S-NUCA, D-NUCA and SP-NUCA[9][10][5] in terms of average access latency without a significant reduction in the hit rate. The complete set of L2 banks is divided into various zones. Each core belongs to one particular zone which is the closest to it. Consequently, adjacent cores are grouped into the same zone. Each zone individually follows the SP-NUCA scheme [5] for maintaining core isolation and sharing common blocks. However, blocks that need to be shared by cores which belong to different zones are replicated. This scheme is much more scalable than the SP-NUCA scheme and bounds the maximum on-chip access latency to a lower value as the number of
cores increases.

This paper merely details the proposed scheme. The claims made regarding the benefits of the scheme shall be substantiated through simulations and a detailed comparative study in the future. The intended simulation methodology and architectural framework to be used in this regard have also been mentioned.

Reference

- J. Chang and G. S. Sohi. Cooperative caching for chip multiprocessors. In 33rd
International Symposium on Computer Architecture, 2006

Index Terms

Computer Science
Microprocessor

Key words
Chip Multiprocessor (CMP)
Non-Uniform Cache Architecture (NUCA)

Shared Last level Cache (LLC)