ADAPTIVE MECHANISMS AND POLICIES FOR MANAGING CACHE HIERARCHIES IN CHIP MULTIPROCESSORS

CRITIQUE

By Srivatsan Chellappa

The paper discusses methods for improving cache performance for the emerging chip multicore processors. The methods discussed are introducing another level of cache called the L3 cache for holding the evicted lines from the L2 cache, avoiding write back of clean lines to memory, Maximizing L2-L2 transfers thus using peer L2 lines for writing back.

PROS:
1) The paper introduces a new level of cache which is bound to reduce the miss rate and certainly improves performance.
2) The paper emphasizes on removing unnecessary writes like writing back clean lines. This reduces the memory traffic and decreases memory latency.
3) It suggests ways to improve the on chip memory usage by encouraging writing to peer L2 caches while writing back.
4) The Write Back History table method produces up to 75.3% accurate predictions.
5) The paper has also studied the complexities that arise when the number of cores increase i.e. the modifications needed for the coherence protocols to implement L2 snarfing.

CONS:
1) The complexities involved with introducing a new level of cache have not been analyzed. That is the area and power constraints have not been studied at all. I feel that having another level of cache might not be feasible with the increasing number of cores as on chip area is very costly.
2) Moreover the paper proposes to introduce a history table for each L2 cache. This would greatly restrict the number of cores we can have on a chip or alternatively is bound to increase the chip size.
3) The techniques provided show an average increase of 13% in performance. But the performance impact variations with increasing the number of cores on a chip have not been studied. That is the complexity of extra hardware and greater die area for 13% improvement in performance is debatable.
4) Section 2.1 of the paper talks about the potential negative effects of the use of WBHTs and selective L2 write back scheme. But nothing has been proposed to overcome these.
5) L2 snarfing might decrease the off- chip accesses but definitely increases the on chip accesses and this costs power and heating. These issues have not been dealt with in the paper.
6) The paper would have been better if simulations had been carried out for greater number of cores and had considered power issues.