

Shared Caches in Multicores: The Good, The Bad, and The Ugly

Mary Jane Irwin
Pennsylvania State University

Abstract: As we transition from clock-frequency performance scaling to performance scaling with multicores, the pressure on the memory hierarchy is increasing dramatically. Many different on-chip cache topologies have been proposed/implemented; effective management of these shared caches is crucial to multicore performance.

This talk will begin with a description of a cache miss classification scheme for multicores (compulsory, inter-core misses, intra-core misses) that gives insight into the interactions between memory transactions of the different cores on a chip sharing a cache. Ways to improve the on-chip cache performance with architectural enhancements, compiler enhancements, and runtime system enhancements will then be discussed. If the application thread mapping and the on-chip topology is static (i.e., does not change during runtime), then compiler enhancements that support cache topology aware code optimization can be used to significantly improve an application's performance. Results from such an augmented compiler, where the topology is exposed to the compiler and where the compiler also does thread-to-core mapping assignments, will be presented. If the application thread mapping or the on-chip topology is dynamic, then other alternatives exist. For example, a thread scheduler, or allocator, can make decisions about moving threads to different cores during runtime in the hopes of improving overall cache performance. Initial experiments with the REEact system being developed by researchers at Penn State-UPittsburgh-UVirginia that "reacts" to hardware conditions (such as cache miss rates, hot-spots, etc.) by reallocating threads at runtime will be outlined. Finally, if the on-chip cache topology itself is dynamic (i.e., is designed to be reconfigurable at runtime), large performance benefits might be obtained. However, both hardware and software design challenges to realizing such a dynamic system abound. Some of these challenges will be briefly discussed.

Categories & Subject Descriptors: C.0 Computer Systems Organization; General.

General Term: Design.