SORS: Getting instructions from where they are to where they need to be

Objectives

Abstract: A good supply of instructions as a result of instruction fetching is critical to a processor's performance. This requires the efficient operation of first-level instructions TLBs (iTLBs), branch target buffers (BTBs) and L1 instruction caches. Increasing code sizes for server applications result in frequent misses in these first-level structures, and architects have resorted to a variety of prefetching techniques to bring the information needed by instruction fetch into these structures. We propose a new approach--instruction presending--where the cache blocks holding (static) instructions, and associated BTB and iTLB information, are autonomously sent from the larger, slower secondary structures on chip, where they reside, to the (potentially very) smaller primary structures (L1 cache, BTB, iTLB) close to the processor, in a "just in time" fashion. Instruction presending is extremely effective, especially for applications with a large code footprint (and high base MPKI), achieving significantly lower MPKI (e.g., order of magnitude for L1 cache) than even the best prefetching schemes. Presending opens up the possibility of significant other optimizations to, and of different microarchitectural components for, instruction processing.

Short Bio: Guri Sohi has been a

faculty member at the University of Wisconsin-Madison since 1985 where he currently a John P. Morgridge Professor and a Vilas Research Professor. His research has been in the design of high-performance microprocessors and computer systems. Results from his research can be found in almost every high-end microprocessor in the market today.

Speakers

Speaker: Professor Guri Sohi, University of Wisconsin-Madison
Host: Miquel Moretó, High Performance Domain-Specific Architectures Associated Researcher, CS, BSC

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

---