Researchers solve scaling challenge for multi-core chips

April 16 2012

Researchers sponsored by Semiconductor Research Corporation (SRC), the world's leading university-research consortium for semiconductors and related technologies, today announced that they have identified a path to overcome challenges for scaling multi-core semiconductors by successfully addressing how to scale memory communications among the cores. The results can lead to continued design of ever-smaller integrated circuits (ICs) into computer hardware without expensive writing of all new software from scratch to accommodate the increased capabilities.

Today's announcement involves researchers Professor Daniel Sorin from Duke University, Professor Milo M.K. Martin from University of Pennsylvania and Professor Mark D. Hill from University of Wisconsin. The SRC-guided research significantly extends the path for cores to communicate by reading and writing to a shared space – known as cache-coherent shared memory. In each core, one or more caches hold the subset of memory locations that most recently have been written and read by the core.

Cache coherence protocols are built into hardware in order to guarantee that each cache and memory controller can access shared data at high performance. As computational demands on the cores increase, so do concerns that the protocol will be slow or energy-inefficient when there are multiple cores.

"We have refuted calls for a radical design change by showing that, using

already existing techniques, we can create cache coherence protocols that scale to hundreds and perhaps even thousands of cores," said Sorin.

"Our results allow us to confidently predict that, with these new protocols, on-chip coherence is here to stay. Computer systems don't need to abandon current compatibilities to accommodate even hundreds of cores," Sorin added. "Chip area and energy consumption may limit future multi-core chips, but our research refutes conventional wisdom that multi-core scalability of the memory system would be the primary scaling bottleneck."

The alleged lack of scalability of coherence is attributed to the poor scaling of the storage and traffic on the interconnection network that coherence requires, as well as concerns about latency and energy needs. A popular expectation among industry has projected that future multi-core chips will no longer be able to rely on coherence, but instead will communicate with software-managed coherence or message passing that does not share memory. For the past few years, high costs estimated for support of those alternatives have registered growing concern among computer hardware manufacturers.

The solution described by the research brings together a combination of identified techniques for creation of shared caches augmented to track cached copies, explicit cache eviction notifications and hierarchical design. Scalability analysis of this design confirms that shared memory among multiple cores and its vital benefits for future computational increases can allow a broad range of technologies and industries to maintain their reliance on more powerful, cost-effective roadmaps.

"Chipmakers are not operating in a vacuum and must continue to identify how they'll enable their partners on the hardware side," said SRC Executive Vice President Steven Hillenius. "As we collectively grapple with how to keep costs low and performance high for the next

generation of computational technologies, our announcement today is that one of the key problems for scaling can be solved."

This news means that not only will the computer industry be able to avoid radically changing the programming paradigm from the mainstream technique of cache-coherent shared memory, but the solution developed by Sorin and his colleagues also facilitates backward compatibility with the vast amount of legacy code written for cache-coherent shared memory. Thus, as the industry plans for the future, it gains a path for scalability without requiring all new software.

Provided by Semiconductor Research Corporation