

Piranha: Exploiting Single-Chip Multiprocessing

Luiz André Barroso, Kourosh Gharachorloo, Tom Heynemann, Dan Joyce, David Lowell, Harland Maxwell, Joel McCormack, Ravishankar Mosur, Jeff Sprouse, Robert Stets, and Scott Smith
Compaq Computer Corp.

Today's microprocessor industry struggles with escalating development and design costs, which arise from exceedingly complex processors that push the limits of instruction-level parallelism. Meanwhile, such designs yield diminishing returns and are ill-suited for commercial applications such as database and Web workloads, which constitute the high-performance servers' most important market. These server applications typically suffer from large memory stall times, exhibit little instruction-level parallelism, and have no use for high-performance floating-point or multimedia functionality.

Fortunately, increasing chip densities provide architects with many options for tackling design complexities while addressing commercial applications' needs. Integrating all system-level components onto the processor die—as the upcoming Alpha 21364 does—enables a more efficient memory hierarchy without further increasing design complexity. Beyond that, exploiting the abundance of thread-level parallelism in commercial workloads through simultaneous multithreading or chip multiprocessing seems promising. The

chip multiprocessing approach is particularly useful for addressing design complexity because it enables using simpler cores.

The Piranha project's¹ primary goals are to

- build a system that achieves superior performance on commercial workloads, and
- effectively address design cost and complexity issues.

Our research prototype aggressively exploits chip multiprocessing by integrating eight simple Alpha processor cores along with a two-level cache hierarchy, memory controllers, coherence protocol engines, and an interconnect router onto a single chip. Combining simple, single-issue in-order processor cores with an industry-standard ASIC design methodology should let us complete our design with a shorter schedule and smaller team and budget than a commercial microprocessor requires.

Although each Piranha processor core is substantially slower than a conventional next-generation processor because of its simpler design and the constraints of an ASIC process, integrating eight cores onto a single chip provides Piranha with a two-fold to threefold performance margin on important commercial workloads. This advantage can approach a factor of five using full-custom instead of ASIC logic.

Most of Piranha's architectural innovations lie in the memory and interconnect

subsystems, including a shared eight-way banked, eight-way associative noninclusive L2 cache; highly specialized microprogrammed coherence protocol engines; and an aggressive two-level coherence protocol. Piranha is particularly efficient for applications with little instruction-level parallelism because it can exploit thread-level parallelism to issue multiple independent misses and better utilize its aggressive memory. In addition, significant constructive cache interference among threads for applications such as databases lets a relatively small, 1-Mbyte L2 cache effectively handle the eight processor cores.

While the exceedingly complex general-purpose architecture of most current processors is not optimal for any given application domain, Piranha's focused design targets an important market segment at the possible expense of other workloads, resulting in superior performance and improved time to market.

Reference

1. L.A. Barroso et al., "Piranha: A Scalable Architecture Based on Single-Chip Multiprocessing," *Proc. 27th Ann. Int'l Symp. Computer Architecture (ISCA 2000)*, ACM Press, New York, June 2000, pp. 282-293.

The authors are affiliated with Compaq Computer Corp. Contact Luiz André Barroso at Luiz.Barroso@compaq.com.