Simulation and Evaluation Techniques for Modern Processors

Kevin Skadron Dept. of Computer Science University of Virginia Charlottesville, VA 22904 Margaret Martonosi Dept. of Electrical Engineering Princeton University Princeton, NJ 08544

Abstract

The field of Computer Architecture is heading toward significant experimental evaluation problems that constitute a fundamental threat to continuing advances in computational speed. These problems are severe in all types of modeling and simulation, but are especially severe in performing simulations that guide the design flow for future computer systems. In particular, microprocessors are doubling in transistor count every 18 months and microprocessors architectures are rapidly becoming more complex. Unfortunately, the ever-more-sophisticated processors and workloads of tomorrow must be evaluated using the slower processors of today. Furthermore, cycle-accurate processor simulation of hypothetical architectures currently requires software modeling, exacerbating the slow speed of simulation. The recent interest in modeling power as well as performance makes the problem even worse.

Techniques for performing CPU simulations are therefore not keeping up. As an example of the speed problems already in existence today, current simulators like Wattch that model both pipeline and power effects can simulate only about 200 million instructions per hour on top-of-the-line computers. Yet the standard benchmarks in use—the SPECcpu2000 programs—run for hundreds of billions or trillions of instructions. At this rate some of these benchmarks would take over a year to simulate to completion! Since architectural studies typically require a diverse workload and often must explore a rich parameter space, techniques for reducing simulation time are essential. Clearly aggressive sampling techniques, aggressive workload-scaling techniques, more sophisticated parameter-search techniques, and analytical models with guaranteed bounds on accuracy are necessary. Unfortunately, despite ongoing research into making simulations faster, there is not yet a consensus in the architecture community on what constitutes valid technique for controlling simulation time. Most architectural studies today use crude sampling techniques to reduce simulation times, or make a variety of simplifying assumptions.

Worse yet, we are not aware of any proposed techniques that can keep up with the relentless pace of processor complexity and continue to provide reasonable simulation times and accuracy 5-10 years out. Over this longer time period, additional problems will arise. It is difficult to predict exactly what these problems will be, but already we can identify several important trends that are expected to make processor evaluation even more difficult. Technological/fab limits will play a more important role in determining feasible architectures. This will likely require yet another level of modeling in addition to performance and power. Furthermore, increasing performance must clearly continue to come not only from improvements in clock speed, but also from exploiting various types of parallelism: instruction-level parallelism, task-level parallelism, and coarse-grained parallelism. Mechanisms for exploiting all these types may appear on the same chip. Yet current simulators for multi-threaded and multi-processor systems are much slower even than today's uniprocessor simulators. Finally, heterogeneous chips, in which different processing units, various specialized computation units, and various I/O drivers appear on the same die, may become increasingly common. This too will complicate modeling, especially if the model for each component is effectively its own simulator.

Solving these evaluation problems is of fundamental importance to computer science and also to the many other fields that depend on continuing increases in computational speed. Without fast, accurate, and efficient evaluation techniques, the pace of innovation in computer architecture is likely to slow as more and more time must be spent on modeling and verification. These questions also constitute one of the vital and intellectually challenging areas of inquiry in computer science today. They touch on the basic questions of what architectural features have the greatest leverage over simulation accuracy; how various architectural features and their effects on simulation are related, what architectural trends are likely to place the most pressure on evaluation speed in the future, and what will constitute a representative workload in the future.

Overall, improving our ability to evaluate future microprocessors will help not only those fields that depend on increased processing speed, but is also likely to yield new insights into the area of computer architecture itself.