ZOOM: A Performance-Energy Cache Simulator

Regina Sam, Krste Asanović, and Jude A. Rivers[1]

Introduction: Caches play a crucial role in narrowing the ever-widening processor-memory performance gap. A cache designer has to balance, often conflicting, demands for power, speed and area. Reaching an optimum compromise between these three dimensions of design requires a firm grasp of the effects of micro-architectural changes. However, exploring the large design space of candidate cache architectures using conventional circuit simulation tools would be extremely tedious and time-consuming. Hence, simplified analytical models are very valuable. ZOOM is a fast, flexible and robust analytical tool for *functional* and *micro-architecture* level simulation of caches. The functional simulator supports a wide range of caches (multi-level separate/unified instruction and data caches with copy-back, write-through, write-allocate and non-allocating caches etc). At the micro-architecture level, the analytical models enhance models proposed by several authors ([1] - [5]) to accomodate low-power design techniques and flexible energy-delay optimization.

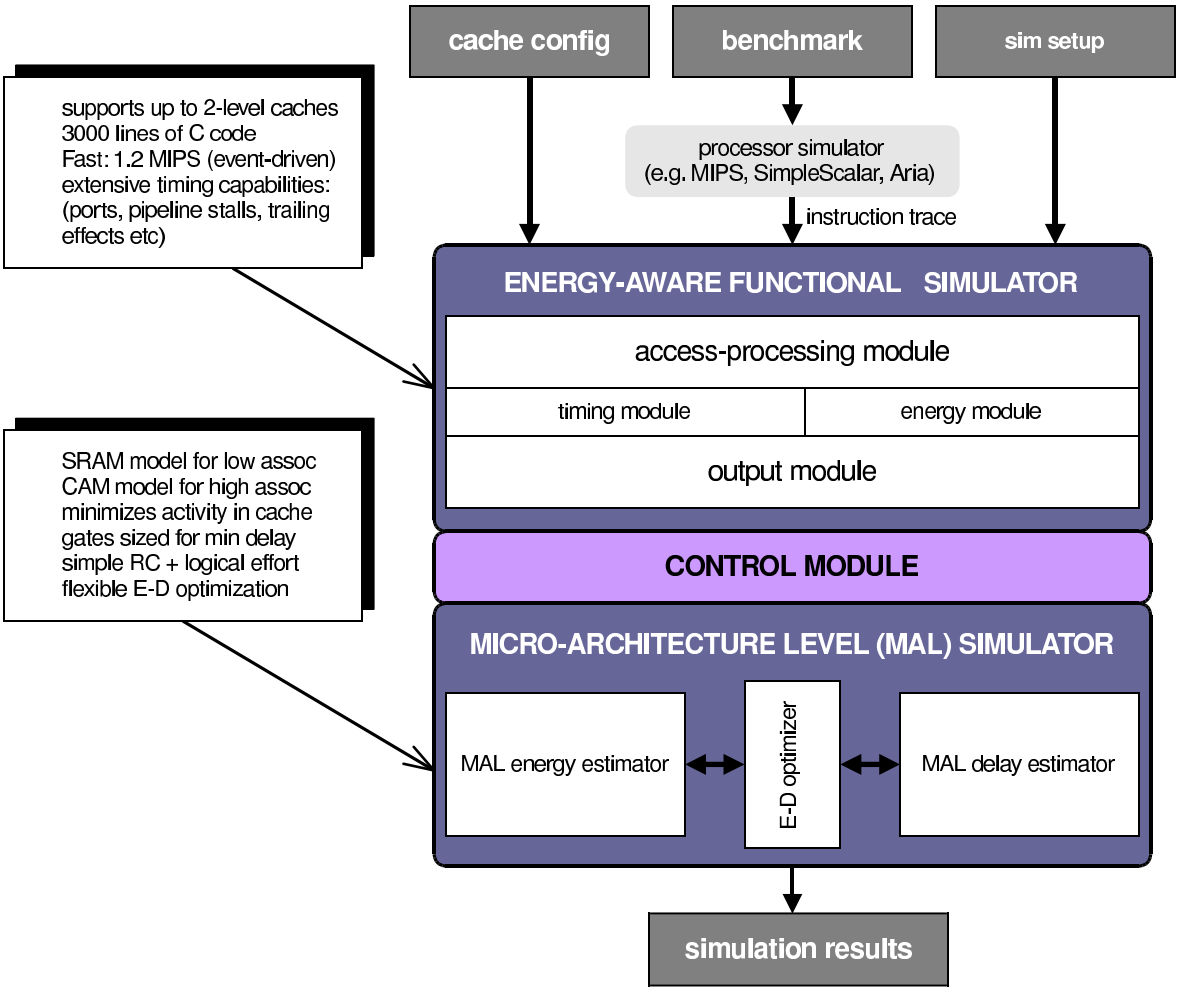

Figure 1: Overview of ZOOM

Approach: Speed and flexibility are fundamental goals at the *functional* simulation level. Since benchmarks may consist of millions, or possibly billions of instructions, only a tractable number of computations may be employed by

[1] Jude A. Rivers is a research staff member at the IBM T. J. Watson Research Center in Yorktown Heights, New York

the simulator. We assume that the typical user of the functional cache simulator is primarily interested in studying the effects of configuration changes (cache size, associativity, block size, buffer sizes etc) on the dynamic metrics of the cache (hits, cache/buffer utility, access patterns/block usage) in the context of benchmarks representative of target applications. Hence, the functional simulator should be easily adaptable to novel multi-lateral cache configurations.

The *micro-architecture* level models used in the simulator are tailored towards low-power cache designs. Although the program user may specify to what extent energy and/or delay influence the cache architecture assumed by ZOOM in its estimates, ZOOM only considers the tractable design space of energy-efficient techniques/architectures commonly used in today's caches. Due to the rapid rate of change of process technology, the ability of the models to adapt to future process technologies with little or no loss of accuracy is crucial. Meeting both of these goals requires extensive parametrization of the analytical models.

Progress and Future: ZOOM is in the advanced stages of development. The functional simulator is complete and working. The micro-architecture level simulator, save the E-D optimizer, is fully implemented. The next immediate step is to validate the energy and delay estimates from ZOOM with simulated estimates from HSPICE. Since the area of a cache is another issue of concern to designers, we hope to add an area model to ZOOM in the near future.

Research Support: This research was supported by an IBM VI-A fellowship, DARPA PAC/C award number F30602-00-2-0562, NSF CAREER award CCR-0093354, and an equipment grant from Intel Corporation.

References:

[1] T. Wada, S. Rajan, and S. A. Przybylski, "An analytical access time model for on-chip cache memories," *IEEE Journal of Solid-State Circuits* vol. 27, pp. 1147-1156, Aug. 1992.

[2] S. Wilton, N. P. Jouppi. "An enhanced access and cycle time model for on-chip caches." *Technical Report 93/5, Compaq Western Research Lab*, July 1994

[3] R. J. Evans, P. D. Franzon, "Energy consumption modeling and optimization for SRAMs," *IEEE Journal of Solid-State Circuits* vol. 30, pp. 571-579, May 1995.

[4] M. Zhang, K. Asanovic. "Highly-associative caches for low-power processors" *Kool Chips Workshop, 33rd International symposium on Microarchitecture,* December 2000

[5] B.S. Amrutur, M. A. Horowitz. "Speed and power scaling of SRAMs," *IEEE Transactions on solid-state circuits,* vol. 35, No. 2, February 2000

[6] B.S. Amrutur, M. A. Horowitz. "Fast low-power decoders for RAMs," *IEEE Journal of solid-state circuits,* vol. 36, No. 10, October 2001