2017 International Symposium on Computer Architecture Influential Paper Award

David Brooks
Harvard University

The International Symposium on Computer Architecture (ISCA) has a tradition of awarding the ACM SIGARCH/IEEE-CS TCCA Influential ISCA Paper Award at the conference each year. This award is conferred on the authors of a paper from the ISCA conference that occurred 15 years prior and which had a substantial impact on the field in terms of research impact and/or industrial influence. The selection process starts by soliciting nominations by members of the current year's ISCA Program Committee. The top papers are then voted on by the full PC (excepting conflicts). The results of this vote are conveyed to a selection committee comprising the current ISCA PC Chair (David Brooks), the ACM Special Interest Group on Computer Architecture (SIGARCH) Chair (Sarita Adve), and the IEEE Computer Society Technical Committee on Computer Architecture (TCCA) Chair (Dean Tullsen). The award includes an honorarium for the authors and a certificate.

At ISCA 2017, the award was presented to the authors of the ISCA paper published in 2002 titled "Drowsy Caches: Simple Techniques for Reducing Leakage Power." The paper was written by Krisztián Flautner, Nam Sung Kim, Steven M. Martin, David Blaauw, and Trevor N. Mudge. At the time of publication in 2002, the computer architecture community was beginning to realize that power concerns would be a major problem for future high-performance and mobile microprocessors. However, most attention was focused on dynamic power consumption, rather than the static, or leakage, power that was growing in importance. In fact, at the time it was projected that leakage power would dominate total power consumption below the 90-nm technology node. "Drowsy Caches" was one of the first papers to address this growing problem area.

Like most influential papers in computer architecture, the key idea of the Drowsy Cache is simple. The authors observe that for fixed periods of time, most cache accesses occur on a small subset of cache lines. The Drowsy Cache is designed to take advantage of this property by splitting the cache lines into active and drowsy states. In the active mode, the cache line can be accessed as normal. In the drowsy mode, the supply voltage to the cache line is reduced to the point where the leakage current is significantly reduced, but the voltage is maintained at a level that allows data retention. A small performance hit is incurred when moving between the drowsy and active states, so the paper proposes architectural policy mechanisms that can be implemented to move lines between the states. The paper shows that with simple policy mechanisms, up to 90 percent of the cache lines can be in the drowsy state without impacting the overall performance by more than 1 percent. The drowsy approach contrasts with prior Gated V_{DD} techniques that turn off cache lines completely, resulting in state loss and the need to fetch the data from lower levels of the memory hierarchy.

A groundbreaking aspect of the paper was the strong collaborative effort between computer architecture and circuit design. This is reflected both in the list of authors and the Drowsy Cache design itself. Memory circuits are notoriously difficult to design because of the tradeoff between array density and susceptibility to process variation and on-chip noise. Thus, one concern with the Drowsy Cache approach is that the techniques needed to create the drowsy mode would be unreliable or require significant chip area. The paper provides comprehensive circuit diagrams to explain how the memory circuits need to be modified to support drowsy operation. The paper also includes detailed HSPICE simulations demonstrating cross-talk analysis of internal nodes of the memory and the expected leakage savings benefits. At the same time, the previously proposed Gated V_{DD} techniques required somewhat complex control algorithms to maintain correctness due

Published by the IEEE Computer Society
0272-1732/17/$33.00

to the loss of state from completely disabling the cache lines. The Drowsy Cache paper describes a relatively simple architectural policy mechanism and evaluates the overall energy savings and performance impact on both in-order and out-of-order microprocessor cores using state-of-the-art architectural simulation approaches across a range of benchmarks. In this regard, the paper is a model for researchers working at the interface between computer architecture and circuit design.

The Drowsy Caches paper has had a substantial impact on the research community and has been cited more than 1,000 times as of September 2017. One can also see the influence of the Drowsy Cache work in modern microprocessors that implement aggressive power optimizations in the cache hierarchy. For example, the Intel Xeon Processor 7100 includes leakage power management in the L3 cache design. The design uses sleep transistors that allow fine-grained control of leakage power in the cache subarray blocks with wake-up counters that can be programmed to balance switching and leakage power. Clearly, the Drowsy Cache paper has withstood the test of time and is a worthy recipient of the 2017 SIGARCH/TCCA Influential Paper Award.

David Brooks is the Haley Family Professor of Computer Science at Harvard University. Contact him at dbrooks@eecs.harvard.edu.