

James Donald
Meta

Margaret Martonosi
Princeton University

I. SETTING

This paper was published in ISCA 2006, as part of a wave of power and thermal modeling research efforts that had swept through the field, particularly in the decade from the late 1990's into the 2000's. One of the first ISCA papers to consider power was "Pipeline Gating: Speculation Control for Energy Reduction" by Srilatha Manne, Artur Klauser, and Dirk Grunwald, which appeared in 1998. Given the lack of modeling tools available, such early works demonstrated the advantages of their ideas using proxy measures, such as the reduction in speculative instructions. With the ISCA 2000 publication of papers about the Wattch and SimplePower power simulation tools, the field was better equipped to perform architecture-level power modeling. Then, in ISCA 2003 the HotSpot tool further demonstrated the importance of distinguishing power and thermal modeling issues.

What happened next was the dramatic march towards multicore parallelism. In fact, ISCA 2006 included several papers about different aspects of multicore design, their shared cache hierarchies, Networks-on-Chip and more. With this shift to multicore design issues, the field also needed a shift to power and thermal tools built to consider them and quantify tradeoffs.

II. THIS PAPER

Framed in that context above, our paper offered several key contributions:

- **Heterogeneous Thermal Management:** While the field was still primarily focused on homogeneous on-chip parallelism, this paper took initial steps towards heterogeneity. In particular, while the cores were homogeneous, we asked and answered the question: should we manage all cores the same or differently?
- **Multi-loop Formal Control:** In prior work, Skadron and others had paved the way regarding formal power/thermal control overall. Our paper proposed a multi-loop formal approach for multi-core thermal management.
- **Thermal estimation and Sensing:** One key question in a thermal management paper is how will we know temperatures in order to know when to invoke some management action. In this paper, we explored multiple options including using OS-level performance counters to estimate application thermal behavior, or assuming the presence of hardware thermal sensors to be read and acted upon.
- **Exploration and taxonomy of thermal management policies:** The next question of thermal management concerns what actions to take when thermal management

is warranted. The paper considered both OS-level and hardware-supported methods for thermal management, and also considered whether to use centralized uniform adjustments or to adopt per-core specialized management. In addition to techniques like Stop-go or Dynamic Voltage and Frequency Scaling (DVFS), one could also adjust the thermal profile by migrating workloads; we evaluated these opportunities as well.

- **Real-system thermal measurements:** One thing that sets this paper apart from others of that era was that we provided real-system measurements in our results. While these represented a lot of hard work for first author James Donald, they set the paper apart from many others, and they were very helpful in validating our ideas and our approaches.

III. WHERE IS THE FIELD TODAY?

Seventeen years later, it is exciting to see how the field has progressed. In particular, processor and system thermal management is universal. Many of the ideas from the research papers of the early 2000's have been broadly adopted for industry use. Likewise, on-chip parallelism is also universal, and the ideas of how to manage multi-core systems have similarly seen broad uptake. This work, along with others of the era, laid some of the groundwork as that transformation occurred.

IV. WHAT HAPPENED TO THE AUTHORS?

James Donald received his Ph.D in 2007 and joined NVIDIA where he worked on three GPU architectures. As of today he is at Meta and has contributed to the Rift and Quest product lines.

Margaret Martonosi is (still!) on the faculty at Princeton University, where she is now the H. T. Adams '35 Professor of Computer Science.