DOI 10.1109/MM.2018.043191128

2018 International Symposium on Computer Architecture Influential Paper Award

Antonio González
Universitat Politècnica de Catalunya, Barcelona, Spain

The International Symposium on Computer Architecture (ISCA) recognizes every year the most influential paper published in this conference 15 years earlier, based on its impact on research, development, products or ideas. This award is sponsored by the IEEE Computer Society Technical Committee on Computer Architecture (IEEE-CS TCCA) and the ACM Special Interest Group on Computer Architecture (ACM SIGARCH). In this year's edition, the candidate papers were those papers published in ISCA 2003 proceedings. The selection process was chaired by Antonio González. Candidate papers for the award were selected by the current year's ISCA Program Committee. The final award selection was made by the Award Chair (Antonio González), the IEEE-CS TCCA Chair (Lieven Eeckhout) and the ACM SIGARCH Chair (Sarita Adve). The award includes an honorarium for the authors and a certificate.

The 2018 award was presented to "Temperature-Aware Microarchitecture" by Kevin Skadron, Mircea R. Stan, Wei Huang, Sivakumar Velusamy, Karthik Sankaranarayanan and David Tarjan. All the authors were at the University of Virginia at the time of publication.

This work was carried out at a time when power dissipation of microprocessors was starting to be regarded as a major concern by the scientific and industrial community. Power dissipation of microprocessors had been increasing exponentially over the past years and this trend was unsustainable. Some pioneering works during the late 1990s and early 2000s observed that power dissipation would soon limit the increase in performance that Moore's law should theoretically provide, meaning that effective use of both the increasing transistor density and their faster switching rates would be jeopardized by their associated power increase. These observations motivated an increasing number of works in the area of power reduction, which has continued as a hot research topic since then.

The awarded paper, building on prior work by Brooks and Martonosi[1], and Skadron, Abdelzaher, and Stan[2], provided a new perspective on the problem by focusing on the thermal issues caused by power dissipation. Chip operating temperature has important implications in performance and energy consumption, but most importantly, it cannot exceed the threshold associated with reliable operation. The paper claimed that many power reduction techniques may have little effect on operating temperature because they may not reduce power density in hot spots. Temperature-specific techniques had been confined up until then to thermal package design (heat sink, fan,

etc.) or costly system-level techniques such as stopping the clock when operating temperature exceeded a safe threshold. On the other hand, this paper made a case for architecture-level thermal management techniques. These techniques can work at a finer granularity, both in terms of time and space, by reacting more quickly and focusing on the particular blocks that are responsible for the hot spots at each moment.

The paper proposed a number of microarchitecture-level techniques such localized toggling of microprocessor blocks (i.e., dynamically adjusting the duty cycle based on the operating temperature), migrating computation to spare units located in different parts of the chip to better distribute the heat, and dynamic frequency scaling based on temperature tracking. This paper also showed the value of formal feedback control methods in managing these techniques.

However, the most important contribution of the paper was a novel approach for modeling temperature at the architecture level. This approach was based on a well-known analogy between heat transfer and electrical phenomena by which heat flow can be described as a "current" passing through a thermal resistance and leading to a temperature difference analogous to a "voltage". Thermal capacitance is used for modeling the delay in temperature change produced by a change in power. Each microarchitecture block of the system represents a node in this RC circuit, whose thermal resistance and thermal capacitance depends of the physical characteristics of the block, mainly its area and thickness, in addition to the physical properties of the material used for each component (e.g., silicon, copper, etc.). Using this approach, the authors developed a tool called *HotSpot*, validated it using a commercial, finite-element simulator of 3D fluid and heat flow, and made it publicly available.

This paper has had a tremendous impact on the computer architecture community. It galvanized the research community to explore architecture-level thermal management to improve performance, as opposed to just using chip-level techniques. It spurred a multitude of new proposals in this area, and the availability of the *HotSpot* tool made it possible to evaluate these ideas in a simple and cost-effective manner. In the years since its publication, this paper has had more than 1,300 citations and the *HotSpot* tool has been used by thousands of students and researchers. I congratulate the authors on their fine work and its high impact.

REFERENCES

1. D. Brooks and M. Martonosi, "Dynamic Thermal Management for High-Performance Microprocessors", *Proc. 7th Int. Symp. High-Performance Computer Architecture,* 2001, pp. 171-182.
2. K. Skadron, T. Abdelzaher and M.R. Stan, "Control-Theoretic Techniques and Thermal-RC Modeling for Accurate and Localized Dynamic Thermal Management", *Proc. 8th Int. Symp. High-Performance Computer Architecture,* 2002, pp. 17-28.

Antonio González is a professor in the Computer Architecture Department at the Polytechnic University of Catalonia (UPC) in Barcelona, Spain. He is the director of the Architectures and Compilers research group, focusing on energy-efficient computing systems. González has a PhD in Computer Architecture from UPC. He is a fellow of IEEE.