DARK SILICON

Michael B. Taylor
Steven Swanson
University of California, San Diego

In the past decade, the semiconductor industry has encountered turbulent times precipitated by a tectonic shift in processor design metrics: before, we focused primarily on frequency and area; now, we focus almost exclusively on power and energy. This shift has transformed our industry and the entire digital design stack. It has determined the features in our chip design tools, driven new architectures for our transistors, made obsolete entire circuit families, shrunk our pipelines, and mutated our instruction sets and programming models.

Although this arsenal of techniques has dampened the power wall, only in the past few years have we come to a widespread understanding of the source of these power problems. As we advance toward the final nodes of Moore's law, we can now estimate power's inexorable impact on scaling, and use it to moderate our expectations of the future fruits of Moore's law. This utilization wall[1] will force us to use only a small fraction of available chip area at full frequency at any one time. The remaining area that must remain unclocked, or underclocked, has become known as "dark silicon."[2] This utilization wall is getting exponentially worse with each process generation.

Although the utilization wall is indeed a dark message for semiconductor scaling, several researchers have started to examine approaches that put dark silicon to work. This is the focus of this special issue. We briefly introduce each of the issue's articles below, and describe how each augments the discourse in this important space. We apologize in advance to the authors for being less articulate than they in describing their work.

In "A Landscape of the New Dark Silicon Design Regime," Michael B. Taylor surveys the emerging dark silicon research area. His article starts with a derivation of the utilization wall, revisits the impact of dark silicon on multicore, and describes what has become known as the "four horsemen"—four key approaches that researchers are examining to address the problems and opportunities presented by dark silicon. The four horsemen, in order, are: first, shrinking chips to maintain high duty cycles; second, using dim operation modes to explicitly manage the duty cycles of a chip; third, exploiting specialization to reduce capacitance per function; and fourth, making use of new, post-MOSFET devices. All four options are entertained in this issue. Beyond the four horsemen, the article presents several near-term evolutionary design principles for dark silicon, and shows how the human brain provides some fresh ideas for more revolutionary approaches for putting dark silicon to better use.

"Utilizing Dark Silicon to Save Energy with Computational Sprinting" by Arun Raghavan et al. looks at dim silicon, the second horseman. The authors propose to exploit the thermal capacitance of materials, allowing chips to exceed their power budget for brief bursts until the chip reaches its peak operating on-chip temperatures. This allows the chip to "sprint" ahead for latency-bound tasks. Interestingly, they discovered that although this approach burns more energy in the short term, it can save energy in the long term for "race-to-idle" computations.

Nathaniel Pinckney and his team from the University of Michigan also examine dim silicon. In "Limits of Parallelism and Boosting in Dim Silicon," they propose the

Published by the IEEE Computer Society

use of more cores than could ordinarily be powered at nominal voltages; instead, they reduce the voltages to levels near the threshold voltage to reduce energy, and then parallelization to recoup performance used by the lowered voltage. They examine the impact of Amdahl overheads, including sublinear speedups, leakage, and memory and interconnect energy. They also examine the benefits of boosted, higher-voltage operation that targets performance-limiting serial regions.

The fourth article, ''Implications of the Power Wall: Dim Cores and Reconfigurable Logic,'' examines the third horseman, specialization. Authors Liang Wang and Kevin Skadron show approaches for modeling trade-offs between the use of accelerators and reconfigurable logic. Among the article's several contributions, they present a model that suggests that, for workloads with ideal parallelism and a collection of diverse tasks, reconfigurable logic beats accelerators because the reconfigurable logic can be reused across many tasks, offering a higher mean speedup.

In the final article, ''Steep-Slope Devices: From Dark to Dim Silicon,'' Karthik Swaminathan et al. focus on the fourth horseman, looking at the incorporation of tunnel field-effect transistors (TFETs) into the computational stack. TFETs are a post-MOSFET device that enables subthreshold slopes below 60 mV/decade, greatly reducing leakage. Although TFETs offer reduced voltage and energy levels, they do not provide as good performance as MOSFETs at higher voltages. As a result, the authors propose a heterogeneous multicore chip that includes both MOSFET- and TFET-based cores.

Although silicon chips of the future are getting darker, this new design regime has created exciting times for researchers and practitioners alike. We trust that you will agree that these articles shine a bright light on the possibilities for exploiting dark silicon in the future. MICRO

References

1. G. Venkatesh et al., ''Conservation Cores: Reducing the Energy of Mature Computations,'' *Proc. 15th Architectural Support for Programming Languages and Operating Systems Conf.,* ACM, 2010, pp. 205-218.
2. R. Merrit, ''ARM CTO: Power Surge Could Create 'Dark Silicon,''' *EE Times,* 22 Oct. 2009.

Michael B. Taylor is an associate professor in the Department of Computer Science and Engineering at the University of California, San Diego, where he leads the Center for Dark Silicon. His research interests include dark silicon, chip design, parallelization tools, and Bitcoin computing systems. Taylor has a PhD in electrical engineering and computer science from the Massachusetts Institute of Technology.

Steven Swanson is an associate professor in the Computer Science and Engineering Department at the University of California, San Diego, where he leads the Non-Volatile Systems Laboratory. His research interests include architectural systems and security issues surrounding nonvolatile, solid-state memories as well as dark silicon architecture and multithreading. Swanson has a PhD in computer science from the University of Washington.

Direct questions and comments about this article to Michael B. Taylor at mbtaylor@ucsd.edu or to Steven Swanson at swanson@cs.ucsd.edu.