Is Moore's Law Slowing Down? What's Next?

Lieven Eeckhout
Ghent University

We are all familiar with Moore's law, which is named after Intel cofounder Gordon Moore. He observed in 1965 that chip technology is advancing so fast that every year we can fit twice as many transistors onto a chip.[1] In 1975, he adjusted the pace of transistor doubling to every two years.[2]

The chip industry has kept Moore's prediction alive, and microprocessor manufacturers such as Intel, AMD, and IBM, to name a few, have devised various microprocessor optimizations to turn the continual supply of extra transistors into higher performance. This has had a tremendous impact on the computer industry and society at large. The continual shrinking of transistors has helped computers become increasingly powerful, leading to numerous enhancements that have dramatically changed our daily lives. Think about smartphones, powerful Internet services, and breakthroughs in various scientific fields such as weather and climate modeling, AI, genetics, and drug design.

For many years, Moore's law went hand in hand with Dennard scaling. Robert Dennard in 1974 observed that supply voltage and current should be proportional to the linear dimensions of a transistor.[3] As a result, as transistors shrank, so did voltage and current, making power proportional to the area of the transistor. In other words, power density remained constant. Conversely, we could operate the transistors at higher frequencies at the same power. Unfortunately, in modern chip technologies we can no longer scale supply voltage without exponentially increasing leakage current. With the end of Dennard scaling, we observe that power density increases.

The current state of Moore's law seems somewhat unclear. In 2016, Intel announced that it is slowing the pace with which it will launch new chip technology nodes.[4] The gap between successive generations of chips with new, smaller transistors will widen, as it is becoming increasingly more difficult to miniaturize transistors in a cost-effective way for production. Just recently, Intel disclosed that, while it took longer for the company to make the leap, it has shrunk transistors for its 10-nanometer manufacturing at a better-than-expected level, putting Moore's law back on track, as they claim.[5]

Is Moore's law dead? Is it slowing down? Is it resting? Or is it still alive and well? The least one can say is that it seems reasonable to assume that keeping up with Moore's law has become much harder than ever before. And maybe at some point companies will have to start pushing back next-generation transistor technologies.

Is this going to affect our daily lives? It doesn't look like it will in the short term for at least some application domains. Many popular and emerging applications including mobile and wearable devices as well as medical implants do not require powerful chips. In contrast, power, energy, and cost efficiency are key in that market segment. Moreover, chip manufacturers for the mobile space are generally a few years behind in terms of shrinking transistors and new manufacturing technologies.

The answer may be (very) different at the opposite end of the computing spectrum in the large-scale datacenters and supercomputers that are crammed with the most advanced microprocessors for optimum performance. Many recent scientific breakthroughs hinge on very complex long-running calculations on some of the world's largest and most powerful supercomputers. Many of the useful things that mobile devices can do rest on the

Published by the IEEE Computer Society

computational power of modern-day datacenters.

One potential solution is to build and integrate accelerators that are specialized for particular workloads. Various well-established companies and start-ups are pursuing this road. Last year, Google announced the Tensor Processing Unit for accelerating neural network computations.[6] Nvidia released a new chip design called Tesla P100 to accelerate AI and deep learning applications.[7] Microsoft's Catapult leverages field-programmable gate arrays (FPGAs) in their datacenters to accelerate networking, security, cloud services, and AI.[8] Intel acquired FPGA manufacturer Altera last year.[9]

In any case, it is crucial for our society and industry that technology continue to innovate. If supercomputers start slowing down, this could have a tremendous long-term impact on society. Slowed-down supercomputer performance is bad news for research programs that rely on supercomputers, such as improved drug design, development of new materials for batteries, and efforts to understand climate change. It is equally bad news for big-data companies such as Google, Facebook, Microsoft, and others that rely on powerful datacenters to deliver online Internet services to several hundreds of millions or even billions of customers.

Continuing progress of supercomputing beyond the scaling of Moore's law is likely to require a comprehensive rethinking of technologies, ranging from innovative materials and devices to circuits, system architectures, programming systems, system software, and applications. This special issue, entitled "Architectures for the Post-Moore Era," includes six articles that provide a perspective on this important issue, ranging from innovations in technology all the way up to advances in application development for better scalability beyond Moore's law.

I want to wholeheartedly thank the guest editors of this special issue, Jeffrey S. Vetter (Oak Ridge National Laboratory), Erik P. DeBenedictis (Sandia National Laboratories), and Thomas M. Conte (Georgia Institute of Technology). They have done an excellent job selecting articles for this special issue. Please read their guest editorial for a detailed description of the selected articles.

The Micro Economics column in this issue is very much in line with the theme of this issue. Shane Greenstein looks at the question of whether economic growth will slow along with Moore's law. Finally, the Micro Law column by Richard Stern continues a three-part series on standardization skullduggery. This column, part two, focuses on Apple's complaint against Qualcomm.

With that, I'd like to conclude and wish you happy reading, as always.

References

1. G.E. Moore, "Cramming More Components onto Integrated Circuits," *Electronics*, vol. 38, no. 8, 1965, pp. 114–117.
2. G.E. Moore, "Moore's Law at 40," *Understanding Moore's Law: Four Decades of Innovation*, Chemical Heritage Foundation, D.C. Brock, ed., 2006, pp. 67–84.
3. R.H. Dennard et al., "Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions," *IEEE J. Solid-State Circuits*, vol. 9, no. 5, 1974, pp. 256–268.
4. T. Simonite, "Intel Puts the Brakes on Moore's Law," *MIT Technology Rev.*, 23 Mar. 2016; www.technologyreview.com/s/601102/intel-puts-the-brakes-on-moores-law.
5. D. Takahashi, "Intel: Moore's Law Isn't Slowing Down," *VentureBeat*, 28 Mar. 2017; http://venturebeat.com/2017/03/28/intel-moores-law-isnt-slowing-down.
6. K. Sato, C. Young, and D. Patterson, "An In-Depth Look at Google's First Tensor Processing Unit (TPU)," *Google Cloud Big Data and Machine Learning Blog*, 12 May 2017; http://cloud.google.com/blog/big-data/2017/05/an-in-depth-look-at-googles-first-tensor-processing-unit-tpu.
7. "Nvidia Tesla P100: Infinite Compute Power for the Modern Data Center—Technical Overview," Nvidia, 2016; http://images.nvidia.com/content/tesla/pdf/nvidia-teslap100-techoverview.pdf.
8. A.M. Caulfield et al., "A Cloud-Scale Acceleration Architecture," *Proc. 49th Ann. IEEE/ACM Int'l Symp. Microarchitecture* (MICRO), 2016; doi:10.1109/MICRO.2016.7783710.
9. "Intel Completes Acquisition of Altera," press release, 28 Dec. 2015; http://newsroom.intel.com/news-releases/intel-completes-acquisition-of-altera.

Lieven Eeckhout is a professor in the Department of Electronics and Information Systems at Ghent University. Contact him at lieven.eeckhout@ugent.be.