ISPD '20, September 20–23, 2020, Taipei, Taiwan
Proceedings published March 29, 2020

Scalable System and Silicon Architectures to Handle the Workloads of the Post-Moore Era

Ivo Bolsens
Senior Vice President & Chief Technology Officer, Xilinx
San Jose, CA, USA

ivo@xilinx.com

ABSTRACT

The end of Moore's law has been proclaimed on many occasions and it's probably safe to say that we are now working in the post-Moore era. But no one is ready to slow down just yet. We can view Gordon Moore's observation on transistor densification as just one aspect of a longer-term underlying technological trend – the Law of Accelerating Returns articulated by Kurzweil. Arguably, companies became somewhat complacent in the Moore era, happy to settle for the gains brought by each new process node. Although we can expect scaling to continue, albeit at a slower pace, the end of Moore's Law delivers a stronger incentive to push other trends of technology progress harder.

Some exciting new technologies are now emerging such as multi-chip 3D integration and the introduction of new technologies such as storage-class memory and silicon photonics. Moreover, we are also entering a golden age of computer architecture innovation.

One of the key drivers is the pursuit of domain-specific architectures as proclaimed by Turing award winners John Hennessy and David Patterson. A good example is the Xilinx's AI Engine, one of the important features of the Versal™ ACAP (adaptive compute acceleration platform) [1].

Today, the explosion of AI workloads is one of the most powerful drivers shifting our attention to find faster ways of moving data into, across, and out of accelerators. Features such as massive parallel processing elements, the use of domain specific accelerators, the dense interconnect between distributed on-chip memories and processing elements, are examples of the ways chip makers are looking beyond scaling to achieve next-generation performance gains.

Next, the growing demands of scaling-out hyperscale datacenter applications drive much of the new architecture developments. Given a high diversification of workloads that invoke massive compute and data movement, datacenter architectures are moving away from rigid CPU-centric structures and instead prioritize adaptability and configurability to optimize resources such as memory and connectivity of accelerators assigned to individual workloads. There is no longer a single figure of merit. It's not all about Tera-OPS. Other metrics such as transfers-per-second and latency come to the fore as demands become more real-time; autonomous vehicles being an obvious and important example.

Moreover, the transition to 5G will result in solutions that operate across the traditional boundaries between the cloud and edge and embedded platforms that are obviously power-conscious and cost-sensitive. Future workloads will require agile software flows that accommodate the spread of functions across edge and cloud.

Another industry megatrend that will drive technology requirements especially in encryption, data storage and communication, is Blockchain. To some, it may already have a bad reputation, tarnished by association with the anarchy of cryptocurrency, but it will be more widely relevant than many of us realize. Who could have foreseen the development of today's Internet when ARPANET first appeared as a simple platform for distributed computing and sending email? Through projects such as the open-source Hyperledger, Blockchain technology could be game-changing as a platform for building trust in transactions executed over the Internet. We may soon be talking in terms of the Trusted Internet.

The predictability of Moore's law may have become rather too comfortable and slow. The future requires maximizing the flexibility, agility, and efficiency of new technologies.

With Moore's Law now mostly behind us, new adaptable and scalable architectures will allow us to further provide exponential return from technology in order to create a more adaptable and intelligent world.

CCS Concepts/ACM Classifiers

- **Computer systems organization~Parallel architectures**
- **Hardware~Reconfigurable logic and FPGAs**
- **Hardware~Very large scale integration design**

Author Keywords

Artificial intelligence, domain specific architectures, scale-out computing, adaptable compute acceleration platform

ISPD '20, March 29–April 1, 2020, Taipei, Taiwan.

ACM ISBN 978-1-4503-7091-2/20/03.
DOI: https://doi.org/10.1145/3372780.3378166

ISPD '20, September 20–23, 2020, Taipei, Taiwan
Proceedings published March 29, 2020

BIOGRAPHY

Ivo Bolsens is senior vice president and chief technology officer at Xilinx, with responsibility for advanced technology development and Xilinx research laboratories. Before he was vice-president embedded systems at IMEC, Belgium. He received his PhD in Electrical Engineering from the Catholic University of Leuven in Belgium.

REFERENCES

[1] Kees Vissers, Versal : The Xilinx Adaptive Compute Platform, Proceedings of the 2019 International Symposium on Field Programmable Gate Arrays