



# Error-Resilient Server Ecosystems for Edge and Cloud Datacenters

**Georgios Karakonstantis and Dimitrios S. Nikolopoulos,**  
Queen's University Belfast

**Dimitris Gizopoulos,** National and Kapodistrian University of Athens

**Pedro Trancoso and Yiannakis Sazeides,** University of Cyprus

**Christos D. Antonopoulos,** University of Thessaly

**Srikumar Venugopal,** IBM Research

**Shidhartha Das,** ARM Research

*The explosive growth of Internet-connected devices that form the Internet of Things and the flood of data they yield require new energy-efficient and error-resilient hardware and software server stacks for next-generation cloud and edge datacenters.*

In the past few decades, aggressive miniaturization of semiconductor circuits has driven the explosive performance improvement of digital systems that have radically reshaped the way we work, entertain, and communicate. At the same time, new

that are magnified as circuits get smaller.

This can result in many fabricated chips not meeting their intended performance and power specifications, thereby endangering the correct functionality of products that use them.

paradigms such as cloud and edge computing, as well as the Internet of Things (IoT), enable billions of devices to interconnect intelligently. These devices will generate huge volumes of data that must be processed and analyzed in centralized or decentralized datacenters located close to users. The analysis of this data could lead to new scientific discoveries and new applications that will improve our lives.<sup>1</sup>

Such advances are at risk, however, because ongoing technology miniaturization appears to be ending, as it frequently causes otherwise-identical nanoscale circuits to exhibit different performance or power-consumption behaviors, even though they're designed using the same processes and architecture (see Figure 1). Such variations are caused by imperfections in the manufacturing process



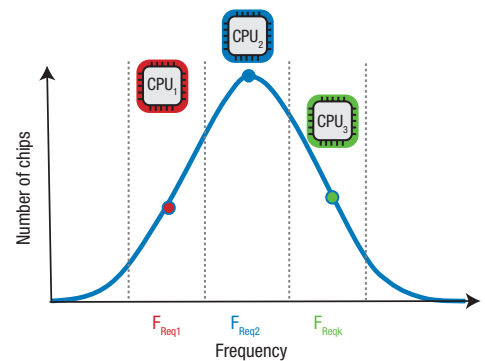
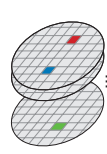
Manufacturers try to deal with the huge performance and power variability in fabricated chips—and hide it from the software layers—by adopting pessimistic safety timing margins and redundant error-correction schemes designed to counteract the worst possible scenario.<sup>2</sup> Typically, such measures are extremely pessimistic because they're based on rare worst-case operating conditions and because the capabilities of the worst performing chips are far inferior to those of the vast majority of identically manufactured circuits. As a result, the majority of the chips are constrained to operate at the low speed and high power consumption of the relatively few worst-case chips, not at the speed and power they could actually achieve.<sup>3</sup>

The use of pessimistic timing margins—along with the inability to save power efficiently by scaling down the supply voltage in nanoscale circuits because that would make circuits even more prone to failure—has elevated energy efficiency's significance.<sup>4</sup>

Reducing processors' energy consumption could not only enable products to meet their tight power budgets but could also let users improve performance by employing more resources or by operating the chips at a higher frequency.<sup>5</sup> This would be particularly important for servers, which will soon have to handle the huge amounts of data that the increasing number of interconnected devices will generate, a total estimated to reach 24.3 exabytes per month in 2019.<sup>6</sup>

## IMPROVEMENTS REQUIRE NEW DESIGN APPROACHES

Substantially improving energy-efficiency requires new types of error-resilient server ecosystems that can handle hardware components' increased power and performance variability more intelligently than conventional pessimistic paradigms.



**Figure 1.** Identical chips, in this case CPUs, might have substantially different performance characteristics, such as operating frequency (Freq), even if they were designed and manufactured using the same processes.

The computing industry should see such heterogeneity not as a problem but as an opportunity to improve energy efficiency. This could be done by not artificially constraining all chips' performance based on a few outliers but rather by letting each chip operate according to its true capabilities. Exploiting such heterogeneity requires a shift away from current approaches and a redesign of next-generation servers' hardware and system software.

## SERVER ECOSYSTEM

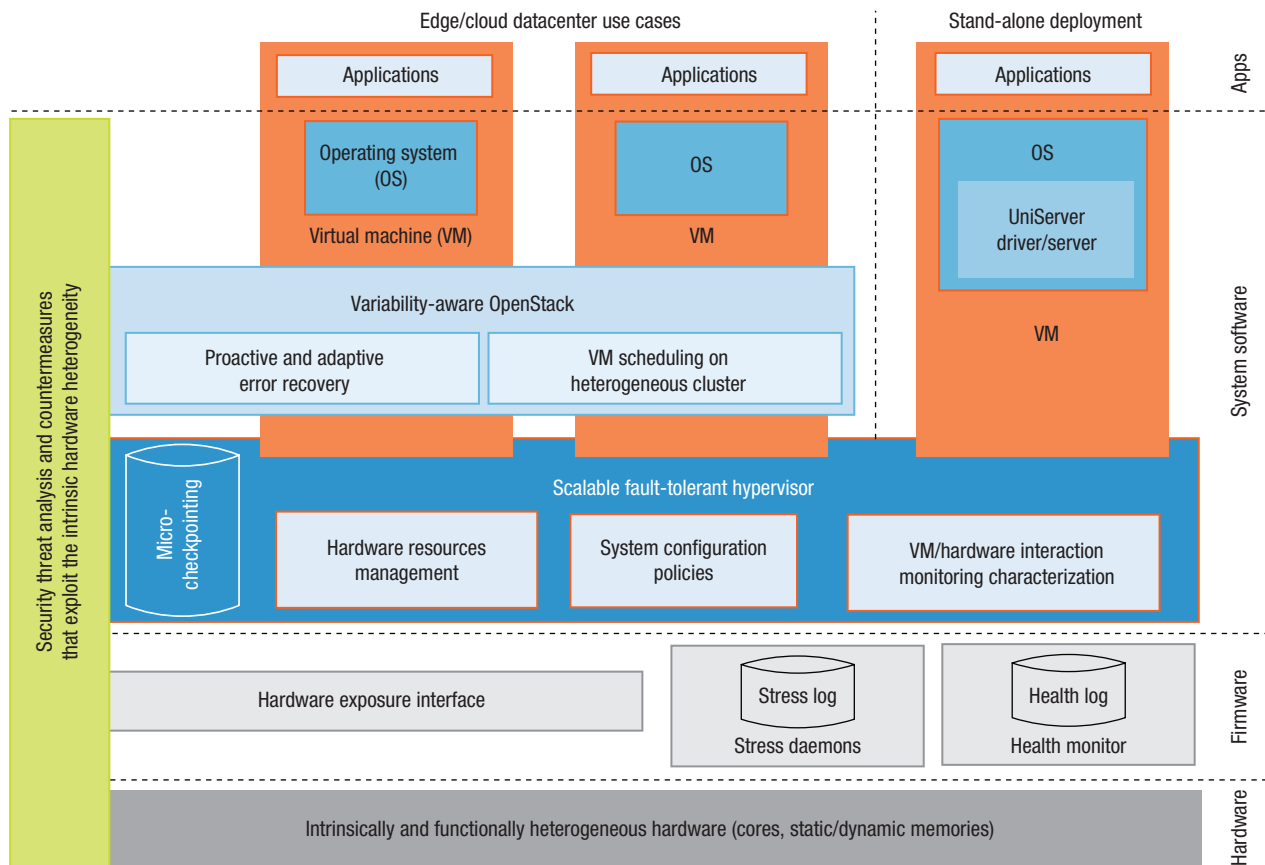
Making the most of heterogeneity requires automated firmware-level procedures to expose each processor's and memory resource's capabilities, even as they change over time. This requires the embedding of diagnostic and health-monitoring daemons in the firmware of any server that evaluates hardware components' operations during the product lifetime (see Figure 2). These daemons would access on-chip sensors and error-detection circuitry to collect and analyze various parameters, such as correctable and uncorrectable errors, performance counters, system crashes and hangs, and thermal and power behavior. This would be similar to the

procedures that the machine-check architecture in x86 systems has adopted ([www.mcelog.org](http://www.mcelog.org)).

The new system would use a hardware exposure interface (HEI) enhanced to collect the required information and communicate it to the software stack, which would identify energy-efficient voltage, frequency, and refresh-rate states for processors and memory subsystems.

We should also rethink the design of all system-software layers—including hypervisors and resource-management frameworks such as OpenStack—used in today's datacenters. These layers should be able to operate hardware close to its performance and power limits. For example, hypervisors could use this capability to allocate processor and memory resources with different reliability, power, and performance-efficiency characteristics to virtual machines so that they improve performance while reducing the server's energy consumption. Cloud-management frameworks could leverage the same capability to provide better quality of service (QoS) to users and applications.

However, operating hardware outside its normal safety margins could introduce critical system-software



**Figure 2.** This server ecosystem spans all layers of the system stack and enhances it with technologies for monitoring hardware health and optimizing overall operation under an extended range of operating points.

errors and even crash servers. We need system-software-resilience approaches to avoid or mitigate such errors and sustain high server availability. In datacenters, this means not degrading the QoS of a customer workload in case of a server crash and adopting mechanisms for speeding up the recovery process.

## EMPOWERING INTERNET EVOLUTION

Manufacturers could integrate our proposed ecosystem into standard high-end servers and the newly introduced microservers. Microservers don't perform as well as mainstream servers yet, but they can service many types of application requests with an appropriate level of performance and with significantly less power consumption.

Integrating our proposed software and hardware ecosystem into servers would help power next-generation datacenters in the cloud and at the network edge, where energy efficiency is particularly critical for minimizing power supply, cooling, and maintenance costs.

Energy-efficient servers would help create a more sustainable Internet. Presently, most Internet processing and storage takes place in the cloud, in massive centralized datacenters that contain tens of thousands of servers, consume as much electricity as a small city, and utilize expensive cooling mechanisms. This won't be practical in the IoT era because the current Internet infrastructure's limited network capacity won't accommodate the exabytes of data that Internet-connected

devices will soon generate. However, using the typical centralized datacenters along with new decentralized datacenters at the network edge, closer to users, could limit the load put on the Internet infrastructure by allowing preprocessing and selective forwarding of data to the cloud. This paradigm, used by both edge and fog computing, has advantages over the cloud paradigm and is being promoted by major companies such as Cisco, Huawei, IBM, and Intel as a way to transform the next-generation Internet.<sup>7</sup>

Finally, edge resources' ability to provide all necessary services within a home or small business improves privacy because the data they carry doesn't have to travel through the public network or reside in third-party datacenters.

**R**ealizing our proposed error-resilient, energy-efficient ecosystem faces many challenges, in part because it requires the design of new technologies and the adoption of a system operation philosophy that departs from the current pessimistic one.

The UniServer Consortium ([www.uniserver2020.eu](http://www.uniserver2020.eu))—consisting of academic institutions and leading companies such as AppliedMicro Circuits, ARM, and IBM—is working toward such a vision. Its goal is the development of a universal system architecture and software ecosystem for servers used for cloud- and edge-based datacenters. The European Community's Horizon 2020 research program is funding UniServer (grant no. 688540).

The consortium is already implementing our proposed ecosystem in a state-of-the-art X-Gene2 eight-core, ARMv8-based microserver with 28-nm feature sizes. The initial characterization of the server's processing cores shows that there is a significant safety margin in the supply voltage used to operate each core. Results show that the some cores could use 10 percent below the nominal supply voltage that the manufacturer advises. This could lead to a 38 percent power savings.<sup>8</sup>

Similarly promising is the characterization of the DRAM memories used on the ARMv8-based microserver showing that the refresh rate and supply voltage could be decreased by 98 percent and 5 percent from nominal levels, respectively. Such reductions could lead to an average power savings of more than 22 percent across a range of benchmarks.<sup>9</sup> ■

## REFERENCES

1. H. Bauer, M. Patel, and J. Veira, *The Internet of Things: Sizing up the Opportunity*, online report, McKinsey & Co., December 2014; [www.mckinsey.com/industries/semiconductors/our-insights/the-internet-of-things-sizing-up-the-opportunity](http://www.mckinsey.com/industries/semiconductors/our-insights/the-internet-of-things-sizing-up-the-opportunity).
2. S. Ghosh and K. Roy, "Parameter Variation Tolerance and Error Resiliency: New Design Paradigm for the

Nanoscale Era," *Proc. IEEE*, volume 98, no. 10, 2010, pp. 1718–1751.

3. P.N. Whatmough et al., "An All-Digital Power-Delivery Monitor for Analysis of a 28nm Dual-Core ARM Cortex-A57 cluster," *Proc. 2015 IEEE Solid-State Circuits Conf. (ISSCC 15)*, 2015, pp. 1–3.
4. H. Wong et al., "Implications of Historical Trends in the Electrical Efficiency of Computing," *IEEE Annals of the History of Computing*, vol. 33, no. 3, 2011, pp. 46–54.
5. H. Esmaeilzadeh et al., "Dark Silicon and the End of Multicore Scaling," *IEEE Micro*, vol. 32, no. 3, 2012, pp. 122–134.
6. Cisco Visual Networking Index: *Global Mobile Data Traffic Forecast Update, 2016–2021*, online white paper, Cisco Systems, March 2017; [www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html](http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html).
7. W. Shi et al., "Edge Computing: Vision and Challenges," *IEEE Internet of Things J.*, vol. 3, no. 5, 2016, pp. 637–646.
8. G. Papadimitriou et al., "Harnessing Voltage Margins for Energy Efficiency in Multicore CPUs," *Proc. 50th IEEE/ACM Int'l Symp. Microarchitecture (MICRO 17)*, 2017, pp. 503–516.
9. K. Tovletoglou, D. Nikolopoulos, and G. Karakonstantis, "Relaxing DRAM Refresh Rate through Access Pattern Scheduling: A Case Study on Stencil-Based Algorithms," *Proc. 23rd IEEE Int'l Symp. Online Testing and Robust System Design (IOLTS 17)*, 2017, pp. 45–50.

## GEORGIOS KARAKONSTANTIS

is an assistant professor in the School of Electronics, Electrical Engineering, and Computer Science (EECS) at Queen's University Belfast and the scientific coordinator of the UniServer project. Contact him at [g.karakonstantis@qub.ac.uk](mailto:g.karakonstantis@qub.ac.uk).

## DIMITRIOS S. NIKOLOPOULOS

is a professor and the head of the School of EECS at Queen's University Belfast. Contact him at [d.nikolopoulos@qub.ac.uk](mailto:d.nikolopoulos@qub.ac.uk).

**DIMITRIS GIZOPOULOS** is a professor in the Department of Informatics and Telecommunications at the National and Kapodistrian University of Athens, where he leads the Computer Architecture Laboratory. Contact him at [dgizop@di.uoa.gr](mailto:dgizop@di.uoa.gr).

**PEDRO TRANCOSO** is an associate professor in the University of Cyprus' Department of Computer Science. Contact him at [pedro@cs.ucy.ac.cy](mailto:pedro@cs.ucy.ac.cy).

**YIANNAKIS SAZEIDES** is an associate professor in the University of Cyprus' Department of Computer Science. Contact him at [yanos@cs.ucy.ac.cy](mailto:yanos@cs.ucy.ac.cy).

## CHRISTOS D. ANTONOPOULOS

is an assistant professor in the University of Thessaly's Electrical and Computer Engineering Department. Contact him at [cda@uth.gr](mailto:cda@uth.gr).

**SRIKUMAR VENUGOPAL** is a research scientist at IBM Research—Ireland. Contact him at [srikumarv@ie.ibm.com](mailto:srikumarv@ie.ibm.com).

**SHIDHARTHA DAS** is a principal research engineer at ARM Research and a Royal Academy of Engineering visiting professor at Newcastle University. Contact him at [sdas@arm.com](mailto:sdas@arm.com).

*This article originally appeared in Computer, vol. 50, no. 12, 2017.*