

# Nanocomputing: Small Devices, Large Dependability Challenges

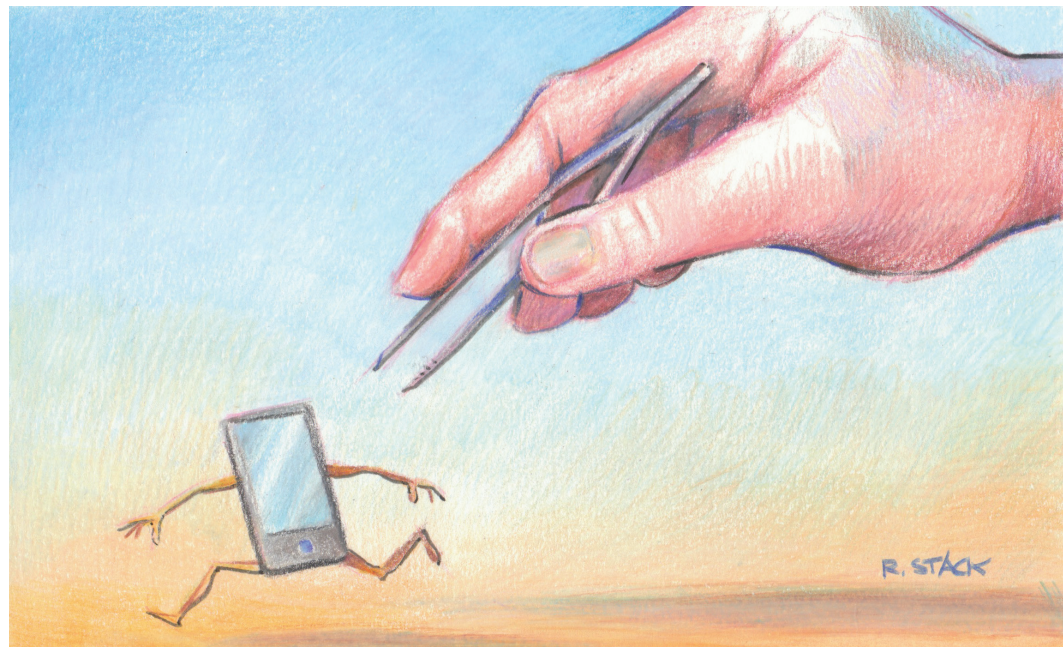
Jean Arlat | LAAS-CNRS

Zbigniew Kalbarczyk | University of Illinois

Takashi Nanya | Canon

**N**anoscale digitization will be an essential lever to foster the emerging cyberphysical systems. Thanks to the widespread presence of IT and communication capabilities far beyond today's Internet and wireless-networking capabilities, future highly pervasive embedded systems will feature smart objects (sensors and actuators) fully merged with the environment in which they're deployed. This will result in enhanced ambient services spanning the everyday life of citizens, thus improving our quality of life, increasing our awareness of resources and the environment, and enriching the user experience.

Silicon technologies' evolution toward nanoscale dimensions (for more on this, see the sidebar "Nanotechnology and Moore's Law") raises serious challenges regarding dependability and security. In particular, such concerns motivate the dependable-computing and fault-tolerance community. By exploiting large-scale integration, we might expect fault-tolerance techniques to offset the limitations of the currently dominant fault-avoidance approaches. Several workshops attest to interest in this area—for example, the Workshop on Dependable and Secure Nanocomputing, the Silicon Errors in Logic—System Effects workshop, and the workshop on hardware



issues held in July 2011 by International Federation for Information Processing Working Group 10.4.

Research in this area involves three main concerns:

- the unreliability and variability characterizing nanoscale-device production,
- accidental disturbances that affect the systems' operation, and
- malicious threats targeting hardware circuit vulnerabilities.

Here, we briefly examine these concerns and their related challenges and offer some perspectives on how these challenges might be met.

## Coping with Massively Defective Devices

In the near future, chips might well incorporate several hundred billions of smaller and smaller transistors. However, this extreme downsizing would result in atomic range dimensions and thus in interdevice and intradevice variability. So, nanoscale electronic devices will become inherently unreliable and unpredictable.<sup>1</sup> The shrinking of the layout geometry will cause manufacturing defects to have a severer impact, creating more error-prone devices.

Research addressing these issues for memory chips has already made significant progress.

## Nanotechnology and Moore's Law

Following Moore's law, the trend toward nanodevices has been continuous: since the early '70s, the number of transistors per die has doubled every couple of years. Beyond this "more Moore" track, which focuses mainly on CMOS device miniaturization, the industry is contemplating a new track, often called "more than Moore." In this track, devices add value by incorporating functionalities that don't necessarily scale according to Moore's law.

Besides this top-down approach, the "beyond Moore" track provides a bottom-up approach for nanoscale computing by departing from classical silicon technologies. This approach features atomic assemblies of nanoscale technologies, encompassing nanowires, carbon nanotubes or organic molecules, and so on, and extending to quantum, optical, and microfluidic and nanofluidic devices. However, these technologies haven't reached the same maturity as CMOS technology. They intrinsically suffer from a significantly high rate of residual defects and fault occurrences. This situation is somewhat reminiscent of digital computers' early days and of the seminal research on improving their reliability by John von Neumann, Claude Shannon, and others. So, in the main article, we concentrate on the "more Moore" track.

Most advanced techniques provide spare elements (connection lines, redundant registers, and memory elements) to dynamically replace defective elements. Proposed techniques cope with not only production defects but also runtime faults. Such techniques aim primarily to achieve high yield, which might require significant overhead.

Researchers have also proposed efficient fault-tolerance techniques for processor chips. For example, Toru Nakura and his colleagues' technique employs a set of fragmented microprocessor units for which redundant fragments are available.<sup>2</sup> Another important issue concerns control logic in processors, which is growing in size and complexity and is basically unprotected. The generalization of multi-core architectures, and potentially another layer of control, could well exacerbate this problem.

### Susceptibility to Transient Disturbances

Another effect of technology scaling is the decrease in energy consumed for each information bit. Consequently, transient disturbances (for example, soft errors)

are increasingly affecting computing systems, owing to the impact of ground-based radiation. This problem, well known in space applications and becoming so in avionics, will likely also affect medical electronics, cell phones, and automotive systems.

In particular, this problem will only worsen and create substantial challenges for designers of automotive electronics who are considering programmable-logic devices, such as field-programmable gate arrays (FPGAs), as a flexible, low-cost solution for their next-generation designs. Such disturbances could already significantly affect mature technologies. Consider a 22- $\mu\text{m}$  SRAM-based FPGA featuring 1-million-gate chips. A simulation using Space Radiation 4.5 ([www.spacerad.com](http://www.spacerad.com)) and assuming operation at a 5,000-foot altitude (for example, Denver) predicts  $1.05 \times 10^{-4}$  upsets per day.<sup>3</sup> A fleet of 500,000 vehicles, each featuring an airbag control system using this technology, would incur 52.5 upsets per day (assuming continuous use). Even a more modest usage profile of one hour per day would still lead to approximately two upsets per day.

Car developers certainly shouldn't ignore that figure!

Solutions for hardened technologies exist and have seen intensive use. However, the high (and often excessive) cost of the fabrication lines for such solutions has restricted their use. So, we increasingly must rely on various fault-tolerance techniques. In particular, Naga Avirneni and his colleagues have proposed efficient, cost-effective mechanisms based on replicated flip-flops (basic electronic circuits to store information).<sup>4</sup> Besides external disturbances, an increasingly important issue is the impact of power, current, and voltage fluctuations.<sup>5</sup>

### Exploiting Hardware Vulnerabilities

Modern integrated circuits (ICs) also increasingly face risks related to hacking and malicious threats. The circuits most exposed to attacks are smart card chips and cryptoprocessors in e-commerce or banking transactions. Devices can be compromised through a variety of methods; among the prevalent ones are side-channel attacks (for example, differential power analysis or electromagnetic analysis).

Embedded-testing support devices (such as scan chains), which aim to obtain high controllability and observability for test engineers, also constitute a security weakness. Attackers can use the scan chain architecture's properties for other kinds of side-channel attacks through malevolent fault injections exploiting the related information leakage. The likelihood of a successful attack depends on both the implementation's leakage and the hacker's skill.

To circumvent such attacks, researchers have proposed enhanced mechanisms beyond the classic tamper-resistant designs or irreversible disconnections. These mechanisms are based on

asynchronous-logic designs, signature checks, or a mix of reliability and security mechanisms. Regarding this last method, researchers have recently explored application-specific techniques to provide customizable levels of reliability and security. For example, Ravishankar Iyer and his colleagues introduced the concept of application-aware security and reliability, which customizes the error-checking mechanisms on the basis of knowledge of the application behavior.<sup>6</sup> This approach extends to multicore processors or virtualized environments in which one core or virtual machine (VM) can be explicitly dedicated to monitor applications executing on the remaining cores or VMs.

In addition, the evolution of systems embedded in critical applications (for example, commercial aircraft and spacecraft) is raising acute security problems. The systems could also impair safety requirements, particularly via their increased openness due to networking capabilities. So, in this domain, we must apply or adapt security and dependability techniques that have proven their efficiency in other contexts. A typical example is the protection scheme that relies on virtualization to execute redundant diversified operating systems to address both safety and security issues in aircraft maintenance laptops.<sup>7</sup>

As in accidental-fault-tolerant computing, in which error-handling mechanisms usually reside at the level where most faults develop and errors manifest themselves, defenses here should target the low levels where most vulnerabilities lie. Examples include basic OS kernel implementation of address space protection and hardware implementation of protection mechanisms that attackers could exploit.<sup>8</sup>

Whether the shrinking of device sizes makes coping with side-channel attacks easier isn't

fully established. The smaller feature sizes make alternative design approaches—for example, using specific (dual-rail) logic to reduce information leakage—cost effective. Also, the lower strength of on-chip signals means that hackers must have more skills or sophisticated tools to break the devices. However, Mathieu Renaud and his colleagues recently showed that the increased variability of emerging technologies might jeopardize protection schemes based on the traditionally assumed leakage models.<sup>9</sup>

### The Way Forward

The commercial availability of many-core chips (which feature several tens of cores) paves the way for processor design that accommodates both increased processing performance requirements and the lowest-possible energy footprint. Targeted applications include tablet PCs, smartphones, and high-performance-computing boards. Among the most recent examples is the forthcoming Adapteva 64-core chip, which will rely on a 28-nm process.<sup>10</sup> Accordingly, the concerns we raised earlier regarding such nanoscale technology will have to be dealt with.

To deal with those concerns, a pragmatic and potentially cost-effective approach would be to use the valid (that is, correctly functioning) cores to ensure chip resiliency (that is, allowing the chip to operate in spite of failed cores). In some cases, manufacturers have sorted processor batches according to their frequency (1.6 GHz, 1.8 GHz, or 2 GHz). Similarly, they could sort manufactured chips according to the achieved MIPS (million instructions per second)—for example, as a function of the number of valid cores. Several manufacturers are applying this principle; for example, Intel “recycles” Core Duo chips featuring a defective device as Core Solo

chips (see [http://en.wikipedia.org/wiki/Intel\\_Core](http://en.wikipedia.org/wiki/Intel_Core)).

As Jacques Collet and his colleagues proposed, a more ambitious approach would apply such a configuration dynamically at runtime rather than simply at production time.<sup>11</sup> This way, you could keep using the available cores on a multicore chip, even when operational faults impair additional cores.

**W**ith the advent of nanoscale low-power devices, cyber-physical systems are becoming pervasive in many application domains and impacting citizens' everyday activities. However, shrinking device dimensions increase the devices' sensitivity to accidental faults and malicious tampering, thus raising concern for the dependable, secure deployment of computerized systems.

So, chip manufacturers, system integrators, and end users must depart from the long-time assumed “fault free or correct” to a more pragmatic “fault prone or vulnerable” point of view on hardware chip design and operation. (For a rationale concerning the IC manufacturer's viewpoint, see the second sidebar.)

To achieve dependable, secure

## IEEE micro Calls for Papers

IEEE Micro seeks general-interest submissions for publication in upcoming issues. These works should discuss the design, performance, or application of microcomputer and microprocessor systems. Of special interest are articles on performance evaluation and workload characterization. Summaries of work in progress and descriptions of recently completed works are most welcome, as are tutorials. IEEE Micro does not accept previously published material.

Visit our author center ([www.computer.org/mc/micro/author.htm](http://www.computer.org/mc/micro/author.htm)) for word, figure, and reference limits. All submissions pass through peer review consistent with other professional-level technical publications, and editing for clarity, readability, and conciseness. Contact IEEE Micro at [micro-ma@computer.org](mailto:micro-ma@computer.org) with any questions.

[www.computer.org/micro/cfp](http://www.computer.org/micro/cfp)



## The Inevitable Shift in Semiconductor Manufacturers' Vision

The growth of many-core chips appears to support semiconductor manufacturers' vision for a shift to the classic IC development paradigm. For example, the 2009 edition of the *International Technology Roadmap for Semiconductors* made this statement:<sup>1</sup>

Relaxing the requirement of 100% correctness for devices and interconnects may dramatically reduce costs of manufacturing, verification, and test. Such a paradigm shift will likely be forced in any case by technology scaling, which leads to more transient and permanent failures of signals, logic values, devices, and interconnects. ... In general, automatic insertion of robustness into the design will become a priority as systems become too large to be functionally tested at manufacturing exit. Potential solutions include automatic introduction of redundant logic and on-chip reconfigurability for fault tolerance, development of adaptive and self-correcting or self-healing circuits, and software-based fault tolerance.

### Reference

1. "Design," *Int'l Technology Roadmap for Semiconductors*, ITRS, 2009; [www.itrs.net/Links/2009ITRS/2009Chapters\\_2009Tables/2009\\_Design.pdf](http://www.itrs.net/Links/2009ITRS/2009Chapters_2009Tables/2009_Design.pdf).

behavior, hardware-level faults and attacks must be tackled both at the hardware layer and through software-level protection and reconfiguration mechanisms. ■

### References


1. W. Haensch et al., "Silicon CMOS Devices beyond Scaling," *IBM J. Research & Development*, vol. 50, nos. 4–5, 2006, pp. 339–361.
2. T. Nakura, K. Nose, and M. Mizuno, "Fine-Grain Redundant Logic Using Defect-Prediction Flip-Flops," *Digest of Technical Papers, IEEE Int'l Solid-State Circuits Conf. (ISSCC 07)*, IEEE Press, 2007, pp. 402–403 & 611.
3. M. Mason, "Cosmic Rays Damage Automotive Electronics," *EE Times*, 31 May 2006; [www.eetimes.com/design/automotive-design/4011077/Cosmic-rays-damage-automotive-electronics](http://www.eetimes.com/design/automotive-design/4011077/Cosmic-rays-damage-automotive-electronics).
4. N.D.P. Avirneni, V. Subramanian, and A.K. Somani, "Low Overhead Soft Error Mitigation Techniques for High-Performance and Aggressive Systems," *Proc. 2009 IEEE/IFIP Conf. Dependable Systems & Networks (DSN 09)*, IEEE Press, 2009, pp. 185–194.
5. C. Constantinescu, "Impact of Intermittent Faults on Nanocomputing Devices," *Proc. 2007 IEEE/IFIP Conf. Dependable Systems & Networks (DSN 07)*, supplemental vol., IEEE Press, 2007, pp. 238–241.
6. R. Iyer et al., "Toward Application-Aware Security and Reliability," *IEEE Security & Privacy*, vol. 5, no. 1, 2007, pp. 64–69.
7. Y. Laarouchi et al., "Connecting Commercial Computers to Avionics Systems," *Proc. 2009 IEEE/AIAA Digital Avionics Systems Conf. (DASC 09)*, IEEE Press, 2009, pp. 6.D.1-1–6.D.1-9.
8. F. Lone Sang et al., "Exploiting an I/OMMU Vulnerability," *Proc. 5th Int'l Conf. Malicious and Unwanted Software (Malware 10)*, IEEE Press, 2010, pp. 9–16.
9. M. Renaud et al., "A Formal Study of Power Variability Issues and Side-Channel Attacks for Nanoscale Devices," *Advances in Cryptology—EUROCRYPT 2011*, LNCS 6632, Springer, 2011, pp. 109–128.
10. M. Feldman, "Adapteva Builds Manycore Processor That Will Deliver 70 Gigaflops/Watt," *HPCwire*, 3 Oct. 2011; [www.hpcwire.com/hpcwire/2011-10-03/adapteva\\_builds\\_manycore\\_processor\\_that\\_will\\_deliver\\_70\\_gigaflops\\_watt.html](http://www.hpcwire.com/hpcwire/2011-10-03/adapteva_builds_manycore_processor_that_will_deliver_70_gigaflops_watt.html).
11. J.H. Collet et al., "Chip

Self-Organization and Fault Tolerance in Massively Defective Multicore Array," *IEEE Trans. Dependable and Secure Computing*, vol. 8, no. 2, 2011, pp. 207–217.

**Jean Arlat** is a director of research at LAAS-CNRS (the Laboratory for Analysis and Architecture of Systems of the French National Center for Scientific Research) and is affiliated with the University of Toulouse. He's a member of the laboratory's research group on dependable computing and fault tolerance. Contact him at [jean.arlat@laas.fr](mailto:jean.arlat@laas.fr).

**Zbigniew Kalbarczyk** is a research professor of design and evaluation of dependable and secure systems and applications at the University of Illinois Coordinated Science Laboratory. Contact him at [kalbarcz@illinois.edu](mailto:kalbarcz@illinois.edu).

**Takashi Nanya** is an adviser on research and development of dependable-computing technology at Canon. Contact him at [nanya.takashi@canon.co.jp](mailto:nanya.takashi@canon.co.jp).

 Selected CS articles and columns are also available for free at <http://ComputingNow.computer.org>.