PANEL: Nano, Quantum, and Molecular Computing: Are we Ready for the Validation and Test Challenges?

Sandeep K. Shukla [*]   Ramesh Karri [♠]   Seth Copen Goldstein [♣]   Forrest Brewer [†]

Kaustav Banerjee [†]

Sankar Basu [◇]

[*]Bradley School of Electrical and Computer Engineering,Virginia Tech
[♠] Electrical and Computer Engineering Department , Polytechnic University
[♣] School of Computer Science, Carnegie Mellon University
[†] Department of Electrical and Computer Engineering, University of California at Santa Barbara
[◇] Design Automation of Micro and Nano Systems, National Science Foundation

Abstract

In the recent years a lot of research effort is being spent in the areas of nanotechnology, quantum computation, and biologically inspired computing. As we are faced with various challenges regarding their implementability, architectural visions, and design automation, not much has been done in the field of high level design and validation in looking further into the future, and ponder about the state of the art in design validation and test in such miniscule technology era. Very few reported research work have surfaced on the design and validation challenges for these technologies. However, this certainly is a matter of concern because the technology of the small will be ridden with random faults and hence architectural design strategies need to change to take into account these stochastic models of failures to build robust designs. Validation of such designs also have to capture the stochastic behavioral models of the technology, and hence traditional validation and testing techniques will not work directly. Are we getting ready with our theory, technology and tools to address these challenges? This futuristic panel asks technology and computer aided design experts, as well as funding agency program managers questions about the technological barriers to be surpassed, as well as how the funding agencies such as NSF are ramping up for this technological future.

1. Introduction

With the continuing advances in the miniaturization of devices on silicon, we are already at the deep-submicron scale device manufacturing. However, nanotechnology and molecular computing, as well as quantum computing are making inroads from pure imagination to realization in research laboratories. In this miniaturized computing era, device sizes are in the range of several nanometers and hence amenable to high degree of failures, due to (i) manufacturing defects, (ii) transient faults resulting from reduced noise tolerance at reduced voltage and current levels, (iii) faults due to aging because of molecular and other kinds of techniques for creating nano-devices. Although nano-scale manufacturing will allow us to pack more devices on a chip, we have to live with these possibilities of defects in the nano-substrate. As a result, "defect-tolerant architecture" is being posed as a way to mitigate the challenge of the inherent unreliability at the nano-scale [7, 8, 10, 12]. Defect-tolerance is built into the architecture in the form of redundancy of devices, and functional units. Abundance of devices due to their feature size miniaturization makes some form of redundancy acceptable. However, it turns out that the redundancy level must be properly designed in order to obtain reliable computation from defect prone devices. We investigate ways to automatically compute these optimal redundancies that can guarantee reliable computation. Some theoretical models have been proposed in the past in [7, 10], and some results from information theory [3, 11] have been used to obtain some theoretical insight into these redundancy-reliability trade-offs. Probabilistic modeling and model-checking tools have also been used [5] to gain insight into the defect tolerant design, and their reliability measures. In 1952, von Neumann studied the problem of constructing reliable computation from unreliable devices (due to unreliable valve based computers at that time), and

he used a redundancy technique namely NAND multiplexing [7, 13]. He showed that if the failure probabilities of the gates are sufficiently small, and failures are statistically independent, computations may be done with high probability of correctness. Later in [1] it was shown that a logarithmic redundancy is necessary for some Boolean function computation, and sufficient for all Boolean functions. In [11] Pippenger showed that von Neumann's construction works only when the probability of failure per gate has a limit strictly less than $\bullet \cdot \bullet \bullet \bullet \bullet$, and that the computation in presence of noise (which can be seen as presence of defect), requires more layers of redundancy.

Given these theoretical results that did not consider nanoscopic devices but more about faulty components, one can go about building theory and tools that would allow one to evaluate various architectural options to implement defect-tolerance. However, one has to consider the scale as well. At the nano or molecular level the amount of redundancy one can obtain is huge, and hence the state space size is immense. With deep submicron technology itself, validation of a chip is not within the capacity of any existing tools, with many orders of magnitude of scaling, we need breakthroughs in validation technology for sure. Some other important questions are:

- Even if we know how to build around defects, how do we quantify defects (e.g., probability distribution) accurately and use it at architecture level to build the defect tolerance?

- What are the new technological characteristics that will affect our high level (system level or higher) conceptualization of digital systems?

- Are there any inherent information theoretic bounds that would bring forth some uncertainty elements which cannot be circumvented in these technologies?

- How will post manufacturing testing methodologies change?

2. Panelist Position Statements

2.1 Ramesh Karri

A key question that is yet unclear is what nanotechnology (ies) will be the dominant technology (ies). On one hand, researchers investigating system level architectures have rightly identified defect tolerance (manufacture time defect detection and reconfiguration) and fault-tolerance (operational time fault detection and reconfiguration, recovery etc.) as an enabling mechanism for the successful deployment of these emerging nanotechnologies. On the other hand, defect and fault models in nanotechnologies have not been thoroughly investigated. Based on this status, what should these system level fault and defect tolerant architectures be based on? Is functional defect and fault testing appropriate at this preliminary stage of understanding of nanotechnology fault models. Are techniques based on traditional stuck-at-fault models appropriate? Based on the large failure and defect rates encountered in nanotechnologies, what defect and fault detection and defect and fault tolerance mechanisms are appropriate? These and other questions will be addressed in this special session.

2.2 Seth Goldstein

Reconfigurable Nanoelectronics and Defect Tolerance: Future computing systems will inevitably be built using nanoelectronics, i.e., from devices and wires with feature sizes below thirty nanometers. The SIA roadmap [14] predicts that traditional silicon-based systems will have feature sizes of below 40nm within the decade. There are also advances being made in building computing systems using new technologies, such as molecular electronics [15]. Successfully harnessing nanoelectronics requires a rethinking of the ways in which we design, manufacture, and test computing systems. In this paper we examine how current trends in semiconductor manufacturing as well as how new technologies, e.g., molecular electronics and self-assembly make reconfigurability an inherent part of any nanoscale computing system.

Perhaps the greatest impact of the nanoscale on electronics will be the reduced ability to arbitrarily determine the placement of the components of a system. The most extreme example of this is to be found in chemically assembled electronic nanotechnology (CAEN), a form of molecular electronics which uses bottom-up assembly to construct electronic circuits out of nanometer-scale devices. Large-scale molecular electronics requires some form of self-assembly [15]. When using self-assembly, individual devices and wires are first manufactured, and only later assembled into a circuit. While self-assembly promises to be a very economical process (compared with the cost of silicon fabrication equipment), it cannot be used to create the arbitrary patterns that can be formed using photolithography. Only simple, crystal-like structures, can be created using self-assembly. Furthermore, defect densities of self-assembled circuits are projected to be orders of magnitude higher then in silicon-based devices. Thus, self-assembled circuits and architectures will have to be designed for defect tolerance.

To a lesser extent these same problems will appear as traditional semiconductor technology continues to scale. The complexity of generating a reliable mask set which produces reliable chips is already limiting the ability to create arbitrary patterns of wires. This can be seen in the trend towards "structured" ASICs, which allow custom chips to

share many of the same masks [16]. As devices scale down it is also harder to maintain constant characteristics for all the devices on a single chip [17]. Some argue that process variation will essentially eliminate the performance gains typically expected when feature sizes shrink [18].

Economics will also play a role in changing how chips are manufactured. As mask sets become more expensive it becomes more economical to reuse the same chip for different tasks, i.e., to use programmable (reconfigurable) hardware such as a field programmable gate array (FPGA). The need to decrease time-to-market is also accelerating the trend towards using FPGAs for every higher volume applications. The extremely low defect densities in today's semiconductor manufacturing come at an economic cost. In the future it may be more economical to design defect tolerant systems in order to reduce this cost.

In order to implement useful reliable functionality on top of crystal-like structures, post-fabrication customization is required; this customization will be used for two purposes (1) to implement the desired functionality and (2) to eliminate the deleterious effects of the defects [19–21]. Reconfiguration provides defect tolerance by configuring the desired circuit around the defects, thus creating a reliable system from an unreliable substrate.

Defect tolerance by reconfiguration is a two step process. First, the defects are discovered and mapped. Then, when configuring the device to implement a particular circuit, the map is consulted to avoid the defects. This approach requires three things from a reconfigurable device: it must be reprogrammable, it must have a rich fine-grained interconnect, and it should allow us to implement a particular logic function in many different ways. All three of these attributes are necessary for both defect detection and defect avoidance. During defect detection, we reprogram different test circuits on the device. Each different instance of a test structure gives us information about different sets of components on the device. The latter two attributes are most necessary during defect avoidance. They allow a particular circuit to be implemented without requiring us to use any of the defective components.

Nanoelectronics holds the promise of continuing technology scaling to feature sizes below 40nm. However, to harness the abundance of resources and new constraints concomitant with this feature size requires reconfigurability at the core of computing systems.

2.3 Kaustav Banerjee

The continuous trend in CMOS scaling towards the 10 nm node, as seen in the International Technology Roadmap for Semiconductors [9], is inviting concerns about our long-term demands for small, ultra-fast, ultra low-power, reliable devices. This evokes a serious thinking about the role

that nanoelectronics and non-traditional technologies can play in future electronic systems. Along with nano-scale CMOS, several new nano-technologies have emerged, with each of them having its own merits and demerits in terms of performance and manufacturability [22]. However, till date, only few nanotechnologies, like single electron transistors, carbon nanotubes, molecular electronics and spin transistors have shown some potential for information processing and/or data storage [22, 24]. Therefore, it seems unlikely that any of the emerging nanotechnologies on the horizon will completely replace CMOS. Hence, hybridization of CMOS technology with other (compatible and relatively mature) nanotechnologies might provide an attractive option for overcoming some of the inherent limitations of nano-scale CMOS technologies [23, 25, 26]. Hence, these CMOS-nano hybrid designs must also be evaluated from a circuit and system perspective, in terms of the four key metrics: power, performance, robustness and reliability. Since, all nanoscale technologies are expected to suffer from significant parameter variations and high defect densities, understanding should be developed about how these metrics are going to be affected in presence of such variations and defects. This will be essential for developing techniques and algorithms for achieving design optimization of nanoscale circuits and systems.

From the point of view of high level design validation and testing of systems built upon such purely-nano or hybrid CMOS-nano technologies, there are two major concerns that need to be addresed: validation or verification of large system and completeness of the verification strategy. Any nanoscale integration of devices will certainly give rise to unexpectedly high functional density of components. Hence, it is important to make sure that the high level design strategy is capable of verifying a nano-system with large number of functional components integrated into it. It is instructive to note that formal verification may not be feasible for such a system with even larger number of components than those found in traditional CMOS designs. On the other hand, simulation will not be sufficient for providing complete coverage of the system, nor for understanding of some intrinsic faults (for example, impact of charge-redistribution on dynamic logic gate performance or delay of adder) at the circuit level. Hence, a pertinent quest should be for a validation strategy that is essentially a trade-off between simulation and formal verification techniques. Moreover, the test strategy should encompass ways to address the entire possible range of defects/faults for a given nanotechnology. Furthermore, it is important to understand that the circuit or system level manifestation of the nano-scale defects is going to be different for different nanotechnologies. Therefore, each nanotechnology must be modeled in terms of the faults or variations that may affect the higher level performances. From the above discussion, it is apparent

that a new and comprehensive set of fault models for each of the nanotechnologies (or at least, for the more promising ones) is highly desirable. Moreover, high level test vectors should not only be able to evaluate the output against a desired characteristic but also be able to detect the faults pertaining to a specific nanotechnology, which may have percolated to the system level.

2.4 Forrest Brewer

It is impossible to ignore the potential impact of nanoscopically-scaled design components on the design paradigm at all scales. Nanoscopic devices promise enormous improvements in implementation density and fundamental improvements in computation power. Rapid progress has been made in devices and logic circuit implementation using technologies as disparate as organic molecular switching, quantum switching of optical states, or nanoscopic implementations of conventional FET's such as carbon nano-tube transistors. Clearly, the potential scale at which these devices can be built offers unprecedented opportunities for system design. On the other hand, a device does not make a system and scalable systems need to exist and co-exist in the practical world. Despite the enormous potentials of these technologies, there remain several issues which must be solved to apply these technologies (or to apply conventionally scaled CMOS [9] or its descendent) in practical applications.

The primary issue in applying any nanometer scale technology is to make an interface to larger scale systems. In the near term, this interface is almost certainly going to be comprised of conventional interconnect wires. This is true even for technologies in which the local distribution is quantum based such as molecular/polymer switches, electron spin or single-electron transistor (SET) circuits. In these circuits, the local interconnection is quantum limited and does not scale or fan-out. Thus, locally nanoscopic and globally submicron information distribution is a likely scenario. This paradigm has the further advantage of a ready and existing infrastructure for system applications. One might wonder if these scaled systems can support the bandwidth necessary to make use of the underlying nanotechnology. Again, it is likely that very highly scaled systems will not see the fundamental decrease in switching time which has accompanied CMOS scaling. There are several reasons for this: If the technology is operated at room temperature, the same thermodynamic trade-offs that occur in FET based switches will also occur in nanoscopically scaled ones. In particular, the trade-off between switch operation voltage and leakage depends in order of magnitude on the dimensionality and effective temperature of the carriers— i.e. one might be able to integrate devices at 1000x times the currently achievable density, but power densities will still scale with switching

time. A similar scaling issue occurs with quantum uncertainty principle: as the devices scale to smaller sizes, the energy of switching must decrease to enable integration. Since energy and time are coupled variables, this energy scaling will eventually lead to an increase in the switching time. The upshot of these arguments is that nanoscopically scaled systems can have very high performance, very high connectivity or very high density, but not all three.

For these reasons, it seems likely that initial nanoscopic applications will appear as compact integrated subsystems such as extremely dense memories or content-addressable storage. These applications were indeed the first for the single-electron transistor circuits, for which large memories have already appeared and will continue to be scaled in the near future. Several other possibilities exist such as special purpose functional and pipelined execution units, and LUT-based programmable logic. It is likely that these local units will be composed of regular arrays of local devices and interconnect with personality being programmed or designed in at a higher level of connectivity. It is not likely that future applications can be composed only of these components. In particular, conventional architectures require substantial fan-out, signal amplification and broadcast. Redundant architectures require even more communication fanout. For these reasons, my guess is that these nanoscopic subsystems will incorporate local redundancy, error detection and correction as integral to the particular subsystem design – so that erroneous results need not be communicated to higher levels of the design abstraction. What is very likely is that these subsystems will not communicate synchronously to the FET based higher scale infrastructure.

From the system level, these conjectures seem to imply that at the compositional high-level, these systems will not exhibit substantial errors or probabilistic behavior, but they will very likely be essentially self timed in terms of I/O events. This is a current trend [6] in high performance design and is likely to continue into the future. The problem from the high-level point of view is maintaining coherent design behavior in the face of substantial temporal uncertainty. I think this will lead to a paradigm based on confluent design in which data and control information is combined and sent (possibly redundantly) on multiple point to point connections. A key metric of the design is the number of required confluence points where information traveling on multiple paths must be locally synchronized (or at least temporally ordered and/or selected). Each required confluence point creates constraints on the information flow (and thus on the system performance). System test and evaluation must be designed in to such systems since state comparisons are only meaningful at confluence points. This model of the system appears very similar to that currently employed by GRID and large scale network computations. I suspect that similar models are applicable, with only rela-

tive scale changes in the paradigm.

These arguments do not imply that the computation model is inevitably limited to scaling of large parallel systems. Nanoscopic systems offer the potential for much higher local connectivity than is currently possible, but the incorporation of these technologies is likely to be limited by the very real system infrastructure constraints that currently exist. In particular, molecular computation schemes seem to offer the possibility of local interconnection Rent [2,4] parameters that are much greater than are feasible for conventionally wired integrated circuits, connectivity that is rivaled only by biological neural systems.

2.5 Sankar Basu

Information on funding in different areas of the DA program - a bit about the Nano-science and technology Initiative (not the whole NSF wide Nano initiative but as it is related to the DA program in CISE) will be provided in my panel statement.

References

[1] R. Dobrushin and E. Ortyukov. Upper bound on the redundancy of self-correcting arrangements of unreliable functional elements. *Problems of Information Transmission*, 13(3):203–218, July-September 1977.

[2] W. E. Donath. Placement and average ineterconnect requirements of computer logic. *IEEE Transactions on Circuits and Systems*, 26:272–277, 1979.

[3] W. Evans and N. Pippenger. On the maximum tolerable noise for reliable computation by formulas. *IEEE Transactions on Information Theory*, 44(3):1299–1305, May 1998.

[4] M. Feuer. Connectivity of random logic. *IEEE Transactions on Computers*, 31:29–33, 1982.

[5] G. Norman and D. Parker and M. Kwaitkowska and S. Shukla. Evaluating the Reliability of Defect-Tolerant Architectures for Nanotechnology with Probabilistic Model Checking. In *15th International VLSI Conference, Mumbai, India*. IEEE Press, January 2004.

[6] P. Gronowski, W. Bowhill, R. Preston, M. Gowan, and R. Allmon. High performance microprocessor design. *IEEE Journal of Solid-State Circuits*, pages 676–686, 1998.

[7] J. Han and P. Jonker. A system architecture solution for unreliable nanoelectronic devices. *IEEE Transactions on Nanotechnology*, 1:201–208, December 2002.

[8] J. Heath, G. S. P. Kuekes, and R. Williams. A defect tolerant computer architecture: Opportunities for nanotechnology. *Science*, 80:1716–1721, June 1998.

[9] The international technology roadmap for semiconductors (itrs). Technical report, SRC, 2001.

[10] K. Nikolic, A. Sadek, and M. Forshaw. Architectures for reliable computing with unreliable nanodevices. In *Proc. 1st IEEE Conf. on Nanotechnology, (IEEE-NANO)*, pages 254–259, October 2001.

[11] N. Pippenger. Reliable computation by formulas in the presence of noise. *IEEE Transactions on Information Theory*, 34(2):194–197, March 1988.

[12] M. Ratner and D. Ratner. *Nanotechnology: Gentle Introduction to the Next Big Idea*. Prentice Hall, 2003.

[13] J. von Neumann. Probabilistic logics and synthesis of reliable organisms from unreliable components. *Automata Studies*, Princeton University Press:43–98, March 1956.

[14] *International Technology Roadmap for Semiconductors*. Sematech, 2001.

[15] M. Butts, A. DeHon, and S. Goldstein, 'Molecular electronics: Devices, systems and tools for gigagate, gigabit chips,"in *ICCAD-2002*, Nov. 2002.

[16] R. Wilson, '"structured' asics arrive,"*EETimes*, May 5 2003.

[17] S. Nassif, 'Delay variability: Sources, impacts and trends," in *Proceedings of the ISSCC*, 2000, pp. 368–9.

[18] K. A. Bowman and J. D. Meindl, 'Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration," *IEEE JSSC*, vol. 37, no. 2, Feb. 2002.

[19] C. P. Collier, E. W. Wong, M. Belohradský, F. M. Raymo, J. F. Stoddart, P. J. Kuekes, R. S. Williams, and J. R. Heath, 'Electronically configurable molecular-based logic gates," *Science*, vol. 285, pp. 391–394, July 16 1999.

[20] S. Goldstein and M. Budiu, 'NanoFabrics: Spatial computing using molecular electronics," in *Proceedings of the 28th Annual International Symposium on Computer Architecture*, June 2001, pp. 178–189.

[21] M. Mishra and S. C. Goldstein, 'Defect tolerance at the end of the roadmap,"in *Proceedings of the International Test Conference (ITC), 2003*, Charlotte, NC, Sep 30 – Oct 2 2003.

[22] A. M. Ionescu and K. Banerjee. *Emerging Nanoelectronics: Life with and after CMOS*. Kluwer Academic Publishers, 2003.

[23] A. M. Ionescu, M. J. Declercq, S. Mahapatra, K. Banerjee, and J. Gautier. Few electron devices: Towards hybrid cmos-set integrated circuits. In *Proc. Design Automation Conference (DAC 2002)*, pages 88–93, 2002.

[24] M. Lundstrom. Nanoelectronics the future of microelectronics? In *International Symposium on Low Power Electronic Design*, page 172, 2002.

[25] O. Trel and K. Likharev. Crossnets: Possible neuromorphic networks based on nanoscale components. *International Journal of Circuit Theory and Applications*, 31:37–52, 2003.

[26] M. Ziegler and M. Stan. Case for cmos-nano co-design. In *International Conference on Computer-Aided Design*, pages 348–352, 2002.