

Impact of Process Variation on NASIC Nanoprocessors with 2-Way Redundancy

Michael Leuchtenburg, Pritish Narayanan, Teng Wang, and Csaba Andras Moritz*, *Member, IEEE*

Abstract- Process variation is expected to persist in the various novel nanoscale fabrics being proposed to replace CMOS. Logic circuits built using non-traditional and bottom-up techniques will need to meet new design rules, such as tolerance of high defect rates and use of regular structures in layout. One circuit fabric type that meets these requirements is grid-based logic, with built-in fault resilience provided by 2-way redundancy. In this work, we show that this fabric design also is able to tolerate substantial process variation in addition to its defect resistance.

I. BACKGROUND

In CMOS, process variation will reach as high as $3\sigma=12\%$ [1]. Current manufacturing tests that process variation should be expected to be even higher with nanoscale manufacturing, with reports in the literature reaching as high as $3\sigma=30\%$ [2-6]; much higher variation than that is possible depending on exact manufacturing techniques used for grid assembly (e.g., in-site versus ex-situ), alignment techniques, and targeted feature sizes.

Self-assembly and other bottom-up manufacturing techniques can be used to make regular structures, but arbitrary shapes range from difficult to impossible to build. One regular structure which lends itself to implementing logic at nanoscale is the grid. Several grid-based nanoscale circuit fabrics have been proposed, such as NASIC, CMOL, and crossbar latches [7-9].

It is widely expected that in order to scale down manufacturing to the nanoscale, we will have to accept substantially higher defect rates. Defect rates in 90nm CMOS processes are as low as 0.4 defects/cm². In contrast, the expected rates for nanoscale manufacturing processes are as high as 5-15% of devices, or 10⁹ defects/cm². In order to make functional circuits with the high defect rates expected, defect tolerance of some sort will be necessary. One class of techniques for providing defect tolerance is built-in fault masking.

This is a class of design-time techniques in which the circuits are modified to mask the faults caused by permanent defects. Typically this is accomplished through some form of redundancy, where multiple copies of each gate are made, and the inputs are similarly duplicated. The circuit is structured such that if one copy of a gate fails, the correct functioning of

the redundant copy (or copies) will still result in correct functioning of the circuit as a whole. This class of techniques can also mask faults from sources other than permanent defects, since the techniques are oriented around masking faults rather than specifically eliminating or mapping around defects.

Given that this fault tolerance must already be included in order to achieve acceptable yield, it is important to understand whether it could also be used to mask timing faults caused by process variation. Since the built in fault resilience techniques are generic, they should be able to mask timing faults, allowing the circuit to be run at a faster speed than it might have otherwise been restricted to.

We have been working on many techniques of this type; in this work we show that 2-way redundancy not only prevents performance degradation from process variation but actually increases mean performance.

II. FAULT RESILIENCE

One grid-based nanoscale circuit fabric is NASIC, or Nanoscale Application Specific Integrated Circuits. NASIC is based on crossed nanowire grids with FETs selectively implemented at the crosspoints of the nanowires. Clocking and power are provided by surrounding microwires, which are driven by reliable CMOS logic.

A fault resilience method which is particularly suited to grid-based logic is 2-way redundancy. In a circuit with 2-way redundancy, each gate is duplicated, as is each input to each gate. In a NAND-NAND circuit, this means that any single 0-to-1 (i.e. sub-critical, for NAND) fault will be corrected in the next stage as there will also be a correct 0 generated by the redundant gate and received as an input by the following gate. This is particularly suited to grid-based logic because it can be configured to make 0-to-1 faults far more likely than 1-to-0 faults, making it an efficient and effective technique for defect

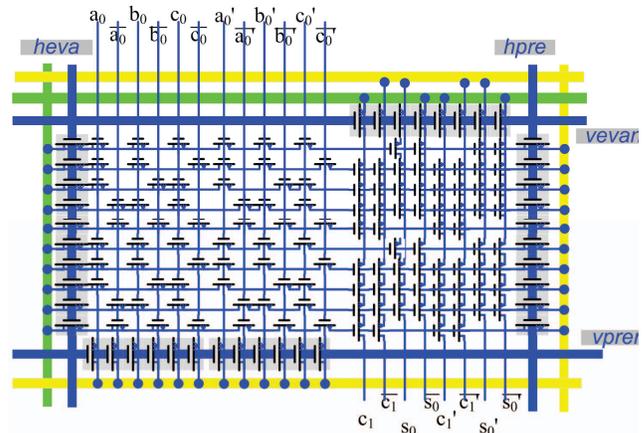


Fig. 1 1-bit adder in NASIC with 2-way redundancy

We acknowledge support from the Focus Center Research Program (FCRP) - Center on Functional Engineered Nano Architectonics (FENA). This work was also supported by the Center for Hierarchical Manufacturing (CHM) University of Massachusetts Amherst and NSF award number 0541066.. Teng Wang is with Qualcomm Inc. All other 3 authors are with the Electrical and Computer Engineering Department, College of Engineering, University of Massachusetts Amherst. (*contact author: phone: (413)-545-2442, email: andras@ecs.umass.edu)

tolerance. For instance, in tests on the WISP-0 nanoprocessor design built on the NASIC fabric, 2-way redundancy was shown to provide a yield of 80% with 3% of transistors faulty [7]. An example of a 1-bit adder implemented with 2-way redundancy is shown in Fig. 1.

This fault tolerance method helps both with permanent defects and with faults caused by process variation. Each gate is duplicated, and one of the two gates will be faster than the other. In the case of a missed deadline, a 0-to-1 fault will result. This is due to the use of dynamic circuits in NASIC. Each output is first pre-charged to a value of 1 and may then be discharged to a value of 0. If the value should be discharged to 0 but the time given by the clock selected is too short, then the output will remain a 1, or at some ambiguous value. Only if the time for precharging was insufficient would a 1-to-0 fault occur, and since the precharge period is much shorter than the evaluate period in actual circuits, this can easily be avoided.

Because 0-to-1 faults result from missed deadlines, the faster gate's output will always mask the incorrect result from the slower gate if the circuit is run at the speed allowed by the faster gate. This allows the redundant gates to serve a dual purpose, enabling reasonable yield despite high defect rates and additionally speeding up operation in the case where both gates of a pair are functional.

III. EVALUATION

A. Evaluation Platform

WISP-0 is a NASIC stream processor implementing a 5-stage pipeline. It is used as a prototype for evaluating NASICs. In this work, we use WISP-0 with 2-way redundancy on all tiles to test the impact of 2-way redundancy on performance with process variation induced delay variation.

The design of WISP-0 is illustrated in Fig. 2.

B. Parameters for Evaluation

Delay calculations are done using RC equivalent circuits to model nanowires with each transistor being changed into ON (low resistance) and OFF (high resistance) modes by its gate input. There is also the resistance of the nanowire interconnect and contact resistance with the CMOS wires that drive the gates. Capacitance sources include inter-wire and junctions between the wires where transistors are formed.

The values of resistive and capacitive elements are calculated using their geometry together with several parameters which vary between devices. The nominal value of the transistor length evaluated is 4nm and the width is 4nm, the square aspect ratio being due to the crossed-nanowire devices. The width of the nanowire depends on e.g., the size of the catalyst nanoparticles used as seeds for Vapor-Liquid-Solid nanowire growth [2]. Variations in nanoparticle sizes therefore directly correlate with variations in nanowire width. The standard deviation of wire widths has been shown to be around

10% of the mean [3][4]. In addition, the width of the nanowire is assumed to be uniform along its length.

The transistor ON resistance is also varied by the gate geometry – i.e., length and width – that are determined by the width of the nanowires. This allows us to make predictions of delay that are based on experimental work and geometric variation. We do not use such traditional CMOS parameters as V_{th} and L_{eff} . This is primarily due to the lack of settled device models for cross-nanowire FETs for circuit simulations. Instead, we abstract them into a variation in transistor resistance R_{ON} .

All device parameters are taken from the literature on manufacturing of nanoscale devices. N-type FETs are used in these calculations. The n-type devices are assumed to be Silicon Nanowires (SiNW) lightly doped with Phosphorus. The nominal ON resistance for the specified geometries for n-type devices (R_{ON}) has been calculated to be 3.75Ω based on experimental work [2]. The overall standard deviation of transistor ON resistance has been found to be $\sim 20\%$ [6], including variation of gate geometry. After removing the variation in gate geometry, this reflects a variation of 10% in transistor ON resistance for a square transistor. The nominal contact resistance with CMOS wires delivering V_{DD} and GND has been found to be $\sim 10k\Omega$ [5].

Interconnect variation is modeled geometrically. The diameter, resistivity, and contact resistance of each nanowire varies independently. Interconnect is assumed to be made by transforming silicon nanowires into nickel silicide (NiSi) via silicidation. The resistivity of the resulting nanowires after thermal annealing averages $9.5\mu\Omega\text{-cm}$ [3]. Non-uniform metallization can lead to variations in the resistivity of interconnect.

The capacitance is calculated geometrically using the standard expressions for cylindrical wires. These are calculated using the wire diameter and pitch together with circuit layout. The dielectric is SiO_2 with a dielectric constant of 3.7. The nominal value for parallel nanowire capacitance is calculated to be 53.49 pF/m and for junction overlap capacitance is 0.602aF . Variations in the spacing of nanowires, due to self-assembly based techniques for creating parallel arrays, will lead to variations in parallel nanowire capacitance. The thickness of the dielectric layer between wires (nominal = 5nm) is determined by the inter-wire spacing, as the space between them is filled with SiO_2 during manufacturing.

TABLE I
PARAMETERS USED FOR TIMING SIMULATION

Parameter	Nominal Value	Standard Deviation
Wire resistivity of NiSi (ρ_{NiSi})	$9.5\ \mu\Omega\text{-cm}$	10%
Wire diameter (d)	4nm	10%
Wire pitch	10nm	10%
Contact Resistance	10Ω	10%
Transistor ON Resistance (R_{ON})	$3.75k\Omega$	10%
Oxide dielectric constant (ϵ_r)	3.7	None
Oxide dielectric thickness (t_{ox})	5nm	10%

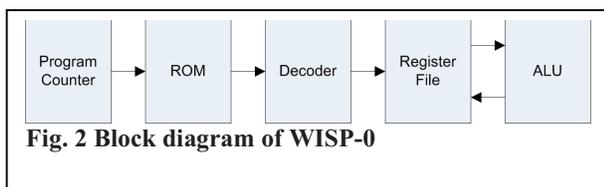


Fig. 2 Block diagram of WISP-0

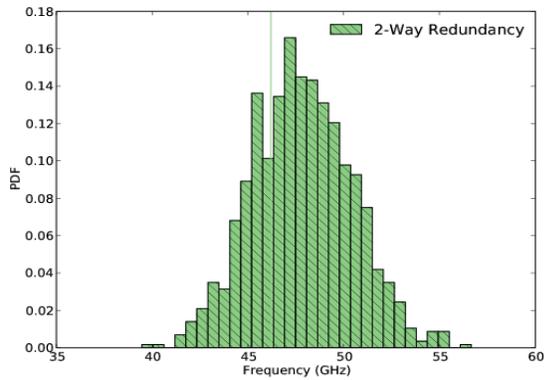


Fig. 3 Distribution of frequencies for WISP-0 with 2-way redundancy. Vertical line shows nominal frequency.

These parameters are summarized in Table I.

C. Method of Evaluation

In order to evaluate the effects of parameter variation on WISP-0, we built a simulator that incorporates physical design parameters and is capable of handling both parameter variation and device faults. The simulator uses Monte Carlo techniques to capture statistical distributions.

The simulator takes as input a circuit design and a set of manufacturing parameters as described above, including the amount of variation to apply. The characteristics of each wire, cross-nanowire transistor, and so forth are varied independently by the simulator based on the variation model given as input.

Additionally, it takes settings for defects. The defect model consists of a set of defect types and probabilities for each, plus a description of any clustering behavior [7]. For instance, the defect model for a given test might be “5% probability of each transistor being stuck on, with uniform distribution of defects”.

The simulator then generates a series of designs based on these parameters, with the values of each individual circuit element changing based on the parameters defined. The speed is then determined by simulating the WISP-0 design and measuring whether it generates correct output, searching to find the fastest speed at which correct output is generated by that test WISP-0 nanoprocessor. If correct output is not generated at any speed due to defects, then the nanoprocessor is noted as faulty and no speed is recorded for that sample. This is repeated many times and each maximum speed is recorded, giving an overall statistical distribution of operating speeds.

It may seem that this could be more quickly done by simply measuring the speed of each gate and taking the speed of the slowest gate as the speed of the nanoprocessor *in toto*, but this would be inaccurate. This is because in many cases, the fault caused by a gate working incorrectly will be masked by the built-in fault tolerance. Therefore the only accurate way to determine the speed is through simulation of the complete

nanoprocessor including all its circuits.

Additionally, the nominal frequency of the design, equal to the maximum frequency of the design with all parameters set to their nominal (i.e., zero variation) values, is measured.

IV. RESULTS

For these results, 1,000 test WISP-0 nanoprocessors were simulated and the operating speed of each measured. The nominal frequency of the WISP-0 nanoprocessor design was found to be 46.2GHz with chosen device and fabric parameters, while the mean operating frequency of the test nanoprocessors is higher at 47.8GHz, with a standard deviation of 2.6GHz. It can be clearly seen from Fig. 3 that the majority of nanoprocessors run faster than the nominal one. Process variation typically results in a decrease in performance, but in this case a moderate increase is seen instead.

This demonstrates that the built-in fault resilience, which is required for an acceptable yield in nanoscale fabrics, also provides a benefit in providing higher performance under process variation. Counter-intuitively, average performance is actually higher with process variation than without.

V. CONCLUSION

We have shown that 2-way redundancy is effective for mitigating the effects of process variation on performance and, in fact, that WISP-0 nanoprocessors with 2-way redundancy have higher performance with process variation than without. This shows promise for other similar redundancy techniques which are more tailored to provide resilience against faults caused by process variation. We are currently evaluating many such techniques, which will be reported on in future works.

REFERENCES

- [1] *The International Technology Roadmap for Semiconductors (ITRS) Reports*. Available <http://www.itrs.net/>.
- [2] W Lu and CM Lieber, “Semiconductor nanowires,” *J. Phys. D: Appl. Physics*, vol. 39, pp. R387-R406, October 2006.
- [3] Y Wu, J Xiang, C Yang, W Lu, and CM Lieber, “Single-crystal metallic nanowires and metal/semiconductor nanowire heterostructures,” *Nature*, vol. 430, pp. 61-65, 2004.
- [4] E Garnett, W Liang, and P Yang, “Growth and electrical characteristics of platinum-nano-particle-catalyzed silicon nanowires,” *Advanced Materials*, vol. 19, pp. 2946-2950, 2007.
- [5] J Kim, DH Shin, E Lee, and C Han, “Electrical characteristics of singly and doubly connected Ni silicide nanowire grown by plasma-enhanced vapor deposition,” *Applied Physics Letters* 90, 253103, 2007.
- [6] Y Cui, Z Zhong, D Wang, WU Wang, and CM Lieber, “High performance silicon nanowire field effect transistors,” *Nano Letters*, vol. 3, no. 2, pp. 149-152, 2003.
- [7] CA Moritz, T Wang, P Narayanan, M Leuchtenburg, Y Guo, C Dezan, and M Bennisner, “Fault-tolerant nanoscale processors on semiconductor nanowire grids,” *IEEE Trans. On Circuits and Systems I, special issue on Nanoelectronic Circuits and Nanoarchitectures*, vol. 54, iss. 11, pp. 2422-2437, November 2007.
- [8] K Likharev, DB Strukov, “CMOL FPGA: A reconfigurable architecture for hybrid digital circuits with two-terminal nanodevices”, *Nanotechnology* 16, no 6, pp 888-900, 2005.
- [9] PJ Kuekes, DR Stewart, RS Williams, “The crossbar latch: Logic value storage, restoration, and inversion in crossbar circuits”, *J. Appl. Phys.* 97, 034301, 2005.