Toward Holistic Modeling, Margining and Tolerance of IC Variability

Andrew B. Kahng
UC San Diego CSE and ECE Depts., La Jolla, CA 92093-0404 USA
abk@ucsd.edu

***Abstract*—The 2013 edition of the *International Technology Roadmap for Semiconductors* [10] highlights a slowdown of traditional pitch and density scaling in leading-edge patterning technologies. Through the foundry N5/N7 nodes, the roadmap also projects unfavorable scaling of device and interconnect electrical performance (drive vs. leakage, resistivity, capacitive coupling, etc.). IC product value is also challenged by increasingly dominant *variability* mechanisms ranging from lithography and planarization in manufacturing, to dynamic voltage droop and aging in the field. Design teams compensate variability with margin (guardbanding), but this substantially reduces the value of designs at the next technology node. In this context, it is increasingly critical to deliver *design-based equivalent scaling* through novel design technologies. This paper reviews recent research directions that seek to improve *modeling*, *margining* and *tolerance* of IC variability. Collectively, these design methods offer new means by which product companies can extract greater value from available technologies, even as traditional scaling slows for patterning, devices and interconnects.**

I. Introduction

The recent history of the semiconductor industry shows that even as patterning technologies have delivered "available" Moore's Law scaling (i.e., geometric pitch scaling) per the *International Technology Roadmap for Semiconductors* (ITRS) [10] at least through the year 2013, the "realized" transistor density scaling in actual products has *since 2008* slowed to 1.6× per node instead of the traditional 2× per node [15]. The inconsistency between "available" density scaling and "realizable" density scaling is a clear challenge to Moore's-Law scaling of IC value; it may be attributed to a *design capability gap* [15] that stems from patterning constraints, reliability constraints, variability in process and operating conditions, signoff analysis pessimism, and other costly trends at the technology-design leading edge.

The design capability gap is compounded by BEOL challenges that include resistivity and manufacturability of damascene copper interconnects, poor design-level ROI from new technologies with heavily restricted layout ground rules, and wide parasitic extraction corners in multi-patterning technologies. These have led the 2013 ITRS to dramatically slow the roadmap for interconnect pitch scaling, compared to the previous 2011 edition.[1] Aside from any slowdown of scaling, it has been apparent for nearly a decade that benefits from scaling have been stuck in a "20-20-20" world, where 20% speed, 20% power, and 20% (or slightly better) density scaling comprise the overall design benefit from a given next technology node. Examples of "20-20-20" abound, e.g., [20], [21], [24], [25], [27]–[30]. In this context, there is an urgent need for novel IC design tools and methodologies that can achieve "equivalent scaling" of product value.

This paper reviews several recent research directions at UCSD that address *modeling*, *margining* and *tolerance* of IC variability in future design methodologies. The driving motivation for this research is: "How can designers scale the value of IC products, even as scaling slows in patterning, device and interconnect technologies?" Collectively, the methods reviewed offer the potential for *design-based equivalent scaling* that enables product companies to extract greater value from existing process nodes, even as traditional scaling slows in the roadmap.

[1]In the 2013 ITRS, Metal-1 half-pitch (M1HP) has been 'reset', and scaling has been slowed to a three-year cycle over the next two technology nodes, as opposed to the two-year cycle previously projected in the 2011 ITRS roadmap.

TABLE I
Example works reviewed.

Topic	Modeling	Modeling	Tolerance
BEOL Corner Optimization	[4]		
Process-Aware V_{dd} Scaling		[5]	
{BTI, EM}-AVS Interactions		[2], [3], [7]	
Overdrive Signoff		[6]	
Min Cost of Resilience			[13]
Approximate Arithmetic			[12]

II. Modeling

We use the term *modeling* to broadly encompass the capture of process variations in design implementation and timing signoff methodologies. In advanced technology nodes, variation is main threat of design and signoff methodologies – and modeling pessimism (including model-to-silicon miscorrelation) is an increasingly dominant root cause of design guardbands. (The substantial costs of design guardbands has been quantified in works such as [11].) This section presents one example of potential improvement of variation modeling, in the BEOL context.

Example: Tightening BEOL Corners in Signoff

To ensure functional correctness, conventional implementation methodology signs off a given system-on-chip (SOC) design at extreme process, voltage and temperature (PVT) conditions. At the 20nm foundry node and below, (double-patterned, highly resistive) BEOL layers have become major sources of variation, which must be accounted for by signoff at BEOL corners. Existing signoff methodology defines the BEOL corners such that all BEOL layers vary in the same way, e.g., all BEOL layers simultaneously have worst-case capacitance. Although the BEOL corners can capture most of the timing violations at the extreme BEOL condition, the BEOL corners are unnecessarily pessimistic. This is because the BEOL layers are not perfectly correlated, and the likelihood of a worst-case (or, best-case) condition on all layers is vanishingly small (if not a physical impossibility). The conservatism in BEOL corners results in longer chip implementation schedules (time spent on design closure steps), as well as area and power overheads (buffering and sizing to meet timing and signal integrity constraints).

To reduce conservatism in *conventional BEOL corners* (CBCs), we have recently studied the potential of "tightening" the CBCs [4]. This is based on the observation that most timing-critical paths are routed on multiple BEOL layers. Figure 1 shows the wirelength ratio of (setup) critical paths extracted from a design implemented in a 45nm foundry library. From the figure, we can see that the critical paths are mostly routed on layers 2 to 6, and that the proportion of wirelength on a single layer never exceeds 50%. When the process variations of the BEOL layers are not perfectly correlated, the BEOL-induced timing variations are also not perfectly correlated, and timing variations on a critical path will be much smaller due to averaging of uncorrelated variations. More specifically, if we model BEOL-induced timing variation of a critical path, p_j, as a normal distribution, the timing variation of the path ($\sigma_{delay_eff_j}$) becomes

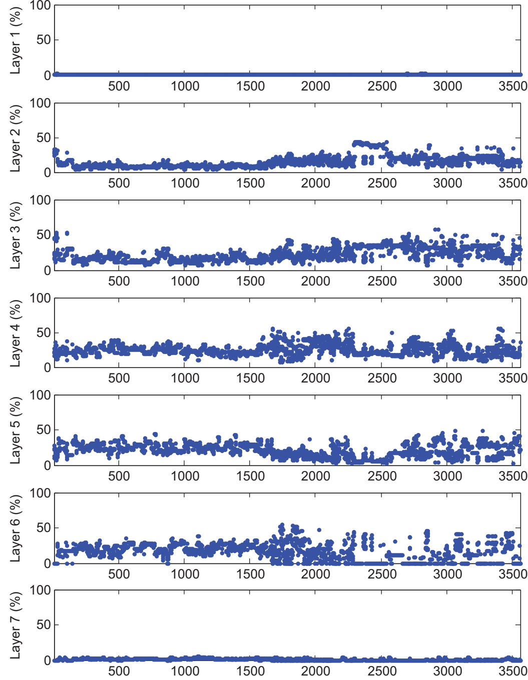

Fig. 1. Wirelength distribution across BEOL layers of setup-critical timing paths.

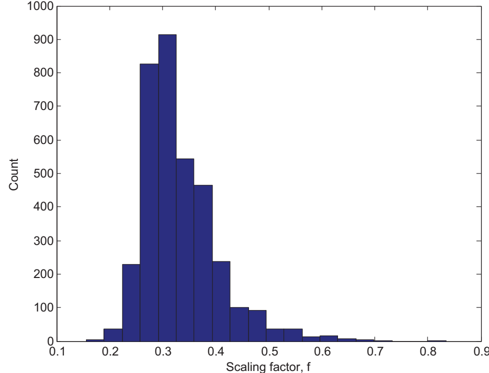

Fig. 2. Histogram of scaling factors for critical paths.

$$\sigma_{delay_eff_j} = \sqrt{\sum_{m=1}^{N} (\sigma^2_{delay_j_m})} \qquad (1)$$

where $\sigma_{delay_j_m}$ represents the standard deviation of the jth critical path delay due to BEOL variation on layer m, and N is the total number of BEOL layers. Using Equation (1), we may calculate $\sigma_{delay_eff_j}$ and compare it with the *delay variation*, Δd_j (defined as the path delay difference between a CBC and the typical BEOL corner), obtained from CBCs. Further, we may calculate a *BEOL scaling factor*, f_j, which is defined as the ratio of $3 \cdot \sigma_{delay_eff_j}$ to Δd_j. In other words, f_j is the scaling factor such that the delay variation is same as three times $\sigma_{delay_eff_j}$ (assuming that delay variation scales linearly with f_j). The results in Figure 2 show that most of the setup-critical paths have scaling factors < 0.7. This suggests that the delta delay from CBCs is quite pessimistic, and that the CBC corners can be tightened by up to 30%. Potential benefits of such corner tightening include avoidance of unnecessary timing fixes which currently delay the SOC tapeout schedule and severely compromise power-performance-area (PPA) product metrics.

Separately, [4] observes that critical paths can be classified as C- or R-dominated based on their delay sensitivities to delta capacitance and resistance in BEOL layers. The C- vs. R-dominance dichotomy has many implications for future physical implementation methodology, e.g., (i) to reduce clock skew across BEOL corners, clock tree synthesis should avoid creating different kinds of 'gate-wire balance' in clock distributionpaths; and (ii) design implementation and signoff must correctly choose the 'dominant' BEOL corner that captures the impact of BEOL variations on a given critical path.

III. Margining

We use the term *margining* to broadly indicate the process of defining proper PVT conditions in timing analysis or runtime so as to minimize design overheads at tapeout and during runtime. This section reviews several works that focus on interactions between margining and adaptivity mechanisms. Process and runtime variations can be compensated with adaptive voltage scaling (AVS); below, we describe a process monitor design that is enabling to *process-aware voltage scaling* (PVS). We also review works that investigate how signoff criteria should change given the interaction between adaptivity mechanisms and design reliability requirements. Last, an optimization of overdrive signoff criteria is described which reduces resource waste due to overdesign.

Example: Process-Aware Voltage Scaling (PVS)

VLSI circuits usually allocate excess margin to account for worst-case process variation. Since most chips are fabricated at process conditions better than the worst-case corner, adaptive voltage scaling (AVS) is commonly used to reduce power consumption whenever possible. A typical AVS setup relies on a *performance monitor* that replicates critical paths of the circuit to guide voltage scaling. However, it is difficult to define appropriate critical paths for an SOC which has multiple operating modes and IPs. Chan and Kahng [5] propose a methodology for AVS which matches the voltage scaling characteristics of a circuit, rather than the delays of critical paths. This fundamental change in monitoring strategy simplifies the monitoring circuitry as well as the calibration flow of conventional monitoring methods. At the heart of the methodology is *process-aware voltage scaling* (PVS) *ring oscillators* (ROs) which are designed so that they require a relatively higher supply voltage compared to critical paths of a SOC, enabling safe compensation of process variation-induced frequency drift. In other words, any SOC manufactured in the given process can safely perform a closed-loop AVS by using these ROs as hardware performance monitors.

Two basic usage scenarios for the proposed ROs [5] are shown in Figure 3. At the design stage, the PVS ROs are designed using SPICE models and standard cells. At the silicon characterization stage, sample test chips at different process corners are provided by the foundry. The ROs' frequencies are measured at nominal operating voltage (V_0). In **Scenario 1**, the frequencies measured at the signoff corner (e.g., SS corner) are used as the target frequencies of the ROs during AVS. In this scenario, the ROs have no information about the design, and are designed to guardband for worst-case voltage scaling characteristics. Hence, AVS guided by our ROs will always overestimate the supply voltage needed for a chip to meet its operating frequency.

Excess supply voltage (and hence power consumption) can be reduced when the chip maximum frequency f_{max} is also measured during the silicon characterization stage (**Scenario 2**). In this latter scenario, designers can tune the voltage scaling characteristics of the ROs such that the supply voltage suggested by the AVS (in the silicon characterization stage, guided by the ROs) for each chip is slightly higher than the minimum voltage (V_{min_chip}) needed for that chip to meet its required operating frequency. When all test chips manufactured for silicon characterization can safely operate at their

respective operating frequencies using AVS guided by the PVS ROs, designers record the configurations of the ROs. We note that in the characterization step of Scenario 2, the test chips are manufactured at biased process corners, so that calibration with these test chips will configure the ROs to account for circuit performance across a wide spectrum of process variation. The sampling of test chips at different process corners is important because this allows the configurations of the ROs to be applied in the subsequent production stage without additional calibrations. During mass production, the previously obtained ROs' configurations will be stored in every production chip. Then, designers run AVS tests with the stored ROs' configurations and RO target frequencies. If a chip fails to meet its target frequency with the AVS guided by PVS ROs, the silicon characterization step can be modified if necessary (e.g., adjust ROs' configurations so that the AVS becomes less aggressive in reducing supply voltage).

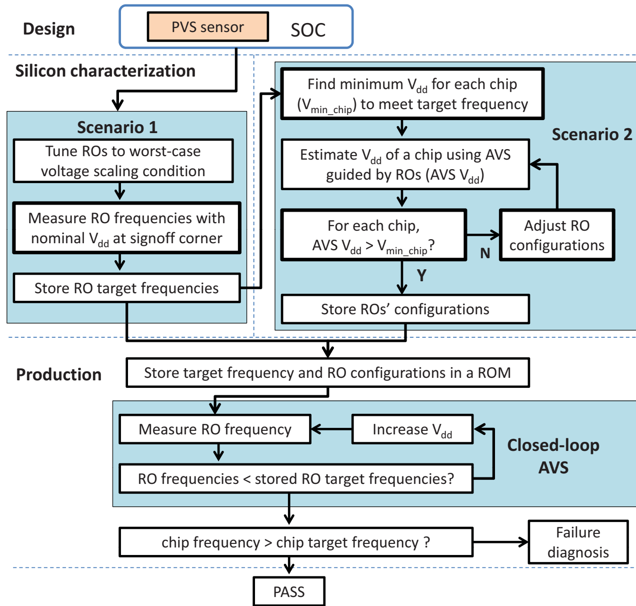

Fig. 3. Application example for the proposed tunable (PVS) ROs.

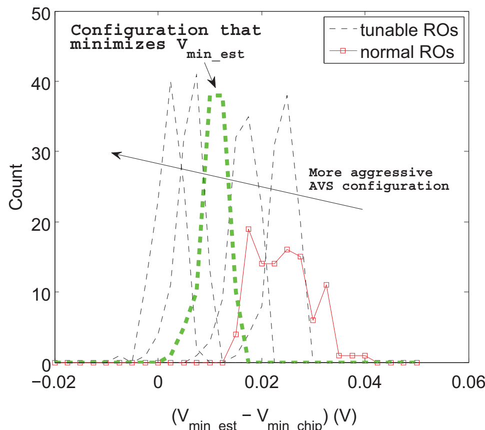

Fig. 4. Every line represents the distribution of ($V_{min_est} - V_{min_chip}$) for 100 Monte Carlo samples with different process variations. By tuning the configuration of the PVS ROs, designers can change the voltage scaling characteristics (V_{min_est}). Shown: an optimized configuration can reduce V_{min_est} by 13mV (on average) compared to standard non-tunable ROs.

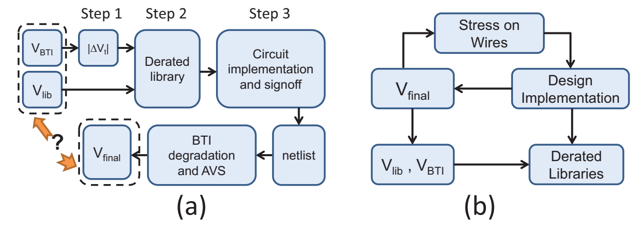

Fig. 5. (a) "Chicken-and-egg" loop due to inconsistency between voltages used to model supply voltage (V_{lib}) and BTI stress (V_{BTI}) when characterizing a derated library for BTI-aware signoff. (b) Additional loop with awareness of EM.

Figure 4 shows an example of the application of the PVS methodology. The minimum voltage of PVS ROs and of a design implemented using a 65nm GP foundry library are denoted by V_{min_est} and V_{min_chip}, respectively. When designers use the conservative configuration (i.e., the rightmost dashed black lines) V_{min_est} is always smaller than V_{min_chip}. By tuning the configuration of the PVS ROs, designers can obtain a more aggressive AVS configuration for voltage reduction. The maximum voltage reduction configuration (green color) enables $\sim$13mV voltage reduction on average compared to the non-tunable ROs.

Example: AVS-Aware Signoff

Reliability signoff is critical in modern sub-22nm technology nodes to guarantee operation at minimum acceptable performance throughout product lifetime [14]. The dominant reliability mechanisms at leading-edge nodes are bias temperature instability (BTI) and electromigration (EM) [14] [18] [19]. BTI degrades chip performance by slowing down device switching speeds in critical paths. EM can increase wire resistance, which can cause voltage drop resulting in device slowdown; it can also cause permanent failures in circuits due to shorts or opens.

Since BTI is worsened at higher supply voltages, designers must add more guardband in the implementation flow to meet product lifetime requirements. Conventional signoff flow relies on technology-dependent libraries to define a set of signoff corners according to operating conditions, e.g., designers can specify supply voltage V_{lib} and the BTI stress voltage V_{BTI} to characterize delay under BTI degradation at a fixed voltage. However, this approach does not correctly guardband BTI in modern designs that implement *adaptive voltage scaling* (AVS), because variable voltage causes a "chicken-and-egg" loop in the signoff flow. The upper part of Figure 5(a) depicts a signoff flow using a derated library, while the lower part depicts how AVS increases the supply voltage to compensate for BTI degradation. As a result, the circuit ends up with a voltage at the end of lifetime (V_{final}) which does not match the voltages (V_{lib}, V_{BTI}) used for library characterization. This inconsistency leads to design overheads.

When the impact of EM on interconnects is considered in designs with AVS, there is a second chicken-egg loop as shown in Figure 5(b). To understand the impact of EM degradation, we study two models – Black's [1] and Mishra-Sapatnekar [17]. To quantify the power and area cost of designs that require to handle the two "chicken-and-egg" loops to meet lifetime requirements, we create eight implementations of the *AES* design from OpenCores [22] using V_{lib} and V_{BTI} as shown in Table II. Figure 6 compares area and power of these eight implementations for three design flows: (i) non-EM aware (only addresses BTI degradation), (ii) with fix of EM violations based on [17] and (iii) with fix of EM violations based on [1]. Designs can consume up to $\sim$6% additional power (e.g., implementation #6) when EM violations are fixed.

TABLE II
SIGNOFF CORNERS FOR BTI. V_{min} = 0.90V, AND V_{max} = 1.10V.

	Implementation #							
	1	2	3	4	5	6	7	8
V_{lib}	V_{min}	V_{min}	V_{max}	V_{min}	0.98V	0.97V	0.96V	0.95V
V_{BTI}	V_{min}	V_{max}	V_{max}	N/A	0.98V	0.97V	0.96V	0.95V

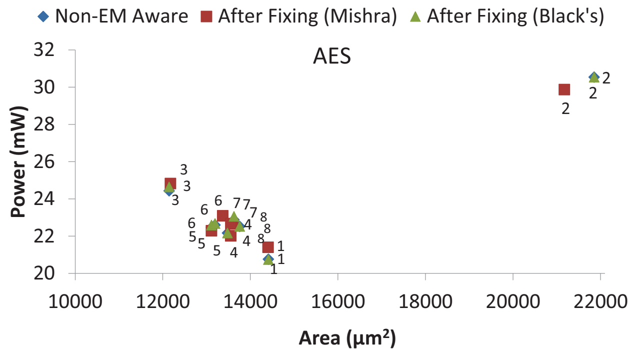

Fig. 6. Power and area after EM fix in AVS systems using Black's Equation [1] and Mishra-Sapatnekar [17] models. Design: AES.

At the system level, there is a limit on maximum allowable supply voltage (V_{max}); hence, initial voltage V_{init} of a chip must be smaller than a threshold ($V_{critical}$), such that the final supply voltage of the chip with AVS is less than V_{max}. [3] uses cell chains to develop estimates of $V_{critical}$ for early-stage evaluation in the implementation flow. Figure 7 shows that there can be up to 29% area overhead if V_{init} is $0.08V$ lower than the $V_{critical} = 0.98V$. The area overhead decreases when designers use a higher V_{init}, and further diminishes when designers use $V_{init} = V_{critical}$.

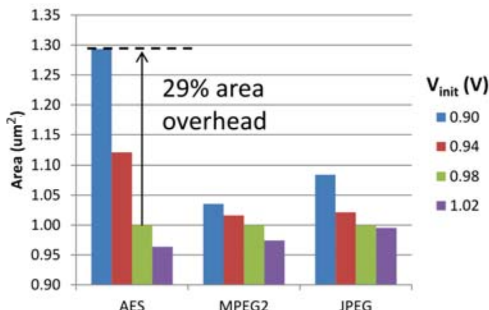

Fig. 7. Area of circuits implemented with non-derated library and zero timing margin, showing area overheads when $V_{init} < V_{critical} = 0.98$V.

Example: Optimization of Signoff Modes

In the era of heterogeneous multi-core SOCs, single-threaded performance limits the overall speedup of applications. Frequency overdrive at elevated voltages is used to obtain better performance in consumer electronic devices. However, the selection of signoff modes has significant impact on circuit area, power and performance. Figure 8 shows that average design power, for given modes (i.e., frequency-voltage pairs) and corresponding duty cycles, can vary by up to 26% across 40 different definitions of the overdrive mode, with a fixed nominal mode. Even when the overdrive frequency is fixed, the average power can vary by up to 12% for different overdrive voltages. This suggests that significant margin reductions can be achieved by careful optimization of signoff modes.

Chan et al. in [6] study the *signoff mode optimization problem*, which seeks the optimal nominal and overdrive modes with respect to optimization objectives and constraints. Based on the frequency-voltage tradeoff, [6] proposes concepts of *design cone* and *equivalent dominance* and demonstrates that multi-mode signoff at modes which do not exhibit pairwise equivalent dominance will lead to overdesign. Motivated by this, a model-based adaptive search methodology is proposed to efficiently explore the design space of signoff modes. Experimental results show up to 8% increase in performance, given V_{dd}, area and power constraints, compared to a traditional "signoff and scale" methodology. Further, the obtained signoff modes are reported to incur $< 6\%$ power overhead compared to optimal signoff modes.

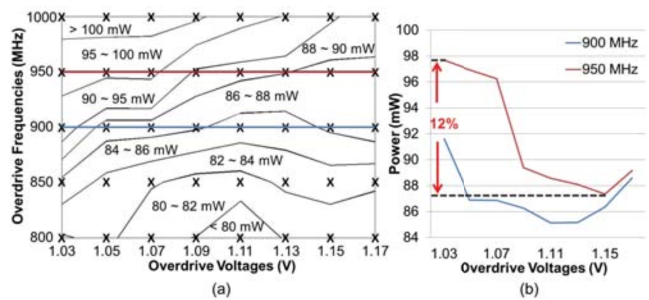

Fig. 8. Average power of circuits signed off at the same nominal mode (500 MHz, 0.9V) but 40 different overdrive modes. Design: *AES* [22]. Technology: foundry 65nm GP. Corner: FF/125°C. Duty cycle of overdrive mode = 10%.

IV. TOLERANCE

We use the term *tolerance* to indicate mechanisms by which timing errors or approximate computation results are permitted, so as to achieve a net improvement of design QOR. Variations will consume design margin and then increase the chance of inaccurate behavior in hardware. Error resilience circuits such as Razor flip-flops [9] [16] can detect and recover from errors at the cost of more stringent hold time requirements. In this section, we describe a "MinRazor" optimization to balance the overheads of error-resilient flip-flops and the overheads of tighter datapath timing constraints, so as to achieve resilience with minimum total overhead. We also note the possibility of gaining additional timing margin by allowing bounded errors in hardware modules.

Example: Low-Cost Design Resilience

It is well-recognized that designing for worst-case conditions incurs considerable power and performance overheads. Resilient designs [8] [9] [16], which can detect and recover from timing errors, enable typical-case optimization and can significantly reduce design margins (e.g., energy) as compared to traditional methodologies. However, large overheads (e.g., 15% energy) are incurred by existing resilient design techniques. For instance, resilient designs require additional circuits to detect and correct timing errors. Further, when there is an error, the additional cycles needed to restore a previous correct state degrade throughput, which diminishes the performance benefit of using resilient designs. The error-detection network also can consume large amount (e.g., up to 9%) of wiring resources. In addition, error-tolerant registers typically require large hold margin, which will lead to power overhead due to short-path padding.

To minimize the resilience overhead and manifest the 'true' benefits of margin reduction from resilient designs techniques, Kahng et al. [13] propose a new methodology for resilient design implementation. The proposed methodology integrates two effective optimization techniques – *selective-endpoint optimization*, and *clock skew (useful skew) optimization* – in an iterative optimization flow which comprehends toggle rate information and the tradeoff between cost of resilience and margin on combinational paths.

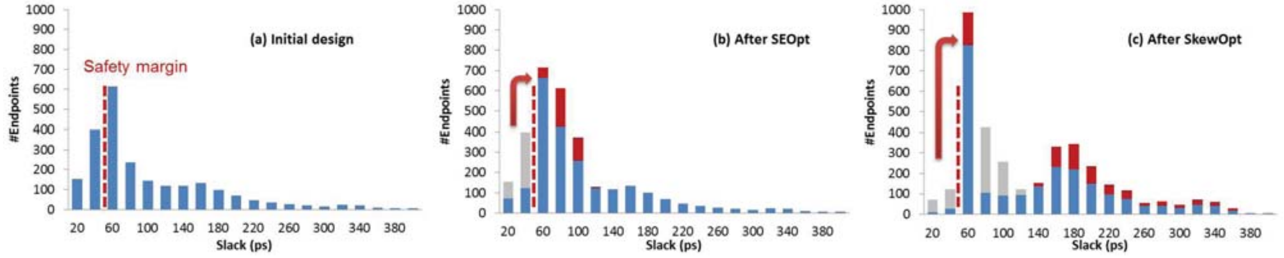

Fig. 9. Endpoint slacks in (a) original design; (b) design after selective-endpoint optimization; and (c) design after useful skew optimization. Red dotted lines indicate required safety margin. Design: FPU (*OpenSPARC T1*) [23]. Technology: foundry 28nm FDSOI.

Figure 9 illustrates the basic idea of the optimization approach. In the initial resilient design (a), a large number of endpoints have timing violations at the target frequency (with respect to the safety margin), and error-tolerant registers or error-masking circuits are used for those endpoints. In (b), the selective-endpoint optimization tightly optimizes a set of selected endpoints to reduce the resilience overheads. In (c), clock skew optimization increases timing slacks of endpoints having timing violations by optimizing the clock-arrival time at individual endpoints, further reducing the resilience overheads.

Experimental results in Figure 10 (*OpenSPARC T1* at 28nm FDSOI technology) show energy comparison with different process variations between resilient designs and conventional ones. In the figure, small, medium and large margins respectively indicate 1σ, 2σ and 3σ for SS corner. Note that resilient designs are signed off at typical case; while conventional ones use worst-case signoff. Results show that an optimized resilient design can achieve up to 10% energy reduction compared to that from a brute-force implementation where error-tolerant flip-flops are simply applied to the most timing-critical paths. Up to 20% energy reduction is achieved in resilient designs as compared to conventional pure-margin designs. The results also show that with larger process variation, resilient designs with brute-force implementation have larger energy cost mainly due to throughput degradation (e.g., FPU and EXU), while the optimized designs are able to jointly minimize the number of error-tolerant flip-flops and error rate, thus achieving greater improvement over brute-force. In addition, the additional circuits for error detection typically cause large area overhead in resilient designs. This example also shows that an optimized implementation is able to significantly reduce such area overhead (e.g., by an average of 45%).

Kahng et al. also study the energy reduction of resilient designs in an adaptive voltage scaling context. Figure 11 compares the energy of conventional designs and resilient designs. The conventional designs are implemented at several supply voltages with pure-margin insertion. The resilient designs are implemented with both the brute-force implementation and the proposed optimization (denoted as "CombOpt"). Results show that optimized resilient designs are able to achieve significant energy reductions with voltage scaling. This is because the optimization comprehends the toggle information and tradeoff between power consumption on combinational cells and error-tolerant registers; this results in less energy penalty from throughput degradation and additional circuits. Therefore, reduced energy is achieved at lower supply voltages.

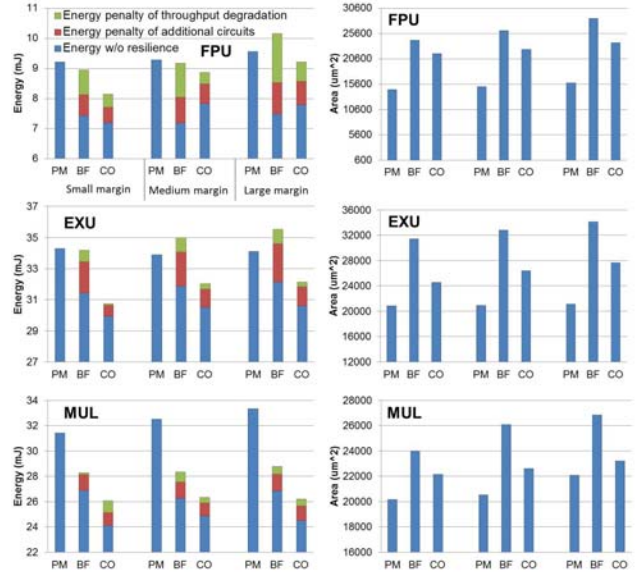

Fig. 10. Energy and area results from different implementation methodologies – pure-margin (PM), brute-force (BF) and proposed optimization (CO).

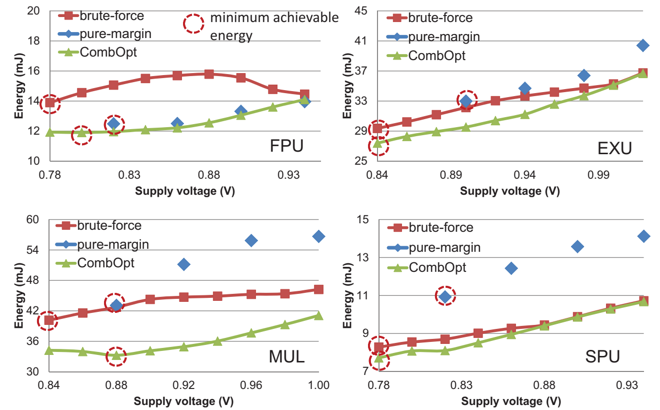

Fig. 11. Energy consumption with voltage scaling, and minimum achievable energy for each method. Design: *OpenSPARC T1 [23]*. Technology: 28nm FDSOI.

Example: Approximate Arithmetic

Approximation can increase performance or reduce power consumption with a simplified or inaccurate circuit in application contexts where strict requirements are relaxed. For applications related to human senses, approximate arithmetic can be used to generate sufficient results rather than absolutely accurate results. Approximate design exploits a tradeoff of accuracy in computation versus performance and power as illustrated in Figure 12. However, required accuracy varies according to applications, and 100% accurate results are still required in some situations. As a final example in this overview of "modeling, margining and tolerance" approaches, we point out an *accuracy-configurable approximate (ACA) adder* for which accuracy of results is configurable during runtime [12].

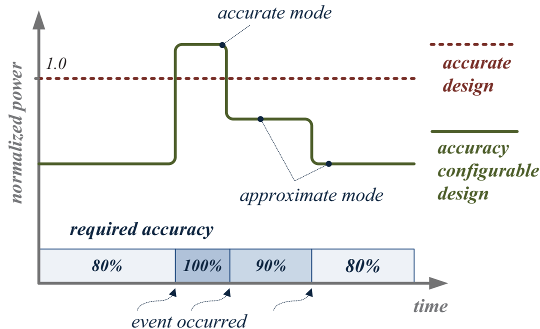

Fig. 12. Power benefits from accuracy-configurable design.

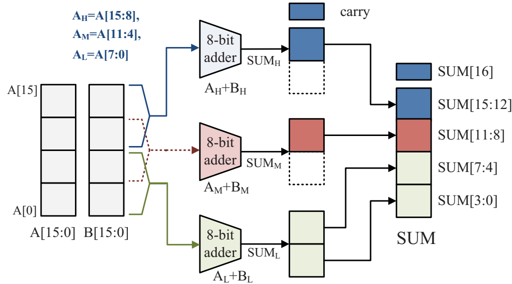

Fig. 13. Implementation of the ACA adder – 16-bit adder case.

The general implementation of ACA is shown in Figure 13. As detailed in [12], the ACA adder can adaptively operate in both approximate (inaccurate) mode and accurate mode (e.g., in an accuracy-configurable application). With its throughput improvement and total power advantages over conventional adder designs, ACA offers yet another dimension in which achievable tradeoffs between performance/power and design quality can be extended. Figure 14 depicts how the ACA adder achieves approximately 30% power reduction (in its approximate modes 2, 3 and 4) versus the conventional pipelined adder (accurate mode 1) when executing SPEC benchmarks [26] with a relaxed accuracy requirement.

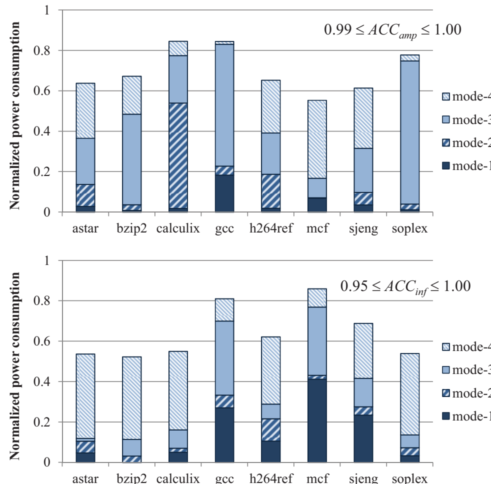

Fig. 14. Normalized power consumption versus conventional pipelined design when accuracy requirement is varied uniformly over the interval 0.99 $\leq ACC_{amp} \leq$ 1.00 (accuracy for amplitude) and 0.95 $\leq ACC_{inf} \leq$ 1.00 (error significance as Hamming distance). Mode 1 is the accurate mode, and Modes 2-4 have different levels of accuracy relaxation.

ACKNOWLEDGMENTS

Many thanks are due to Tuck-Boon Chan, Wei-Ting Jonas Chan, Jiajia Li and Siddhartha Nath for their invaluable help with this paper. Thanks are also due to other coauthors of works cited here. To the extent that my own cited publications have been excerpted, I thank copyright holders for their indulgence.

REFERENCES

[1] J. R. Black, "Electromigration Failure Modes in Aluminum Metallization for Semiconductor Devices", *Proc. of IEEE* 57(9) (1969), pp. 1578-1594.
[2] T.-B. Chan, W.-T. J. Chan and A. B. Kahng, "Impact of Adaptive Voltage Scaling on Aging-Aware Signoff", *Proc. DATE*, 2013, pp. 1683-1688.
[3] T.-B. Chan, W.-T. J. Chan and A. B. Kahng, "On Aging-Aware Signoff for Circuits with Adaptive Voltage Scaling", *IEEE TCAS I*, to appear.
[4] T.-B. Chan, S. Dobre and A. B. Kahng, *manuscript*, 2014.
[5] T.-B. Chan and A. B. Kahng, "Tunable Sensors for Process-Aware Voltage Scaling", *Proc. ICCAD*, 2012, pp. 7-14.
[6] T.-B. Chan, A. B. Kahng, J. Li and S. Nath, "Optimization of Overdrive Signoff", *Proc. ASP-DAC*, 2013, pp. 344-349.
[7] W.-T. J. Chan, A. B. Kahng and S. Nath, "Methodology for Electromigration Signoff in the Presence of Adaptive Voltage Scaling", *Proc. SLIP*, 2014, to appear.
[8] M. Choudhury, V. Chandra, K. Mohanram and R. Aitken, "TIMBER: Time Borrowing and Error Relaying for Online Timing Error Resilience", *Proc. DATE*, 2010, pp. 1554-1559.
[9] D. Ernst, N. S. Kim, S. Das et al., "Razor: A Low-Power Pipeline Based on Circuit-Level Timing Speculation", *Proc. MICRO*, 2003, pp. 7-18.
[10] *International Technology Roadmap for Semiconductors*, 2013 edition, http://www.itrs.net
[11] K. Jeong, A. B. Kahng and K. Samadi, "Impacts of Guardband Reduction on Design Process Outcomes: A Quantitative Approach", *IEEE TSM* 22(4) (2009), pp. 552-565.
[12] A. B. Kahng and S. Kang, "Accuracy-Configurable Adder for Approximate Arithmetic Designs", *Proc. DAC*, 2012, pp. 820-825.
[13] A. B. Kahng, S. Kang, and J. Li, "A New Methodology for Reduced Cost of Resilience", *Proc. GLSVLSI*, 2014, to appear.
[14] A. B. Kahng and S. Nath, "Optimal Reliability-Constrained Overdrive Frequency Selection in Multicore Systems", *Proc. ISQED*, 2014, pp. 1-8.
[15] A. B. Kahn "Design Capability Gap", *UCSD CSE Department technical report* CS2013-1002, 2013.
[16] S. Kim, I. Kwon, D. Fick et al., "Razor-Lite: A Side-Channel Error-Detection Register for Timing-Margin Recovery in 45nm SOI CMOS", *Proc. ISSCC*, 2013, pp. 264-265.
[17] V. Mishra and S. S. Sapatnekar, "The Impact of Electromigration in Copper Interconnects on Power Grid Integrity", *Proc. DAC*, 2013, pp. 1-6.
[18] B. Geden, "Understand and Avoid Electromigration (EM) & IR-drop in Custom IP Blocks", *white paper*, Synopsys, Inc., 2011.
[19] A. Biddle, "Design Solutions for 20nm and Beyond", *white paper*, Synopsys, Inc., 2012.
[20] GLOBALFOUNDRIES, *GF 14nm XM (eXtreme Mobility) process*. http://www.globalfoundries.com/technology-solutions/leading-edge-technologies
[21] GLOBALFOUNDRIES, *GF 20nm LPM (low power mobility) process*. http://www.globalfoundries.com/technology-solutions/leading-edge-technologies/20lpm
[22] OpenCores. http://opencores.org
[23] OpenSPARC T1. http://www.oracle.com/technetwork/systems/opensparc/opensparc-t1-page-1444609.html
[24] Samsung Electronics Inc., *Samsung 32nm HKMG (high-k metal gate) process*. http://www.samsung.com/global/business/semiconductor/foundry/process-technology/32-28nm
[25] Samsung Electronics Inc., *Samsung 45nm process*. http://www.samsung.com/global/business/semiconductor/foundry/process-technology/40-45nm
[26] http://www.spec.org/cpu2006
[27] Taiwan Semiconductor Manufacturing Company, *TSMC 20nm process*. http://www.tsmc.com/english/dedicatedFoundry/technology/20nm.htm
[28] Taiwan Semiconductor Manufacturing Company, *TSMC 28nm process*. http://www.tsmc.com/english/dedicatedFoundry/technology/28nm.htm
[29] Taiwan Semiconductor Manufacturing Company, *TSMC 40nm process*. http://www.tsmc.com/english/dedicatedFoundry/technology/40nm.htm
[30] United Microelectronics Corporation, *UMC 28nm process*. http://www.umc.com/English/process/a.asp