

Containing the Nanometer “Pandora-Box”: Cross-Layer Design Techniques for Variation Aware Low Power Systems

Georgios Karakonstantis, *Member, IEEE*, Abhijit Chatterjee, *Fellow, IEEE*, and Kaushik Roy, *Fellow, IEEE*

Abstract—The demand for richer multimedia services, multifunctional portable devices and high data rates can only be envisioned due to the improvement in semiconductor technology. Unfortunately, sub-90 nm process nodes uncover the *nanometer Pandora-box* exposing the barriers of technology scaling—*parameter variations*, that threaten the correct operation of circuits, and increased *energy consumption*, that limits the operational lifetime of today’s systems. The contradictory design requirements for low-power and system robustness, is one of the most challenging design problems of today. The design efforts are further complicated due to the heterogeneous types of designs (logic, memory, mixed-signal) that are included in today’s complex systems and are characterized by different design requirements. This paper presents an overview of techniques at various levels of design abstraction that lead to low power and variation aware logic, memory and mixed-signal circuits and can potentially assist in meeting the strict power budgets and yield/quality requirements of future systems.

Index Terms—Logic, low power, memory, mixed-signal design, variation tolerant design, voltage scaling.

I. INTRODUCTION

TODAY, we live in the era of miniaturization where there is an increasing demand for small multifunctional devices able to support numerous diverse services [1]. To support all the diverse applications, today’s complex systems consist of heterogeneous components that must meet tight energy and performance requirements [2]. The integration of such complex systems is possible due to aggressive technology scaling that has enabled increased processing capabilities/performance, lower energy consumption per transistor, and lower die cost [1], [2]. However, the decreasing size of devices makes their fabrication more difficult, leading to fluctuations in transistor

length (L), width (W), flat-band voltage, and oxide thickness (T_{ox}) that pose a threat on the correct functionality of each subsystem. In general, parameter variations can be classified into *inter-die* and *intra-die* variations [3]–[5]. *Inter-die* variations refer to parameter variations that have similar parameter values across a single die. Such variations include gate-length variations due to fluctuations in the exposure time during fabrication, variations in thin-oxide, flat-band voltage, and width of transistors. On the other hand, *intra-die* (or random) variations refer to parameter variations in each transistor within a single die. This further complicates the design and modeling of such variations since a large number of random variables are required to represent them [5]. Sources of such variations include fluctuations of geometric dimensions due to lithographic and etching techniques used during fabrication and variations in the number and location of dopant atoms in the channel region, resulting from lithography, etching, and chemical mechanical polishing (CMP). In turn, such variations result in large spread in the threshold voltage (V_{th}) of transistors that affect the various components (digital, analog, memory) in different ways, complicating overall system design. In digital logic for instance, parameter variations have an unpredictable impact on the speed/power-consumption of devices that can ultimately lead up to 30% in delay variation [4]. Such delay variations can result in delay failures due to incomplete computations, degrading yield/output-quality of logic functions [3], [4]. On account of such fluctuations, memories experience stability failures [6], while mixed-signal systems suffer significant loss of yield or significant loss of performance due to increased design guardbands. Note that apart from parametric variations, temporal variations (environmental and aging variations) [7], [8] can also threaten the circuit functionality by leading to delay increase over time. Although this paper focuses on parametric tolerance, the presented mechanisms can be utilized for addressing temporal variations as well.

Depending on the type of design, different techniques have been proposed that try to ensure correct operation under variations [9], [10]. In general, a conservative design approach (i.e. scaling-up voltage or upsizing logic gates), or adding extra hardware (for example, redundant columns/rows in memories) can be beneficial for meeting the various design constraints in the nanometer regime. However, such techniques come at the cost of increased die area and do not allow full utilization of performance gains obtained by technology scaling. More significantly, such techniques further increase power consumption which poses another design challenge in scaled technologies. Conventional power management techniques (supply voltage

Manuscript received October 22, 2010; revised January 24, 2011; accepted February 05, 2011. Date of publication May 12, 2011; date of current version May 25, 2011. This work was supported by the FCRP-MARCO Gigascale Research Center (GSRC) under Grant 2009-DT-2049, by the National Science Foundation under Grant CCF 0916926, and by Qualcomm, Inc., San Diego, CA.

G. Karakonstantis was with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA. He is now with Swiss Federal Institute of Technology (EPFL), Lausanne, CH-1015, Switzerland (e-mail: georgios.karakonstantis@epfl.ch).

K. Roy is with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: kaushik@purdue.edu).

A. Chatterjee is with the Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: chat@gatech.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JETCAS.2011.2135590

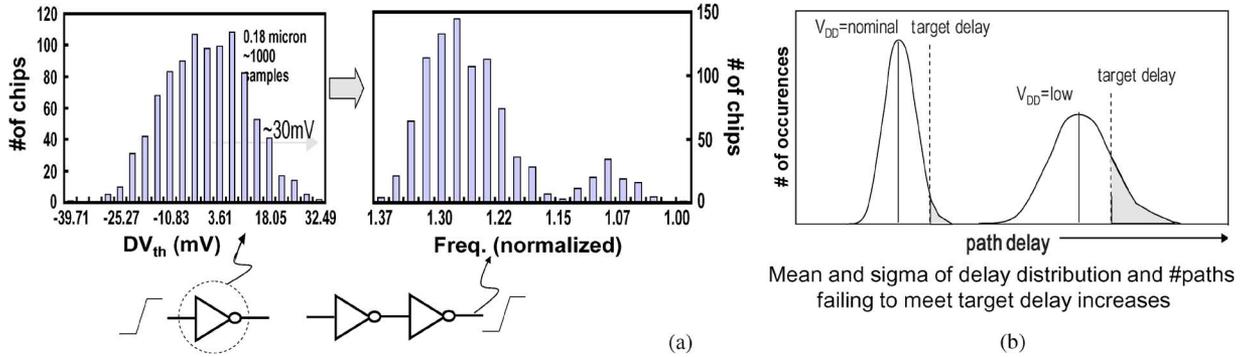


Fig. 1. (a) V_{th} variation and corresponding frequency variation. (b) Impact of supply voltage scaling on path delay distribution.

scaling, power gating, multi- V_{DD} and V_{TH} designs) might reduce power but further magnify the problems associated with process induced variations [2], [5].

It is evident from the above discussion that variation tolerance and low power design techniques may have contradictory requirements (i.e. one requires voltage up-scaling and the other voltage down-scaling) and hence, they should not be addressed in isolation. Consequently, recent variation tolerant and low-power design methodologies have been focused on addressing the two issues simultaneously at different levels of design abstraction. Proposed approaches can be classified based on: i) the type of designs that they target—logic, memory, mixed-signal and ii) the type of applications that they target—general purpose or application-specific. This paper presents an overview of some representative techniques that can potentially lead to the design of low power and variation tolerant systems. Techniques that target general purpose logic/systems include RAZOR [11], [12], CRISTA [14], [15] and stochastic processors [17]–[19], while approaches based on algorithm-architecture co-design include Algorithmic Noise tolerance (ANT) [22]–[26], Significance Driven Approach (SDA) [27]–[31] and system level power management techniques [33]–[37]. Design of new bit-cells that can isolate the conflicting requirements of read and access failures in embedded memories can potentially help in improving memory yield significantly [44]–[47]. Finally, techniques for mixed signal include the design of power-conscious built-in test and tuning algorithms driven by a digital signal processor [59] or autonomous digitally assisted self-tuning mixed-signal systems that react to variations to improve yield [60].

The rest of this paper is organized as follows. Section II focuses on digital logic/system design challenges in the nanometer regime and presents techniques that address process variation tolerance and low power at various levels of hardware stack from circuits to architecture. Section III briefly describes design issues in memory and techniques for addressing them. Section IV overviews design efforts for low power and robust operation of mixed signal systems. Conclusions are drawn in Section V.

II. LOW POWER AND ROBUST LOGIC AND SYSTEM DESIGN

Digital logic is the core of today’s systems and thus needs to operate reliably under the most aggressive power/performance

envelope. In this section the challenges in logic design are discussed and techniques at various levels of design abstraction for addressing them are described.

A. Challenges in Logic Design

As we mentioned earlier, parametric variations and increased power consumption are considered the main barriers of technology scaling [3], [4]. Inter-die and intra-die variations have an unpredictable impact on transistor characteristics such as delay and leakage and are commonly captured by the threshold voltage (V_{th}) of transistors since V_{th} depends on almost all of the transistor parameters [5]. The dependence of gate delay on V_{th} and the possible effect on delay can be represented by the alpha power model [2], where the delay is approximated by

$$D_G = \frac{K \cdot V_{dd}}{(V_{dd} - V_{th})^a} \quad (1)$$

where K and a are technology and circuit dependent parameters. Results in [4] showed a $20\times$ variation in leakage current for a $1.3\times$ variation in performance. This clearly indicates that a circuit designed using nominal V_{th} transistors may not meet the delay target, leading to delay failures [Fig. 1(a)]. Hence, parametric yield of a circuit (probability to meet the desired performance or power specification) is expected to suffer significantly. Apart from parametric variations, nano-scale circuits suffer from static (P_{stat}) and dynamic power (P_{dyn}) consumption due to high transistor densities and high operating frequencies. In other words, parametric variations and increased power consumption unlock the nanometer “Pandora box” unveiling the issues that need to be addressed in the next decades.

Traditionally, fault-tolerant computing and low power design techniques addressed the power and reliability problems in logic functions independently. Several works exist on modeling and addressing the effects of variations at the circuit and logic level. The majority of such approaches such as adaptive body biasing (ABB), adaptive supply voltage scaling (ASVS) and clock tuning [5], [6], [9] aim at providing proper timing margins, while assuming worst-case process conditions. Specifically, at design time, worst case conditions and their effect on the performance of various parts of the circuit are estimated. Based on such estimations transistors are up-sized or voltage is up-scaled to ensure adequate timing margins that subsequently guarantee correct operation under worst case parametric variations [5]. However, worst case conditions are usually rare, and hence, traditional worst-case design approaches lead to

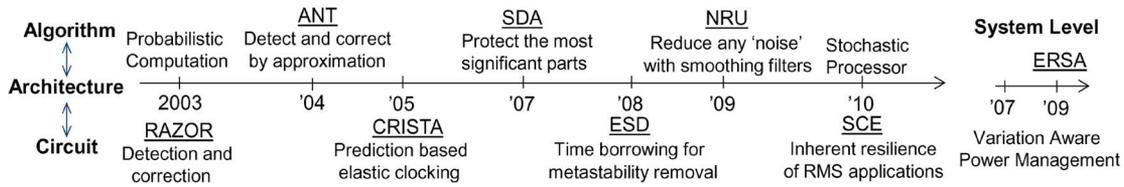


Fig. 2. Timeline of cross-layer techniques for variation aware and low power logic design.

pessimistic design, or in other words, they lead to over-design, not allowing the full utilization of performance gains obtained due to technology scaling. Furthermore, traditional approaches such as transistor up-sizing or voltage up-scaling increase power dissipation. To circumvent the over-design imposed by the traditional worst-case approaches, statistical timing analysis has been widely investigated [5], [11] at the logic level, where a circuit parameter is modeled as a statistical distribution and the circuit is designed to meet a given yield with respect to a target value of the parameter.

Techniques for power reduction try to reduce the switching activity [2], the frequency, or the supply voltage of the circuit. Among them one of the most effective method for power minimization is voltage scaling or over-scaling (VOS) since it leads to large improvements in power consumption due to quadratic dependence of power on supply voltage [2]. VOS extends the concept of voltage scaling beyond the critical voltage value at which the critical path delay (clock period) imposed by architecture and application is just met. Scaling beyond such critical voltage is referred to as VOS and can make the circuit more sensitive to noise [22]. Specifically, VOS increases the delays in all computation paths [Fig. 1(b)] and can result in incorrect computation of certain operations, leading to drastic degradation of output quality at the application layer. In addition, scaled voltages not only increase the delay mean but also the standard deviation of the overall path delay distribution, thereby degrading the timing yield. As we mentioned earlier, scaling up the supply voltage reduces variation at the cost of power consumption. It is evident that meeting a desired power specification with certain degree of process tolerance is becoming an extremely challenging task due to the contradictory nature of the two design requirements.

B. Cross-Layer Design Techniques

It would be ideal if techniques at the circuit and logic levels could solve the problems of increased power and those introduced by variations. However, there is significant concern in the semiconductor industry that techniques at a single level of abstraction (such as circuit or devices) might not be sufficient for solving the low power/error tolerance dilemma under a fixed area/power budget [2]. Consequently, there has been a lot of interest in recent years in addressing variations and power consumption by co-designing at various levels of design abstraction (Fig. 2). High level techniques can take advantage of information that is not easily available at the lower levels of abstraction, provide designers feedback about the impact of errors (due to VOS or variations) early in the design cycle and facilitate better design decisions. Interestingly, voltage over-scaling and parametric variations (inter die) are two sides of the same coin. Both of them result in delay failures at the circuit level that translate to

logic errors. The logic errors can be seen as computation errors at the architecture/algorithm level and may translate to output quality degradation at the system/application level. Therefore, it might be effective to address the symptoms (delay failures) of process variations while meeting the strict power budgets of today's systems (by tuning the supply voltage) at high levels of design abstraction. In the next section we present some of the systematic approaches that address both issues simultaneously. Such approaches can be broadly classified into two categories. Techniques that target general purpose applications need to always ensure exact/correct operations. Such techniques address the issues at the circuit and architecture level trying to detect and correct potential errors. On the other hand, techniques that target application-specific circuits such as digital signal processing (DSP) can relax the requirement of exactness and provide trade-offs with the system level metric of quality through algorithm-architecture co-design.

1) *Circuit-Architecture Co-Design*: Researchers have investigated circuit-architecture co-design approaches that are robust with respect to process variations and at the same time, suitable for aggressive voltage scaling.

a) *Timing error detection and correction*: One such technique, called RAZOR [12] uses dynamic detection and correction of circuit timing errors by using a shadow latch. The key idea of Razor is to tune the supply voltage by monitoring the error rate during circuit operation, thereby eliminating the need for voltage margins and exploiting the data dependence of circuit delay. A RAZOR flip-flop double samples the pipeline stages, once with a fast clock and again with a time-borrowing delayed clock (Fig. 3). A meta-stability tolerant comparator then validates the latched values sampled with the fast clock. If there is any occurrence of timing error, a modified pipeline mis-speculation recovery mechanism restores correct program state. Researchers report up to 64% energy savings with only 3% performance penalty due to error recovery with 3% energy overhead. The micro-architecture level implementation shows that the conversion of 8% of the flip-flops into RAZOR flops is sufficient in order to achieve the desired trade-off between error rate and supply voltage. Recently, the RAZORII [13] technique was proposed in order to circumvent the tight timing constraints and complexity of RAZOR flip-flops. In the RAZORII mechanism, the flip-flop detects potential errors and correction is performed through architectural replay.

Authors in [48] proposed two other timing error detection and recovery circuits that lower clock energy and eliminate the metastability problems of RAZOR based designs by utilizing time borrowing latches. Silicon measurements show that such designs can lead to at least 17% voltage scaling resulting in 37% power reduction.

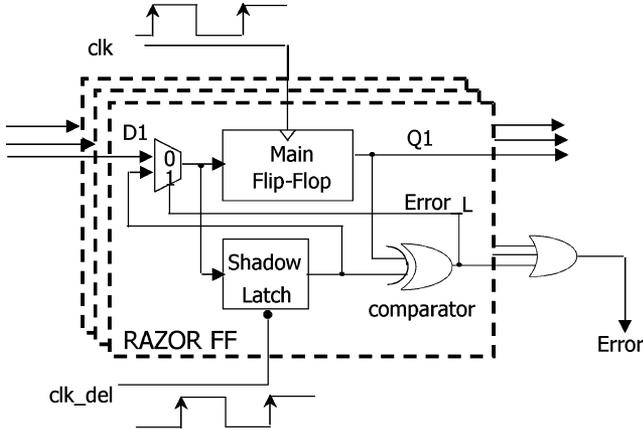


Fig. 3. RAZOR flip-flop.

b) Prediction based elastic clocking: Different from RAZOR, [14] proposes a design time technique called CRISTA to allow VOS while meeting the desired frequency and yield. CRISTA isolates the critical (long) paths of the design, and provides an extra clock cycle for those paths. The CRISTA design methodology ensures the activation of long paths to be rare, thereby, reducing any performance penalty. This allows lowering the supply voltage below specified value and exposes the timing slack of off-critical (short) paths. The concept of CRISTA is shown in Fig. 4(a) with an example of three pipelined instructions where the second instruction activates the critical path. CRISTA can be performed by gating the third clock pulse during the execution of the second instruction for correct functionality of the pipeline at scaled supply voltage. The CRISTA design methodology is applied to random logic by hierarchical Shannon expansion [14] and gate sizing [Fig. 4(b)]. Multiple expansions reduce the activation probability of the paths. In the example shown in Fig. 4(b), critical paths are restricted within CF_{53} (by careful partitioning and gate sizing). The critical paths are activated only when $x_1 x_2 \bar{x}_3 = 1$ with activation probability of 12.5% assuming that each signal can be logic '1' 50% of the time. CRISTA allows VOS improving power consumption by $\sim 40\%$ with only 9% area and small throughput penalty for a two-stage pipelined ALU [14] compared to the standard approach for designing circuits. Note that CRISTA can be applied to circuits as well as at the micro-architectural level.

Elastic-clock latency execution units [15] are example of CRISTA based circuits. Input patterns that activate the critical path are predicted using a low overhead decoder and when they occur, are allocated an additional clock cycle for correct completion of computation, while the rest of input patterns are allotted a single clock cycle. The variable latency units resulted in over 45% power savings while their application in a 5-stage pipeline resulted in 4% penalty for various SPEC2K benchmarks.

Trifecta [16] is a variable latency processor based on CRISTA that completes common-case sub-critical path operations in a single cycle but uses two cycles when the critical (long) paths are exercised. Trifecta increases slack for both single and two-cycle operations and offers a unique advantage under process

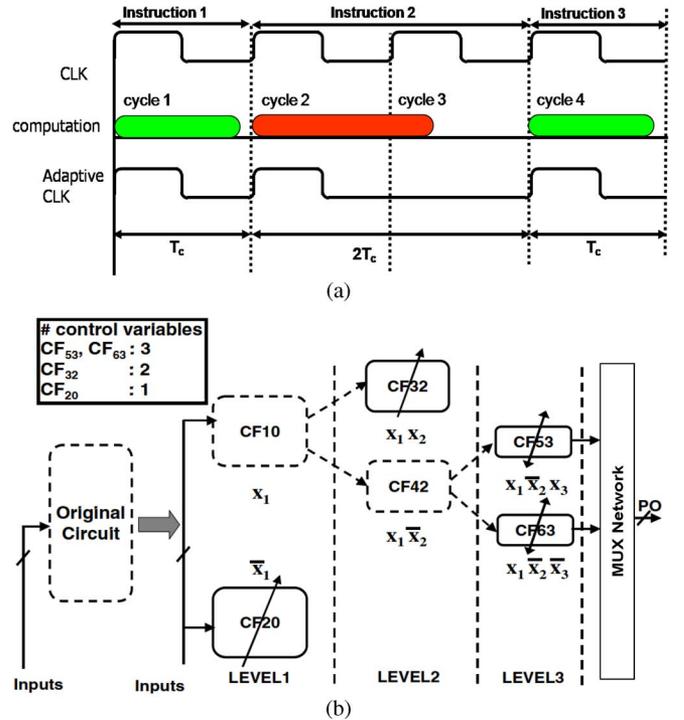


Fig. 4. (a) CRISTA Timing diagram. Long paths are reactivated rarely and they are evaluated in two cycles. (b) Shannon expansion based CRISTA design methodology.

variations. Authors in [16] applied this technique to the critical pipeline stages of a superscalar out-of-order and a single issue in-order processor, namely the instruction issue and execute portions, respectively. The experiments show that the rare 2-cycle operations result in a small decrease (5% for integer and 2% for floating point benchmarks of SPEC2000) in IPC. However, the increased delay slack causes an improvement in yield-adjusted throughput by 20% (12.7%) for an in-order (out-of-order) processor configuration. Note that CRISTA is well suited for pipeline stages that have been designed to be little unbalanced, and the longer stage(s) are designed using the concept of CRISTA. This leads to efficient VOS, with no impact on the frequency of operation (albeit, there is little throughput penalty due to rare/occasional pipeline stalling).

The advantages of elasticity in tolerating errors in synchronous and asynchronous designs are also mentioned by other researchers in [49]–[52].

2) Algorithm-Architecture Co-Design: The above presented techniques such as RAZOR and CRISTA attempt not only to estimate but also to correct any potential error providing always correct outputs. However, in application specific systems such as DSP it is possible to carry out *useful* computations that can lead to *good-enough* output quality rather than dealing with results that are always correct. Such an idea, known as graceful degradation was introduced by Breuer in 1967 [21] and influenced recent design techniques that trade-off *quality-of-results* for tolerating potential errors induced either by parametric variations or VOS.

a) Algorithmic Noise Tolerance (ANT): One such technique is ANT [22] which is shown in Fig. 5. An ANT based

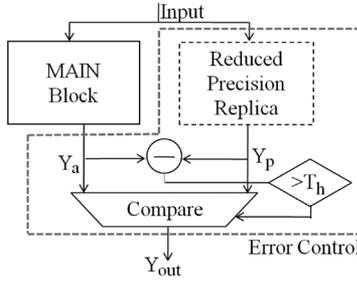


Fig. 5. Algorithmic noise tolerance (ANT) scheme.

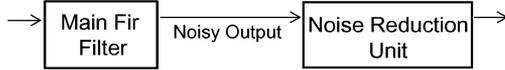


Fig. 6. Noise reduction based technique (NRU).

system consists of a main processing unit that correctly computes most of the time, but its output ($Y_a[n]$) is susceptible to parametric variation and VOS induced errors. The main idea of ANT is the correction of these errors by an estimator/error-control block that produces a statistical replica $Y_p[n]$ of the error-free main block output $Y_{out}[n]$. The main challenge in ANT-based systems is to discover a low-complexity estimator with a much smaller critical path delay. This eventually ensures that the estimator output is error-free even though the main block may exhibit timing errors, affected by parametric variations or VOS.

Several estimation techniques have been proposed and utilized in the design of ANT-based systems. These include prediction-based error correction (PEC), adaptive error cancellation (AEC), and reduced precision redundancy (RPR) [22]–[26]. Another basic element of an ANT based system is the decision block that is nothing other than a comparator that compares the difference between the outputs obtained from the main block $Y_p[n]$ and from the estimation block $Y_a[n]$. When the difference between the two outputs exceeds a threshold- T_h (pre-specified at design time), the error-control block declares an error. In the event of an error, the error-control block selects the predictor output $Y_p[n]$. The various error-control based ANT techniques (AEC, PEC, PRP) have been applied to popular DSP architectures such as FFT, FIR filter, and Viterbi decoders [22]–[26] leading to $\sim 20\%$ – 50% power savings.

Recently, a noise reduction based technique [32] was proposed, as shown in Fig. 6, mainly targeting broadband frequency-selective filters. Such technique allows errors to affect the main filter and it uses simple smoothing filters [called noise reduction units (NRU)] to compensate for errors at the output of the main filter. Results show that such scheme achieves 7–27 dB performance gain while leading to 11%–52% power savings through VOS.

b) Significance Driven Approach (SDA): SDA [27]–[31] attempts to limit any area overhead by utilizing the inherent error resilience of application specific algorithms. The main idea behind SDA approach lies in the fact that all computations are not equally important in shaping the output response in various DSP applications. For such systems, some computations are critical for determining the output quality (thus called *significant*), while others play a less important role (thus demarked

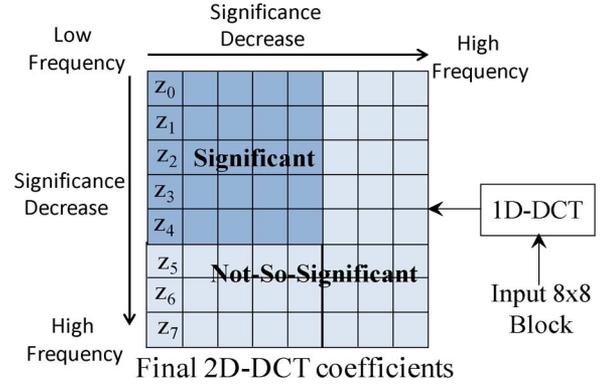


Fig. 7. Significant computations in DCT.

as *less-significant*). The determination of *significant/less-significant* components is performed by a sensitivity analysis that is the first step in SDA methodology. Of course, this classification is only possible based on the quality requirements of the user and application specific characteristics. Based on such classification the *significant* parts are constrained to be computed within a significantly shorter time than the clock period. This is achieved by co-designing algorithm and architecture and imposing tight timing constraints during synthesis. On the other hand, the timing constraints of the less significant ones are less strict and can take longer time (within the clock period) making them susceptible to delay failures. The resultant architecture is characterized by a delay slack between *significant* and *less-significant components* which is utilized to tackle any potential delay increase under VOS and parametric variations. In case of delay errors only the *less-significant* computations are affected and thus have negligible effect on output quality. Other important features of SDA methodology are the maintenance of the same operating frequency at nominal and scaled supply voltage. The SDA approach has been applied to the design of various voltage scalable and variation aware architectures popular in today’s multimedia and signal processing systems such as discrete cosine transform (DCT) [27], color interpolation [28], Fir filters [29], wavelet transform [31], and motion estimation [30].

As an example of SDA, in DCT it was found that the signal energy of the output is concentrated in a few low frequency components (marked as *Significant* in Fig. 7), while most other high frequency components are associated with small signal energy (marked as *Not-So-Significant* in Fig. 7). With this information in mind, an architecture was proposed in [27] that computes the high-energy components of the final DCT matrix faster than the low-energy parts. Results show that even under large process variation and VOS (1 V down to 0.8 V), there is a gradual degradation of image quality (33 dB down to 23 dB in peak-signal-to-noise ratio (PSNR)) with considerable power savings (41% to 62%) compared to existing implementations which fail to operate under such conditions as shown in Fig. 8.

c) Probabilistic computation and stochastic processors: The above techniques (some (ANT, NRU) based on redundancy and other (SDA) on application specific characteristics) in a way, attempt to “correct” potential errors and provide an acceptable output quality. However, there are several works which attempt to take advantage of the inherent resilience of various algorithms and instead provide *good enough* systems

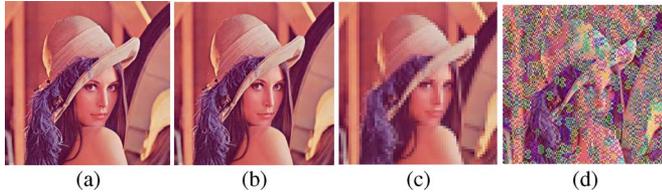


Fig. 8. Output image at (a) nominal conditions; (b), (c) SDA DCT at scaled V_{dd} /slow corner; and (d) conventional design under VOS/PV.

without trying to correct any potential errors rather allowing them to occur. The concept of Probabilistic CMOS or PCMOs, wherein each transistor and logic gate displays a probabilistic rather than deterministic behavior, was proposed as an energy efficient alternative to traditional deterministic computational models [17]. This has led to various probabilistic and approximate computation research, summarized in [18]. Such methods take advantage of the statistical behavior of nano-scale devices and circuits and target emerging killer probabilistic applications such as Recognition, Mining and Synthesis (RMS). Such algorithms typically use iterative and successive refinement techniques, which give them self-healing characteristics, since subsequent iterations can correct errors introduced in the previous iterations. Frequently, these algorithms do not have a single “perfect” result; instead, they may produce any one of many solutions that are equally acceptable.

One such approach is scalable effort (SCE) [20]. The basic idea of SCE is to identify mechanisms at each level of design abstraction (circuit, architecture and algorithm) that can be used to vary the computational effort expended towards the generation of the correct/exact result, and expose them as control knobs in the implementation. The approach was demonstrated through the design of an energy efficient scalable-effort hardware implementation of Support Vector Machines (SVM) for a popular machine learning algorithm. As an extension of such work, a computing platform, referred as stochastic processor, has been proposed in [19]. Scalability in this case is achieved by exposing to the application layer multiple functional units that differ in their architecture. The various available functional units (e.g. adders) achieve different degrees of voltage/frequency scaling at different error rates. Based on performance, power and quality requirements, the application may choose between the available adders while producing outputs that are, in the worst case, stochastically correct.

C. System Level Techniques

It is believed that system level techniques can lead to large power savings and allow better design decisions since they have access to information that is not available to lower levels and at the block level. Therefore, various techniques were proposed recently for addressing power and parametric variations at the system level.

1) *Variation Aware Power Management*: One of the most widely used power reduction technique at the system level is power management [33]. However, conventional power management schemes, designed using nominal power characteristics, can result in substantial power wastage and fail under parametric variations [2], [33]. Therefore, effective power management techniques that can lead to significant energy savings

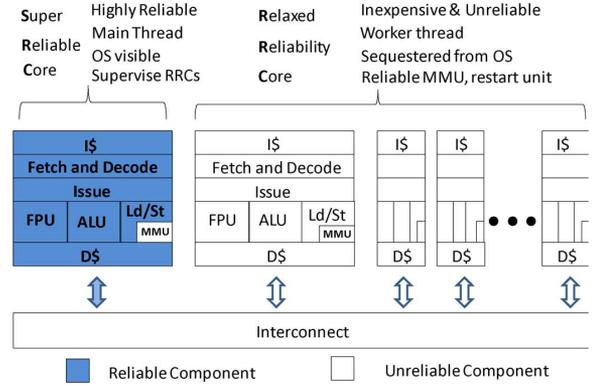


Fig. 9. Error resilient system architecture (ERSA).

even in the presence of parametric variations were proposed in [33]–[37]. Such methods, in general, exploit the variable workload of applications over time in order to adjust the voltage in the various power domains on chip while considering parametric variations. In addition software optimizations that exploit data locality, existing in almost all DSP and multimedia applications, are applied in order to achieve low power and robust operation in embedded processors [2].

2) *Error Resilient System Architecture (ERSA)*: Another example of system level techniques is ERSA [38]. ERSA is a multi-core platform that achieves error resilience to high-order bit errors and in addition controls the low-order bit errors using a combination of three basic ideas: asymmetric reliability, software optimizations and error-resilient algorithms. As shown in Fig. 9, it consists of a limited number of Super Reliable Cores (SRCs) together with a large number of less reliable cores, referred to as Relaxed Reliability Cores (RRCs). The key idea is that the computations of an application are divided into control-intensive (must be error-free) and data-intensive (errors can be tolerated). By applying the idea of asymmetric reliability, the control-related code is assigned to SRCs and the computation intensive code to RRCs. By doing so, ERSA achieves the minimization of processor cores that require high reliability, avoiding conservative system design. This approach was applied to implementation of various probabilistic applications such as K-Means clustering and low density parity (LDPC) decoding. It was demonstrated that even at very high error rates of 2×10^{-4} error/cycle/core, 90% or better accuracy of output results is maintained with minimal impact on execution time.

III. LOW POWER AND ROBUST MEMORY DESIGN

Apart from logic computations, a ubiquitous part of today’s systems is memory [2]. In this section the effect of parametric variations on memories are discussed and novel design techniques are described.

A. Challenges in Memory Design

The inter-die parameter variations, coupled with intrinsic on-die variation in the process parameters (e.g., V_{th} , L , W) of transistors lead to mismatches in the strength of different transistors in a SRAM bit-cell resulting in the failure of SRAM array. In general, parametric failures can take the form of [9]:

- destructive read (i.e., flipping of the stored data in a cell while reading—known as read failure, R_F);

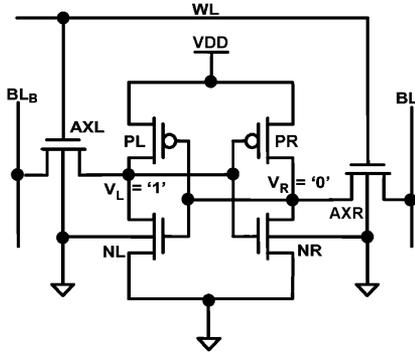


Fig. 10. 6T SRAM bit-cell.

- unsuccessful write (inability to write to a cell—defined as write failure, W_F);
- an increase in the access time of the cell resulting in violation of the delay requirement—defined as access time failure, A_F ;
- destruction of the cell content in standby mode with the application of lower supply voltage (primarily to reduce leakage in standby mode)—known as hold failure, H_F .

The overall failure probability of an SRAM cell is modeled as the union of the individual parametric failures [10]. Interestingly, random variations (such as RDF and LER) impact SRAM cells more than logic since the transistors are of minimum size in order to meet the higher density requirement. Apart from parametric variations, power (dynamic and most importantly static) is another challenge in SRAM arrays. Voltage scaling can effectively reduce power (dynamic and static) but reduces cell stability [36].

Several design techniques that address the requirements of low power and robustness to parametric variations in memories have been proposed. Conventionally, circuit techniques such as sizing of cell transistors and adaptive body biasing (ABB) [7], were employed for the design of low power and robust six transistor (6T) cells. However, the contradictory design requirement of read and write operation in conventional 6T bit-cell (shown in Fig. 10) complicates such approaches. For instance, one can reduce the size of access transistor to decrease the read failure probability (P_{RF}) but this increases the write failure probability (P_{WF}) and access failure probability (P_{AF}). On the other hand, reducing the strength of the pull-up transistors diminishes P_{WF} but increases P_{RF} and P_{AF} .

B. Bit-Cell Configurations

To circumvent the fundamental low reliability of 6T bit-cell, new bit-cells (8T and 10T) were proposed that basically isolate the read from the write for achieving better stability, while enabling low voltage operation. Fig. 11 shows an 8T bit-cell configuration. Adding two FETs to a 6T bit-cell provides a read mechanism that does not disturb the internal nodes of the cell [43]. This requires separate read and write word-lines (RWL, WWL) and can accommodate dual-port operation with separate read and write bit-lines as shown in Fig. 11. Without read disturbs during read, the worst-case stability condition for an 8T cell equals that of two cross-coupled inverters, which provides larger static noise margin (SNM) $\sim 3\times$. A dramatic stability

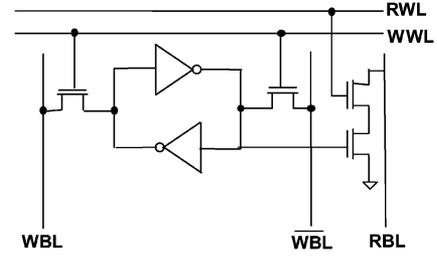


Fig. 11. 8T SRAM bit-cell.

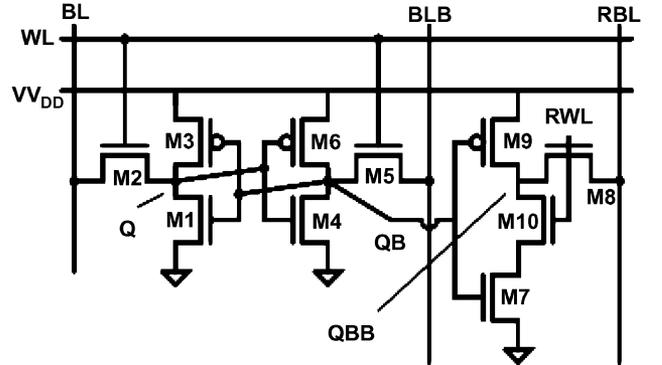


Fig. 12. 10T SRAM bit-cell.

improvement can thus be achieved without a trade-off in performance since a read access is still performed by two stacked nFETs. The area penalty of this bit-cell is shown to be $\sim 30\%$.

To further improve cell stability, a 10T bit-cell was proposed in [44] and is shown in Fig. 12. It uses transistors M7–M10 to improve Read SNM by buffering the stored data during a read access. Thus, the worst-case SNM for this bit-cell is the Hold SNM related to M1–M6, which is the same as the 6T Hold SNM for same-sized M1 through M6. M10 significantly reduces leakage power relative to the case where it is excluded. Another 10T structure uses the Schmitt Trigger (ST) concept [45], [46] that improves both read and write concurrently by improving the inherent stability of the bit-cell. In the ST bit-cell, transistors PL–NL1–NL2–AXL2 form one ST inverter while PR–NR1–NR2–AXR2 form the other ST inverter (Fig. 13). Feedback transistors AXL2/AXR2 raise the switching threshold of the inverter during the $0 \rightarrow 1$ input transition providing the Schmitt trigger action. With same number of NMOS in read path, the read current in ST cell is the same as the 6T cell. During the write operation, both word-lines are ON. Series connected pull down NMOS results in raising the voltage at node storing “0” which is higher than the corresponding 6T cell. As a result ST cell flips at a much higher bit-line voltage compared to 6T cell giving higher ($\sim 2X$) “write-trip-point”. Thus the ST bit-cell design can improve readability and writeability simultaneously unlike the conventional 6T cell. Note that the above techniques also allow operation at sub-threshold or near sub-threshold region, thus leading to large power savings.

C. Circuit-Architecture Co-Design

Apart from circuit techniques and new bit-cells, architectural techniques such as the addition of redundant columns and rows [2], [9] or the use of parity bits for error detection

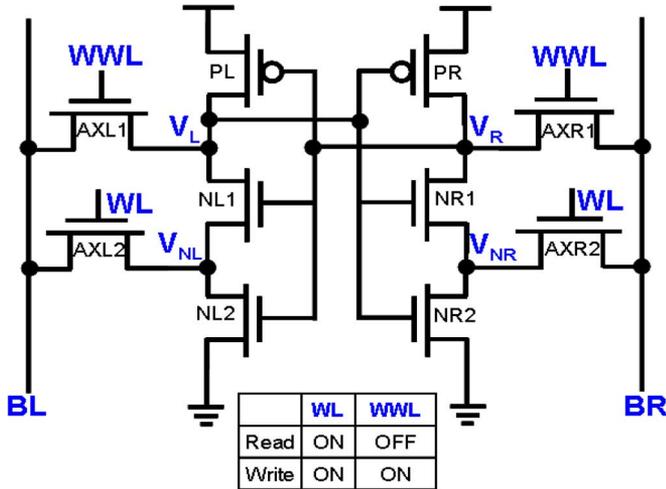


Fig. 13. Schmitt trigger SRAM bit-cell.

and protection (usually reserved for soft errors) [38] have been explored for providing robust memory operation. In an attempt to limit the area overhead of such techniques, memory design techniques have been proposed that exploit the inherent algorithmic resiliency in application specific systems, trading-off power savings and robustness with graceful quality degradation [40]–[42]. One such approach [42] is based on the SDA methodology presented in Section II-C2), where circuit and architectural techniques are combined in order to ensure that significant bits are not corrupted under parametric variations and voltage over-scaling. The basic idea lies behind the observation that 8T SRAM cells are more robust than their 6T counterparts at scaled supply voltages. This feature was exploited to design a hybrid memory for video applications based on a mixture of 8T and 6T SRAM bit-cell configuration, employing a low overhead preferential storage, where the *significant* computations (higher order luma bits) are stored in the more robust 8T-cell memory, while the *less-significant* computations (lower order 5-bits) are stored in a 6T-cell based memory. The fundamental premise of this approach lies in the fact that human visual system (HVS) is sensitive mainly to higher order bits of luminance pixels in video data. As it can be seen in Fig. 14, SDA based SRAM allows operation under VOS (at 600 mV and at slow-fast corner including inter and intra die variations [47]) with insignificant image quality loss as opposed to conventional memory based on 6T bit-cells. Overall, hybrid SRAM results in significant power savings for both read and write (44 ~ 46% at 10 MHz frequency) operations while having only 11% area overhead when the three significant luma bits and all motion vectors are implemented as 8T bit-cells. Note that 600 mV operation of bit-cells implicitly assumes that access times can be met at such low supply voltages. Indeed, in video formats like QCIT [42], the requirement for access time is much less stringent.

IV. ROBUST AND POWER AWARE MIXED SIGNAL DESIGN

Apart from logic and memory blocks, the other basic components of today’s portable systems are mixed-signal circuits. In this section the effect of parametric variations on mixed signal are discussed and novel design techniques are described.



Fig. 14. Output video image of 32 kit SRAM (a) array with 6T-only cells at FS corner and $V_{dd} = 600$ mV, PSNR = 15.27 dB. (hybrid array at $V_{dd} = 600$ mV and FS corner, PSNR = 23.61 dB and (c) hybrid array at $V_{dd} = 600$ mV and SF corner, PSNR = 23.90 dB.

A. Challenges in Resilient Mixed-Signal Design

The push towards design of low power mixed-signal/RF circuits and systems has created new technological challenges particularly in the presence of process variations due to the use of scaled CMOS technologies. As supply voltage scales downwards, it becomes increasingly difficult to maintain performance in the presence of variability induced mismatch, reduced voltage swing and other constraints [53]–[56]. Techniques that allow performance metrics to be maximized in low voltage designs are described in [53]–[58]. There is general consensus that active tuning (post-manufacture) is an effective way of dealing with the effects of process variations. In [58], the use of a measurement and control loop is used to improve the performance of a ring oscillator in the presence of process variations. Extending such techniques to generic mixed-signal/RF circuits across a range of complex specifications requires the use of on-chip test, measurement and tuning infrastructure and is a subject of ongoing research. The key challenges in variation-resilient mixed-signal design are:

- 1) Design of *tunable* mixed-signal/RF modules (data converters, amplifiers, mixers, PLLs, etc). Specifically, we desire to perform post-manufacture tuning of mixed-signal modules in such a way that their performance specifications lie within acceptable limits as determined by *system level specifications*. The tuning “knobs” (current/voltage values at specified nodes) must be selected in such a way that they allow the specifications of the module to be tuned as *independently* as possible.
- 2) Consideration of *multiple static* (e.g., Gain, IIP3, I-Q mismatch) and *dynamic* (e.g. Phase noise) performance parameters of a mixed-signal/RF system under a unified BIST framework for tuning purposes. Current BIST techniques either target individual specifications or require significant computation time to evaluate multiple specifications of the device-under-test (DUT). For efficient tuning, it is imperative to *evaluate multiple specifications of the DUT in as short a time as possible* [61], [62] (to reduce overall tuning time and to solve the tuning problem across all DUT specifications as opposed to a single specification). Consideration of *power* as a critical metric during the tuning process.
- 3) Power should be co-optimized along with multiple performance metrics *across all the modules of the mixed-signal/RF system*. The presence of multiple analog and digital tuning parameters (knobs) allows *superior solutions* to be obtained as opposed to when only limited tuning “knobs” are available for tuning purposes.

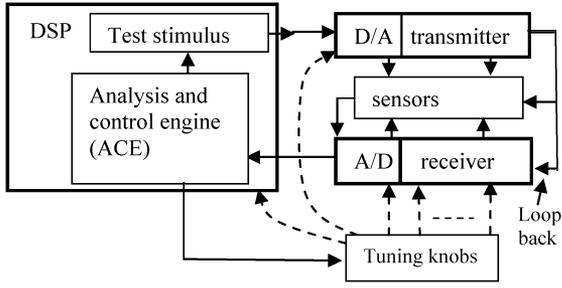


Fig. 15. Test and tuning infrastructure.

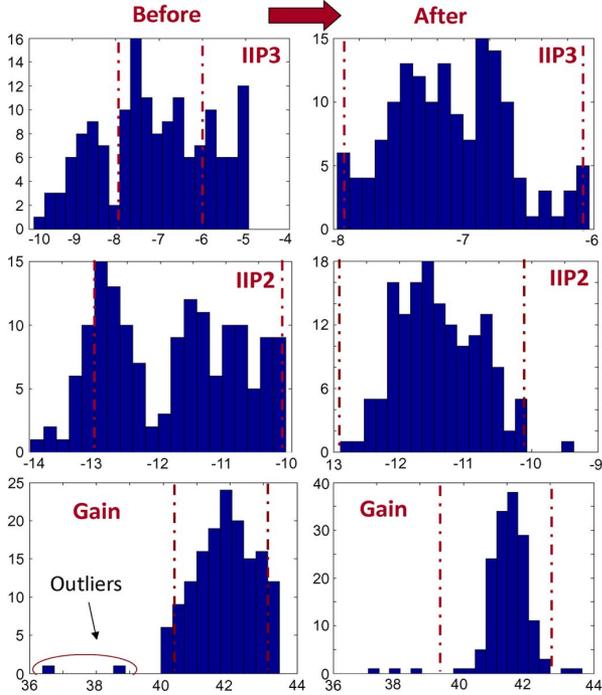


Fig. 16. Yield results for tuning of transmitter using Augmented LaGrange approach.

- 4) Use of a BIST-driven control algorithm that takes results from the applied BIST procedure and changes the values of the analog and digital tuning knobs concurrently to force all the system-level specifications to “move” towards acceptable values. This control algorithm may be implemented using a nonlinear constrained optimizer running on an on-board processor. In specific instances where such a processor is not available, the control algorithm may be implemented using on-chip digital circuitry [63].

B. Mixed-Signal/RF: System Level Tuning

A key challenge is to perform *tuning at the module level while satisfying system-level performance metrics*. As an example, a possible built-in test and tuning architecture for an RF transceiver system is shown in Fig. 15. The transmitter and receiver modules (mixers, LNA, PA are designed with built-in tuning knobs that tradeoff power consumption with performance. Examples of tuning knobs include current and voltage bias control, tunable inductors and capacitors, etc.

(*hardware tuning knobs*) and pre-distortion and post-distortion coefficients (*software tuning knobs*). The transmitter and receiver subsystems are also equipped with built-in sensors for monitoring the quality of signals at intermediate nodes of the RF signal path. While multiple sensors may be used, a single envelope detector at the output of the transmitter can also be used to monitor the health of the transmitter and produces a low frequency signal that is analyzed by the DSP after digitization (the input to the A/D converter of Fig. 15 is multiplexed for built-in test purposes). The test-and-tuning procedures work as follows: The DSP sends a multisine stimulus through the D/A converter to the transmitter which is then “looped back” to the receiver via an external feedback path (through the package or on the DUT load board). The data obtained from the sensor at the output of the transmitter is used to determine *multiple performance metrics* of the transmitter (Gain, IPP2, IIP3, I-Q mismatch, AM-PM distortion, etc.) using *single data acquisition*. The data obtained from the receiver are used to determine similar performance metrics of the receiver. In this context, the data obtained from the sensor allows accurate determination of receiver quality in the presence of *simultaneous* transmitter imperfections. From the test results (obtained in a few milliseconds) software residing in the DSP that implements the analysis and control engine (ACE) determines how to “turn” the hardware and software tuning knobs of the design to improve transceiver performance. At each step, a cost function is evaluated as a weighted mean-square difference between the ideal performance metrics of the transmitter/receiver and the values determined by the built-in test procedure. A nonlinear constrained optimizer running on DSP is used to drive an iterative test-tune-test procedure which terminates when acceptable transceiver performance is achieved.

In specific cases where an on-chip or on-package DSP is not available for running a nonlinear optimizer, a simpler test-tune-test approach can be implemented using on-chip (simplified) analog and digital support circuitry. In [60], a digitally generated stimulus is used to excite analog/RF circuitry and the response of the circuit (downconverted) is digitized into a bit-stream using a 1-bit A/D converter. This bit-stream is analyzed “on-chip” using digital circuitry and a modified SS-LMS algorithm is used to drive the tuning procedure.

Fig. 16 shows the histograms obtained for the Gain, IIP2 and IIP3 specifications of the transmitter of Fig. 15 when an Augmented Lagrange optimizer running on the DSP of Fig. 15 is used to drive the tuning procedure. The tuning is performed in a power-conscious manner as power consumption is formulated within the cost function used for optimization and represents increase of yield from 62% to 96% across all transmitter specifications.

V. CONCLUSION

An overview of several techniques for the design of variation aware and low power systems was conducted. Rather than focusing on a single level of design abstraction, the presented techniques address the contradictory design challenges (power, variations) at various levels of hardware hierarchy. The discussed mechanisms target various types of designs; logic, memory and mixed signal that consist the main elements of today’s systems.

The combination of techniques presented in this paper can effectively allow the design of low power and robust integrated systems in the nano-scale as well as in the post-silicon era.

ACKNOWLEDGMENT

The authors would like to thank S. Ghosh, D. Mohapatra, V. Natarajan, S. Sen, and S. Devarakond for their contributions to the work discussed in Sections III and IV.

REFERENCES

- [1] H. De Man, "Ambient intelligence: Gigascale dreams and nanoscale realities," in *IEEE Int. Solid State Circuits Conf.*, 2005, pp. 29–35.
- [2] J. M. Rabaey, *Low Power Design Essentials*. New York: Springer, 2009.
- [3] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De, "Parameter variations and impact on circuits and microarchitecture," in *IEEE Design Autom. Conf.*, Jun. 2003, pp. 338–342.
- [4] K. A. Bowman, S. G. Duvall, and J. D. Meindl, "Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration," *IEEE J. Solid-State Circuits*, vol. 37, no. 2, pp. 183–190, Feb. 2002.
- [5] A. Shrivastava, D. Sylvester, and D. Blaauw, *Statistical Analysis and Optimization for VLSI: Timing and Power*. New York: Springer, 2005.
- [6] A. Bhavnagarwala, X. Tang, and J. D. Meindl, "The impact of intrinsic device fluctuations on CMOS SRAM cell stability," *IEEE J. Solid-State Circuits*, vol. 36, no. 4, pp. 658–665, Apr. 2001.
- [7] S. Jafar, Y. H. Kim, V. Narayanan, C. Cabral, V. Paruchuri, B. Doris, J. Stathis, A. Callegari, and M. Chudzik, "A comparative study of NBTI and PBTI (charge trapping) in SiO₂/HfO₂ stacks with FUSI, TiN, Re gates," in *Proc. VLSI Circuits*, 2006, pp. 23–25.
- [8] JEDEC Solid State Technology Assoc., "Failure mechanisms and models for semiconductor devices," JEP122-A, 2002.
- [9] J. Tschanz, K. Bowman, and V. De, "Variation-tolerant circuits: Circuit solutions and techniques," in *IEEE Design Autom. Conf.*, 2005, pp. 762–763.
- [10] S. Mukhopadhyay, K. Kang, H. Mahmoodi, and K. Roy, "Reliable and self-repairing SRAM in nano-scale technologies using leakage and delay monitoring," in *IEEE Int. Test Conf.*, Nov. 2005, pp. 1135–1144.
- [11] H. Chang and S. S. Sapatnekar, "Statistical timing analysis considering spatial correlations using a single PERT-like traversal," in *IEEE Int. Conf. Comput.-Aided Design*, Nov. 2003, pp. 621–625.
- [12] D. Ernst, N. S. Kim, S. Das, S. Pant, R. Rao, T. Pham, C. Ziesler, D. Blaauw, T. Austin, K. Flautner, and T. Mudge, "Razor: Circuit-level correction of timing errors for low-power operation," *IEEE Micro*, pp. 10–20, Nov. 2004.
- [13] S. Das, C. Tokunaga, S. Pant, W.-H. Ma, S. Kalaiselvan, K. Lai, D. M. Bull, and D. T. Blaauw, "RazorII: In situ error detection and correction for PVT and SER tolerance," *IEEE J. Solid-State Circuits*, vol. 44, no. 1, pp. 32–48, Jan. 2009.
- [14] S. Ghosh, S. Bhunia, and K. Roy, "CRISTA: A new paradigm for low-power, variation-tolerant, and adaptive circuit synthesis using critical path isolation," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 26, no. 11, pp. 1947–1956, Nov. 2007.
- [15] D. Mohapatra, G. Karakonstantis, and K. Roy, "Low-power process-variation tolerant arithmetic units using input-based elastic clocking," in *Proc. ISLPED*, 2007, pp. 74–79.
- [16] P. Ndai, N. Rafique, M. Thottethodi, S. Ghosh, S. Bhunia, and K. Roy, "Trifecta: A non-speculative scheme to exploit common, data-dependent subcritical paths," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 18, no. 1, pp. 53–65, Jan. 2010.
- [17] K. V. Palem, "Energy aware algorithm design via probabilistic computing: From algorithms and models to Moore's law and novel semiconductor devices," in *Proc. CASES*, 2003, pp. 113–116.
- [18] K. V. Palem, L. N. Chakrapani, Z. M. Kedem, L. Avinash, and K. K. Muntimadugu, "Sustaining moore's law in embedded computing through probabilistic and approximate design: Retrospects and prospects," in *Proc. CASES*, 2009, pp. 1–10.
- [19] S. Narayanan, J. Sartori, R. Kumar, and D. L. Jones, "Scalable Stochastic Processors," in *IEEE DATE*, Mar. 2010.
- [20] V. K. Chippa, D. Mohapatra, A. Raghunathan, K. Roy, and S. T. Chakradhar, "Scalable effort hardware design: Exploiting algorithmic resilience for energy efficiency," in *IEEE Design Autom. Conf.*, 2010.
- [21] M. A. Breuer, "Adaptive computers," *J. Inf. Control*, vol. 11, pp. 402–422, Oct. 1967.
- [22] B. Shim and S. R. Sridhara, "Reliable low-power digital signal processing via reduced precision redundancy," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 12, no. 5, pp. 497–510, May 2004.
- [23] R. Hegde and N. R. Shanbhag, "Soft digital signal processing," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 9, no. 12, pp. 813–823, Dec. 2001.
- [24] L. Wang and N. R. Shanbhag, "Low-power filtering via adaptive error cancellation," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 575–583, Feb. 2003.
- [25] R. A. Abdallah and N. R. Shanbhag, "Error-resilient low-power Viterbi decoder architectures," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4906–4917, Dec. 2009.
- [26] G. Varatkar and N. R. Shanbhag, "Error-resilient motion estimation architecture," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 16, no. 10, pp. 1399–1412, Oct. 2008.
- [27] G. Karakonstantis, N. Banerjee, and K. Roy, "Process-variation resilient and voltage scalable DCT architecture for robust low-power computing," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 18, no. 10, pp. 1461–1470, Oct. 2010.
- [28] N. Banerjee, G. Karakonstantis, J. H. Choi, C. Chacrabarti, and K. Roy, "Design methodology for low power dissipation and parametric robustness through output quality modulation: Application to color interpolation filtering," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 28, no. 8, pp. 1127–1137, Aug. 2009.
- [29] J. H. Choi, N. Banerjee, and K. Roy, "Variation-aware low-power synthesis methodology for fixed-point FIR filters," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 28, no. 1, pp. 87–97, Jan. 2009.
- [30] D. Mohapatra, G. Karakonstantis, and K. Roy, "Significance driven computation: A voltage-scalable, variation-aware, quality-tuning motion estimator," in *IEEE ISLPED*, 2009, pp. 195–200.
- [31] V. Gupta, G. Karakonstantis, D. Mohapatra, and K. Roy, "VEDA: Variations-aware energy efficient discrete wavelet transform," in *IEEE ICCD*, 2010.
- [32] R. Liu and K. K. Parhi, "Noise reduction for low-power broadband filtering," in *Proc. 2009 IEEE ISCAS*, Taipei, May 2009, pp. 1012–1015.
- [33] F. Shearer, *Power Management in Mobile Devices*. New York: Elsevier, 2010.
- [34] N. Bansal, K. Lahiri, A. Raghunathan, and S. T. Chakradhar, "Power monitors: A framework for system-level power estimation using heterogeneous power models," in *IEEE Int. Conf. on VLSI Design*, Jan. 2005, pp. 579–585.
- [35] S. Chandra, K. Lahiri, A. Raghunathan, and S. Dey, "Variation tolerant dynamic power management at the system level," *IEEE Trans. Very Large Scale Integr. (VLSI) Systems*, vol. 17, no. 9, pp. 1220–1232, Sep. 2009.
- [36] S. Garg and D. Marculescu, "System-level process variation driven throughput analysis for single and multiple voltage-frequency island designs," *ACM Trans. Design Autom. Electron. Syst.*, vol. 13, Sep. 2008.
- [37] H. Jung and M. Pedram, "Resilient dynamic power management under uncertainty," in *IEEE DATE*, 2008.
- [38] L. Leem, H. Cho, J. Bau, Q. Jacobson, and S. Mitra, "ERSA: Error-resilient system architecture for probabilistic applications," in *IEEE DATE*, Mar. 2010.
- [39] A. Kumar, J. M. Rabaey, and K. Ramchandran, "SRAM supply voltage scaling: A reliability perspective," in *IEEE ISQED*, 2009, pp. 782–787.
- [40] M. Cho, J. Schlessman, W. Wolf, and S. Mukhopadhyay, "Accuracy-aware SRAM: A reconfigurable low power SRAM architecture for mobile multimedia applications," in *IEEE ASP Design Autom. Conf.*, 2009.
- [41] I. J. Chang, D. Mohapatra, and K. Roy, "A voltage-scalable & process variation resilient hybrid SRAM architecture for MPEG-4 videoprocessors," in *IEEE Design Autom. Conf.*, 2009, pp. 670–675.
- [42] F. J. Kurdahi, A. Eltawil, K. Yi, S. Cheng, and A. Khajeh, "Low-power multimedia system design by aggressive voltage scaling," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 18, no. 5, pp. 852–856, May 2010.
- [43] L. Chang, D. M. Fried, J. Hergenrother, J. W. Sleight, R. H. Dennard, R. K. Montoye, L. Sekaric, S. J. McNab, A. W. Topol, A. D. Adams, K. W. Guarini, and W. Haensch, "Stable SRAM cell design for the 32 nm node and beyond," *VLSI Technol.*, pp. 128–129, 2005.
- [44] B. H. Calhoun and A. Chandrakasan, "A 256 kb sub-threshold SRAM in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 42, no. 3, pp. 680–688, Mar. 2007.

- [45] J. P. Kulkarni, K. Kim, S. Park, and K. Roy, "Process variation tolerant SRAM design for ultra low voltage applications," in *IEEE Design Autom. Conf.*, 2008, pp. 108–113.
- [46] J. P. Kulkarni, K. Kim, and K. Roy, "A 160 mV robust Schmitt trigger based subthreshold SRAM," *IEEE J. Solid-State Circuits*, vol. 42, no. 10, pp. 2303–2313, Oct. 2007.
- [47] IBM Microelectronics Division, "CMOS9SF_V2000-IBM," Apr. 2006 [Online]. Available: <http://www.ibm.com/>
- [48] K. A. Bowman, J. W. Tschanz, N. S. Kim, J. C. Lee, C. B. Wilkerson, S. Lu, T. Karnik, and V. De, "Energy-efficient and metastability-immune resilient circuits for dynamic variation tolerance," *IEEE J. Solid-State Circuits*, vol. 44, no. 1, pp. 49–63, Jan. 2009.
- [49] J. Carmona, J. Cortadella, M. Kishinevsky, and A. Taubin, "Elastic circuits," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 28, no. 10, pp. 1437–1455, Oct. 2009.
- [50] X. Liang and D. Brooks, "Mitigating the impact of process variations on processor register files and execution units," in *Proc. Int. Symp. on Microarchitecture*, 2006, pp. 504–514.
- [51] A. Tiwari, S. R. Sarangi, and J. Torrellas, "Recycle: Pipeline adaptation to tolerate process variation," in *Proc. ISCA*, 2007, pp. 323–334.
- [52] X. Liang, G.-Y. Wei, and D. Brooks, "ReVIVaL: A variation tolerant architecture using voltage interpolation and variable latency," in *Proc. ISCA*, 2008, pp. 191–202.
- [53] M. Steyaert, V. Peluso, J. Bastos, P. Kinget, and W. Sansen, "Custom analog low power design: The problem of low voltage, and mismatch," in *IEEE CICC 1997*, pp. 285–292.
- [54] P. R. Kinget, "Designing analog and RF circuits for ultra-low supply voltages," in *Proc. 37th Eur. Solid State Device Res. Conf.*, pp. 58–67.
- [55] P. Kinget, S. Chatterjee, and Y. Tsvividis, "Ultra-low voltage analog design techniques for nanoscale CMOS technologies," in *EDSSC*, pp. 9–14.
- [56] A. Abidi, G. Pottie, and W. Kaiser, "Power conscious design of wireless circuits and systems," *Proc. IEEE*, vol. 88, no. 10, pp. 1528–1545, October 2000.
- [57] Z. Fu, A. Pappu, and A. Apsel, "Beating the power limit of LC oscillators," in *Proc., Midw. Symp. on Circuits Syst.*, 2007, pp. 441–444.
- [58] X. Zhang and A. B. Apsel, "A process compensated 3-GHz ring oscillator," in *Proc. ISCAS*, 2009, pp. 581–584.
- [59] V. Natarajan, S. Sen, S. Devarakond, and A. Chatterjee, "A holistic approach to accurate tuning of RF systems for large and small multiparameter perturbations," in *IEEE VLSI Test Symp.*, Apr. 2010, pp. 331–336.
- [60] S. Devarakond *et al.*, "Digitally assisted concurrent built-in tuning of RF systems using hamming distance proportional signatures," in *IEEE Asian Test Symp.*, Dec. 2010.
- [61] A. Halder, S. Bhattacharya, and A. Chatterjee, "System-level specification testing of wireless transceivers," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 16, no. 3, pp. 263–276, Mar. 2008.
- [62] R. Voorakaranam, S. S. Akbay, S. Bhattacharya, S. Cherubal, and A. Chatterjee, "Signature testing of analog and RF circuits: Algorithms and methodology," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 54, no. 5, pp. 1018–1031, May 2007.
- [63] Chatterjee, V. Natarajan, S. Devarakond, S. Sen, H. Choi, D. Han, R. Senguttuvan, S. Bhattacharya, D. Lee, A. Goyal, and M. Swaminathan, "Iterative built-in testing and tuning of mixed-signal/RF systems," in *IEEE Int. Conf. Comput.-Aided Design*, Oct. 2009, pp. 319–326.



Georgios Karakonstantis (S'07–M'09) received the Diploma degree in computer and communications engineering from University of Thessaly, Greece, in 2005 and the M.S. and Ph.D. degrees from the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, in 2010.

In the summer of 2008, he was with the Advanced Technology Group, Qualcomm Inc., San Diego, CA, as an intern. Currently he is a research scientist at Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland. His research interests

include cross-layer design techniques for low power and process-variation tolerant application-specific and general-purpose circuits and systems. His work appears in more than 20 refereed journals and conferences.



Abhijit Chatterjee (S'83–M'83–SM'94–F'07) received the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 1990.

He is a Professor in the School of Electrical and Computer Engineering at Georgia Institute of Technology, Atlanta. He currently serves as the Chair of the VLSI Systems and Digital Design (VSDD) Technical Interest Group in the School of Electrical and Computer Engineering. He has published over 350 papers in refereed journals and

meetings and has 12 patents. He is a co-founder of Ardext Technologies Inc., a mixed-signal test solutions company where he served as Chairman and Chief Scientist from 2000–2002. He is currently directing research at Georgia Tech in mixed-signal/RF design and test funded by NSF, SRC, MARCO-DARPA and industry.

Dr. Chatterjee has received the NSF Research Initiation Award in 1993 and the NSF Career award in 1995. He has received four Best Paper Awards and three Best Paper Award nominations. His work on mixed-signal self-healing chips was first cited by the *Wall Street Journal* in 1992. In 1995, he was named a Collaborating Partner in NASA's New Millennium Project. In 1996, he received the Outstanding Faculty for Research Award from the Georgia Tech Packaging Research Center, and in 2000, he received the Outstanding Faculty for Technology Transfer Award, also given by the Packaging Research Center. In September 2007, his research group received the Margarida Jacome Award for work on VIZOR: Virtually Zero Margin RF from the Berkeley FCRP-MARCO Gigascale Research Center.



Kaushik Roy (S'83–M'83–SM'95–F'02) received the B. Tech. degree in electronics and electrical communications engineering from the Indian Institute of Technology, Kharagpur, India, and the Ph.D. degree from the electrical and computer engineering department of the University of Illinois at Urbana-Champaign in 1990.

He was with the Semiconductor Process and Design Center of Texas Instruments Incorporated, Dallas, TX, where he worked on FPGA architecture development and low-power circuit design. He

joined the electrical and computer engineering faculty at Purdue University, West Lafayette, IN, in 1993, where he is currently a Professor and holds the Roscoe H. George Chair of Electrical & Computer Engineering. His research interests include Spintronics, VLSI design/CAD for nano-scale Silicon and non-Silicon technologies, low-power electronics for portable computing and wireless communications, approximate computing, VLSI testing and verification. He has published more than 500 papers in refereed journals and conferences, holds 15 patents, graduated 51 PhD students, and is co-author of two books on *Low Power CMOS VLSI Design* (Wiley, 2000) and *Low-Voltage Low-Power VLSI Subsystems* (McGraw Hill, 2004).

Dr. Roy received the National Science Foundation Career Development Award in 1995, IBM faculty partnership award, ATT/Lucent Foundation award, 2005 SRC Technical Excellence Award, SRC Inventors Award, Purdue College of Engineering Research Excellence Award, Humboldt Research Award in 2010, and Best paper awards at 1997 International Test Conference, IEEE 2000 International Symposium on Quality of IC Design, 2003 IEEE Latin American Test Workshop, 2003 IEEE Nano, 2004 IEEE International Conference on Computer Design, 2006 IEEE/ACM International Symposium on Low Power Electronics & Design, and 2005 IEEE Circuits and System Society Outstanding Young Author Award (Chris Kim), 2006 IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS Best Paper award. He is a Purdue University Faculty Scholar. He was a Research Visionary Board Member of Motorola Labs (2002) and held the M.K. Gandhi Distinguished Visiting faculty at Indian Institute of Technology (Bombay). He has been in the editorial board of *IEEE Design and Test*, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, and IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATED (VLSI) SYSTEMS. He was Guest Editor for Special Issue on Low-Power VLSI in the *IEEE Design and Test* (1994) and IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATED (VLSI) SYSTEMS (June 2000), IET's *IEE Proceedings—Computers and Digital Techniques* (July 2002).