Stochastic Networked Computation

Girish Vishnu Varatkar, Sriram Narayanan, Naresh R. Shanbhag, *Fellow, IEEE*, and Douglas L. Jones, *Fellow, IEEE*

***Abstract*—In this paper, the stochastic networked computation (SNC) paradigm for designing robust and energy-efficient systems-on-a-chip in nanoscale process technologies, where robust computation is treated as a statistical estimation problem is presented. The benefits of SNC are demonstrated by employing it to design an energy-efficient and robust pseudonoise-code acquisition system for the wireless CDMA2000 standard (http://www.3gpp2.org). Simulations in IBM's 130-nm CMOS process show that the SNC-based architecture enhances the average probability of detection ($P_{\rm Det}$) in the presence of process variations by two to three orders of magnitude, reduces power by 31%–39%, and reduces the variation in $P_{\rm Det}$ by one to two orders of magnitude at a typical false-alarm rate of 5% over a conventional architecture. SNC performance in the presence of voltage overscaling and across technology nodes (90, 65, 45, and 32 nm) is also studied.**

***Index Terms*—Code division multiple access (CDMA), low power, nanoscale, process variations, reliability, robust, soft errors, stochastic.**

I. Introduction

MOORE'S LAW, the driving force behind the global semiconductor industry for the past 50 years, is under threat today from artifacts of nanoscale dimensions. Process variations, leakage, and noise in sub-45 nm process technologies [1] are conspiring to make it difficult to reap the benefits of feature size scaling due to reliability concerns. A parallel trend is the growing functional complexity and power of next-generation applications. The result is a power and reliability problem in nanoscale systems-on-a-chip (SOCs). Reliability and power are interlinked problems viewed by the semiconductor industry as the key inhibiters of Moore's law. Not surprisingly, since 2001, the International Technology Roadmap for Semiconductors (ITRS) [2] has stated the achievement of reliability and energy efficiency as important challenges facing the semiconductor industry.

Scaling of feature sizes and voltages in modern semiconductor process technologies have made current and future systems vulnerable to noise [3], process variations [4], leakage, and soft errors due to particle hits [5]. For example, process variations arise from the random dopant fluctuations in a MOS transistor channel [6], which affects the device threshold voltage V_t, and the variations in the width and gate length from the usage of subwavelength lithography. Process variations cause as much as 30% variability in operating frequency in current process technology [1], and this variability is expected to increase to as much as 60% within the next ten years [2]. These variations generate secondary variations in the transistor drive current and gate capacitance, which, in turn, create delay variations in circuits, causing incorrect data to be latched. The resulting logic errors can potentially lead to system failure, especially if these occur in sensitive control-intensive subsystems. In the presence of such increased variations, designing for the worst-case process corner leads to systems consuming substantially higher power than necessary, while designing for the nominal process corner leads to systems that consume less power, but may introduce intermittent errors.

Low-power design [7]–[10] and fault-tolerant computing [11]–[13] are mature areas, but they address the power and reliability problems independently. The idea of jointly optimizing energy and reliability is relatively new. In this latter front, we have proposed a general class of *communications-inspired* techniques such as algorithmic noise tolerance (ANT) [14]. ANT falls into a general class of robust computation techniques referred to as *stochastic computation*, where signal and error statistics are consciously exploited to detect and correct intermittent errors in a nominal-case design, and thereby simultaneously achieve energy efficiency and robustness in nanometer SOCs. These techniques are particularly well-suited for applications where the system-level quality metrics are statistical in nature, such as those in media processing. In fact, a majority of emerging applications are expected to be of this sort. Such applications admit a relaxed definition of correctness of computation by accounting for application-level contextual information. Our past study on ANT and the proposed study exploits this additional degree of freedom afforded by such applications.

Related work includes a statistical analysis of neural computation [15], where the classification property of a feedforward neural network is analyzed in an information-theoretic sense. The work in [15] does not embody any notion of computational errors, or the use of error statistics of the computational fabric to enhance robustness, and is thus very different from the study presented here. The works in [16] and [17] focus on techniques at the logic level. Logic-level techniques tend to have very high overheads [16], which is to be expected as there are very limited degrees of freedom at the logic level. Instead, ANT [14] and the proposed study in this paper takes a system-level view of the problem, and is thus able to exploit additional information (signal and error statistics) and amortize the cost of the overhead.

Manuscript received January 12, 2009; revised April 23, 2009. First published October 06, 2009; current version published September 24, 2010. This work was supported in part by the Gigascale System Research Center, one of five research centers funded under the Focus Center Research Program, a Semiconductor Research Corporation Program, and by the National Science Foundation under Grant CCF 0729092.

The authors are with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: varatkar@illinois.edu; spnaraya@illinois.edu; shanbhag@illinois.edu; dl-jones@illinois.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/TVLSI.2009.2024673

1063-8210/$26.00 © 2009 IEEE

In this paper, we substantially extend the ANT concept into the networked domain via the use of distributed computational units that compute in a stochastic and distributed manner in order to provide robustness and energy efficiency, i.e., the *stochastic networked computation* (SNC) approach [18]. The inspiration for SNC comes from the observation that dense networks [19] are intrinsically robust to the failure/loss of a few constituent nodes because of *information exchange and sharing* in the network. Thus, it may be possible to design a distributed network of highly energy efficient, but unreliable computational nodes that share/exchange enough information to provide a system-level robustness.

This paper presents the key elements of such a computational paradigm and demonstrates its application in designing an energy-efficient and robust pseudonoise (PN) code acquisition system, which is an important power-hungry computational kernel [20], [21] in code division multiple access (CDMA) based wireless systems such as IEEE 802.11 and wideband CDMA (WCDMA).[1] Thus, the SNC design paradigm seeks to embrace randomness and nondeterminism in circuit fabrics as opposed to treating the latter as sources of errors that need to be corrected.

We note that the technique of stochastic communication [22], [23] focuses on robust communication for networks on chip [24], and can be viewed as an appropriate communication fabric for implementing SNCs.

The rest of this paper is organized as follows. In Section II, we describe the SNC approach, followed by its application to design a robust and energy-efficient PN-acquisition system in Section III. Section IV specifies the details of our simulation setup and describes the results. Finally, Section V presents conclusions and future research directions.

II. SNC Paradigm

In this section, we describe the SNC paradigm for the robust and energy-efficient computation. Fig. 1 illustrates the concept of an SNC. Traditionally, the main computation/block generates a desired output $y[n]$, as shown in Fig. 1(a). The centralized nature of this computation makes it vulnerable to localized sources of nonidealities such as particle hits, hot-spots, and across die process variations, and hence, results in hardware errors.

In SNC, the main computation/block is decomposed into M [$M = 4$ in Fig. 1(b)] typically lower complexity computational blocks or *sensors* with a complexity ratio R, where R is the ratio of the complexity of one sensor to that of the main block in Fig. 1(a). Ideally, we would like $R = 1/M$, so that the sum of the sensor complexity equals that of the original computation. Identifying the class of computations that can be decomposed in this manner remains a topic of this on-going research, though we conjecture that most media applications can be decomposed in this ideal manner. However, an M-fold replication ($R = 1$) is also permitted, thereby making SNC applicable to general computations.

A fusion block then combines the sensor outputs to generate a robust estimate $y_f[n]$ of the original/correct output $y[n]$. We employ the term *sensors* to describe the computational units in Fig. 1(b) because errors at the outputs of these blocks can be viewed as a system/architectural-level signature of the variations and noise in the underlying circuit/device fabric.

[1][Online]. Available: http://www.3gpp2.org

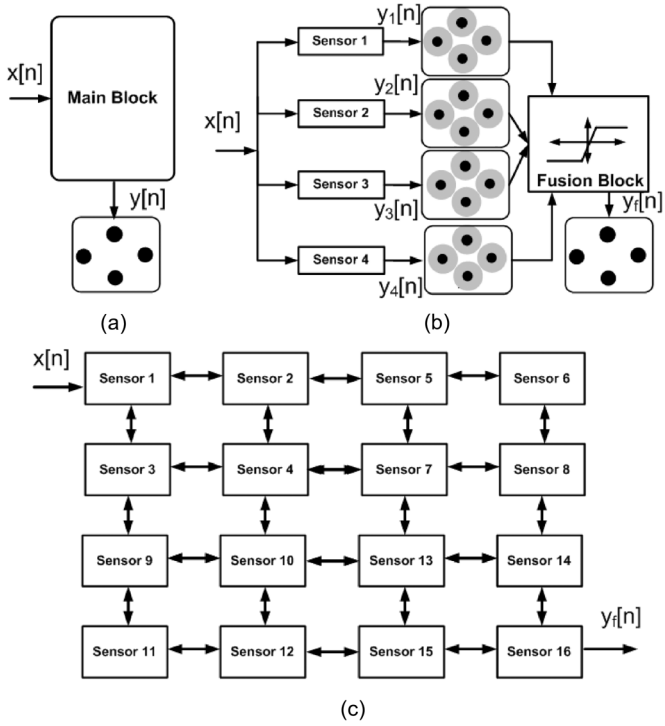


Fig. 1. SNC paradigm. (a) Traditional centralized computation. (b) SNC with the centralized fusion block. (c) General SNC with distributed fusion.

The sensors are designed such that their outputs $y_i[n]$ ($i = 1, \ldots, M$) at time index n are *statistically similar*, i.e., for $1 \leq i \leq M$

$$y_i[n] = y[n] + \eta_i[n] \qquad E[\eta_i[n]] = 0 \tag{1}$$

where $\eta_i[n]$ is the ith sensor error and the expectation is over the ensemble of sensors. Note that statistical similarity implies that the sensor outputs $y_i[n]$ for a fixed value of n equals the correct output $y[n]$ corrupted by additive noise $\eta_i[n]$. We refer to this form of decomposition as *statistically similar decomposition*.

In this paper, we further assume that the noise samples $\eta_i[n]$ are independent and zero mean across the ensemble of sensors. This independence assumption can be easily justified by exploiting

1) *data diversity* where the sensors process independent input data.
2) *architectural diversity* by employing different sensor architectures.
3) *scheduling diversity* by scheduling a different sequence of operations on each sensor.
4) *process diversity* by relying on the random within-die process variations.
5) some combination of the aforementioned.

As an example, Fig. 2 shows the product error distributions and joint error distributions at the output of a 16-bit ripple-carry adder (RCA) and a 16-bit Kogge–Stone (KS) adder with identical inputs and no process variations. It is clear that the *architectural diversity* is sufficient to obtain independence. Indeed,

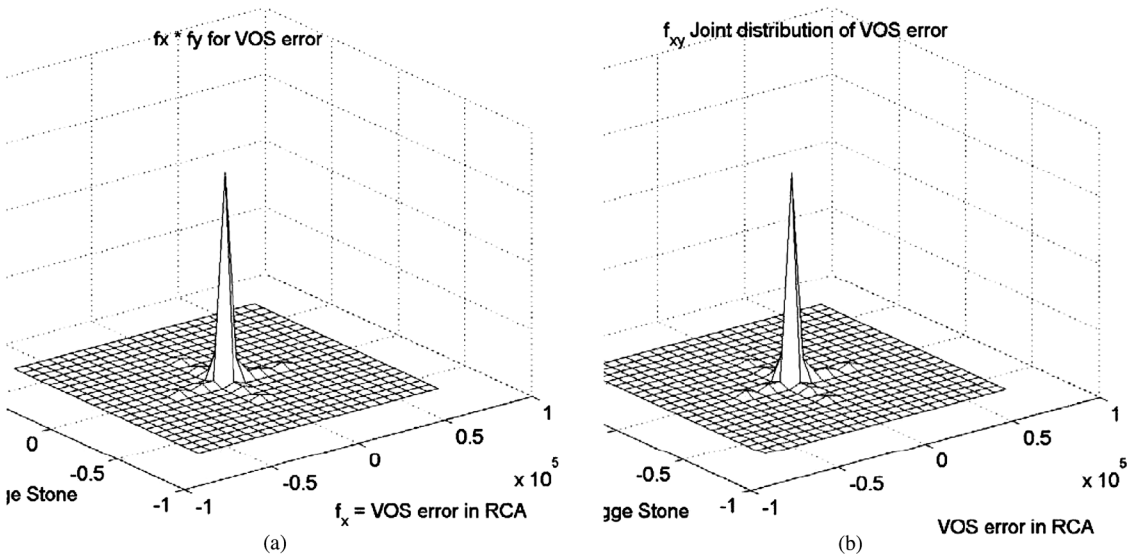


Fig. 2. Error distributions for the 16-bit KS and the 16-bit RCA. (a) Product distribution $[f_{e\mathrm{KS}}(e\mathrm{KS})f_{e\mathrm{RCA}}(e\mathrm{RCA})]$. (b) Joint distribution $[f_{e\mathrm{KS},e\mathrm{RCA}}(e\mathrm{KS}, e\mathrm{RCA})]$.

the Kullback–Liebler distance (KL-distance), a metric that measures the distance between any two distributions, between the distributions shown in the two plots in Fig. 2 was found to be 0.0086. As this value is close to zero, it implies that the product and joint distributions are practically equal, and thus, the errors at the output of the two adders are independent. Note that independent sensor errors though desirable, since it results in low-complexity fusion algorithms, is not essential. Techniques to relax the independence assumption are being developed, but are beyond the scope of this paper.

The gray regions in Fig. 1(b) refer to the fact that instantaneous values of the sensor outputs $y_i[n]$ may not equal the correct output $y[n]$. Thus, an SNC is characterized by two key parameters: the complexity ratio R and the decomposition factor M (or the number of sensors). The most general form of SNC depicted in Fig. 1(c) employs near-neighbor communication between the sensors and a distributed form of fusion. This general form alleviates the need for communicating with a centralized fusion block, as shown in Fig. 1(b), trading off the cost of communication with computation. In this paper though, we will study the robustness and energy efficiency of the basic centralized SNC, as shown in Fig. 1(b). The generalized SNC, as depicted in Fig. 1(c), is a topic of future work.

A number of observations can be made regarding Fig. 1(b).

1) If $R = 1$ *and* the fusion block is a majority voter, then the SNC becomes equivalent to N-modular redundancy (NMR with $N = M$), i.e., the latter falls out as a special case of the former. This also implies that any computation can be formulated as an SNC by replicating the computational units and using a fusion block instead of a majority voter.
2) If $R = 1/M$, then the only hardware overhead in the SNC is the fusion block.
3) R and M can be chosen more or less independently, resulting in a family of SNC architectures.
4) The sensor error $\eta_i[n]$ is composed of two error sources: 1) estimation errors ($\eta_{e,i}[n]$) from the use of low-complexity sensors and 2) hardware errors ($\eta_{h,i}[n]$) due to the nonidealities in the circuit and process, i.e, $\eta_i[n] = \eta_{e,i}[n] + \eta_{h,i}[n]$.
5) As R reduces, $\eta_{e,i}[n]$ increases and $\eta_{h,i}[n]$ generally reduces.
6) The complexity of the SNC $C_{\mathrm{SNC}} = \mathrm{RMC}_{\mathrm{orig}} + C_{\mathrm{fusion}}$, where C_{orig} and C_{fusion} are the complexities of the original/main computation and the fusion block, respectively. Note that ignoring the fusion block complexity, RM can be viewed as the *relative complexity* of the SNC with respect to the main computation.
7) The class of computations that are amenable to statistically similar decomposition is yet to be determined. Most media and digital signal processing algorithms are expected to fall into this class.

Irrespective of and agnostic to the error source, a robust fusion block combines the sensor outputs to produce an output $y_f[n]$, which is statistically close to the correct output $y[n]$. The algorithmic and architectural aspects of the robust fusion block are described next.

A. Robust Statistics

Robust statistics is the science of computing an estimate, which is insensitive to small deviations of the noise statistics from its assumed probability density function (pdf) [25]. An estimate is said to be robust if: 1) it exhibits optimal or near optimal performance when the assumed noise pdf is correct, 2) the performance worsens only slightly when the deviation from the assumption is mild, and 3) large deviations from the assumed model do not cause drastic performance losses.

In the following, for the sake of simplicity, we drop the time index n. Thus,

$$y_i = \theta + \eta_i = \theta + \eta_{e,i} + \eta_{h,i} \quad \text{for } 1 \le i \le M. \tag{2}$$

where M is the number of sensors, y_i is the ith sensor output, $\theta = y[n]$ is the actual output [see Fig. 1(a)], and η_i is the additive noise.

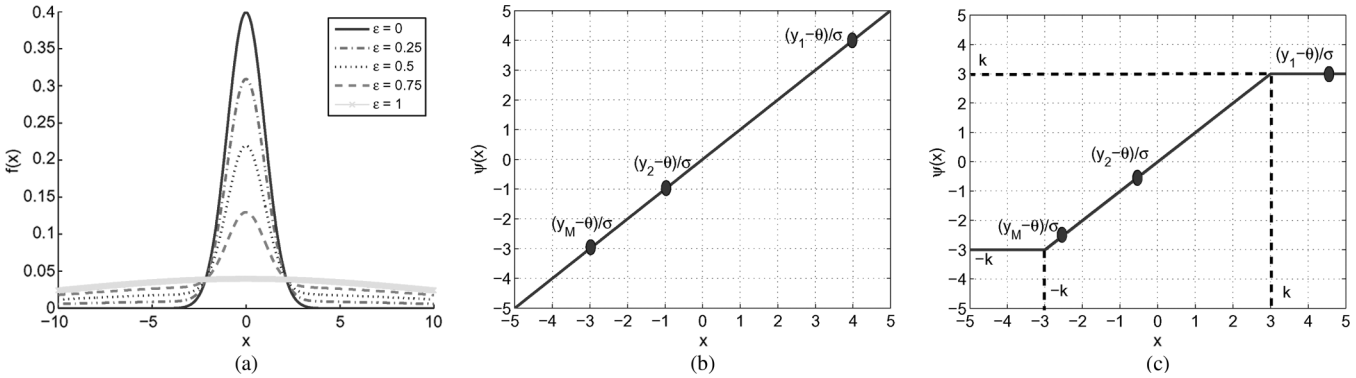


Fig. 3. Examples of: (a) the ϵ-contaminated Gaussian pdf of additive sensor noise η_i, (b) the ψ function resulting in maximum-likelihood estimate for $\epsilon = 0$, and (c) the ψ function resulting in maximum-likelihood estimate for $\epsilon \neq 0$.

While traditional techniques perform poorly if the assumed distribution of noise is incorrect, a robust estimator is designed with a *nominal distribution* in mind and made insensitive to any deviation from the nominal. The idea is to start with an estimator for an assumed distribution, and modify it, so that if our assumed model is violated the quality of the estimate is not excessively sacrificed. Here, sensor output errors η_i are modeled as random variables drawn from a nominal distribution with probability $(1-\epsilon)$ and some *unknown* harsh *contaminant distribution* with probability ϵ for some $0 < \epsilon < 1$. Such a distribution is referred to as an ϵ-*contaminated* distribution, an example of which is shown in Fig. 3(a). Here, note that ϵ corresponds to the frequency of hardware errors and is assumed to be known.

In SNC, the nominal distribution is chosen to be Gaussian because the estimation error $\eta_e[n]$ is, in fact, Gaussian. The contaminant is due to the hardware errors $\eta_h[n]$, whose distribution is unknown and, in fact, whose exact form is unimportant. The additive model for η_i in (2) can be approximated with an ϵ-contaminated model when nonzero values of $\eta_{h,i}$ have much larger magnitude than $\eta_{e,i}$. In this paper, we assume that timing violations are the main error source and the arithmetic unit architectures are array based, resulting to most significant bit (MSB) errors, which are large in magnitude. This assumption is justified in digital systems because the architectural level impact of process variations and noise manifests itself in terms of increased delay. Thus, the use of the ϵ-contaminated model is justified. Simulation results in Section IV further justify this choice.

Huber [25] proposed the robust estimate $\hat{\theta}$ given by the solution of the equation

$$\sum_{k=1}^{M} \psi\left[(y_k - \hat{\theta})/\sigma\right] = 0 \tag{3}$$

where ψ is an odd-symmetric function known as the influence function and σ is the standard deviation of η_e. For the case of ϵ-contaminated $\mathcal{N}(0,1)$ distributions, ψ is given by

$$\psi(y,k) = \begin{cases} y, & \text{if } |y| \leq k \\ k\,\mathrm{sgn}(y), & \text{else} \end{cases} \tag{4}$$

where k is a constant that depends only on ϵ and the standard normal distribution $\mathcal{N}(0,1)$ [25]. Note that as shown in Fig. 3(b), when $\epsilon = 0$, i.e., there are no hardware errors, then $k = \infty$ and $\psi(x) = x$, which implies that $\hat{\theta}$ is the mean of the sensor outputs. For nonzero values of ϵ, $\psi(x)$ is as shown in Fig. 3(c), where $k = 3$. The plot in Fig. 3(c) can be interpreted as stating that the robust estimator limits the influence of outliers. Note further that the distribution of the contaminant is unknown as would be the case in practice as timing violations are the contaminants. The theory of robust statistics provides a solid theoretical foundation for determining the parameters of such a robust estimator.

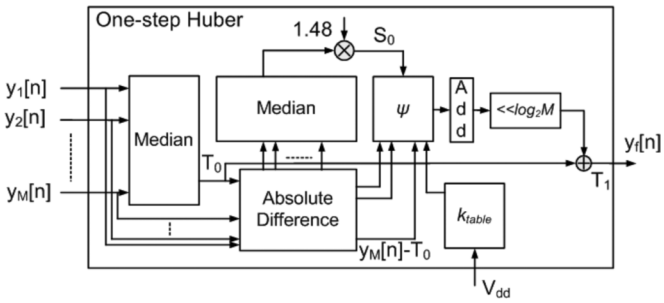


Fig. 4. One-step Huber fusion block architecture.

B. One-Step Huber Algorithm

The solution to (3) requires that both $\hat{\theta}$ and σ be estimated. The *one-step Huber* algorithm [25] can be employed to compute the parameters of the estimator, as shown later.

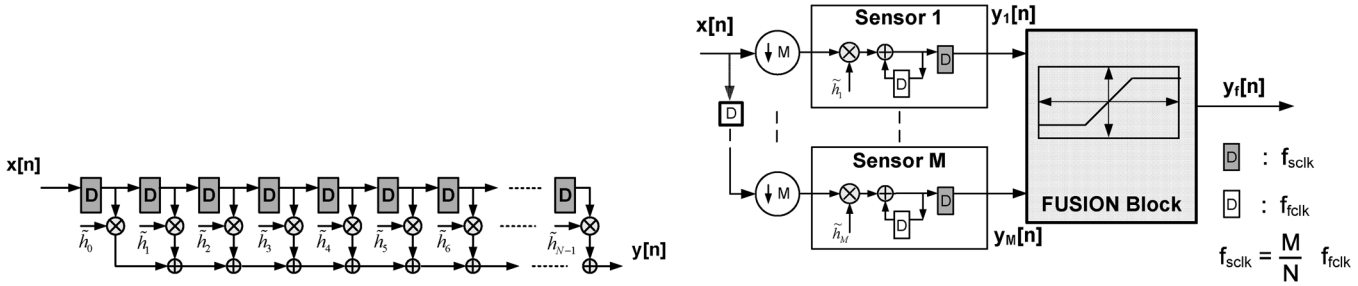
1) SNC(median): Compute scale estimate σ

$$\hat{\theta_0} = \mathrm{median}\{y_i\}$$
$$\hat{\sigma} = 1.4826 \times \mathrm{median}\left\{|y_i - \hat{\theta_0}|\right\}.$$

2) SNC(1-step): Compute location estimate $\hat{\theta}$

$$\hat{\theta_1} = \hat{\theta_0} + \frac{\frac{1}{M}\sum_i \psi(y_i - \hat{\theta_0}, \hat{\sigma} k_{\mathrm{table}})}{0.5}.$$

A one-step Huber architecture is shown in Fig. 4. The median filter for the robust fusion algorithm is implemented using the architecture described in [26] by replacing the analog majority gate with a digital version. The value of k_{table} used in the one-step Huber algorithm depends only on ϵ, and hence, is predetermined for a few expected values of ϵ and stored in a ROM. The gate complexity of the median filter and the one-step Huber



(a) (b)

Fig. 5. Matched filter for PN-acquisition: (a) traditional and (b) an SNC-based architectures.

TABLE I
GATE-LEVEL COMPLEXITY OF VARIOUS SNC COMPONENTS FOR $M = 8$ AND 13-bit SENSOR OUTPUTS

Block	Full Adders	Registers	Other gates
Sensor(Multiply-accumulate)	104	280	72
SNC(median)	104	-	416
SNC(1-step)	772	-	1168

architecture $(M = 8)$ is tabulated in Table I in terms of its various components.

The sensors in SNC in Fig. 1(b) can be viewed as statistical estimators of the main block in Fig. 1(a). Thus, statistical signal processing techniques can be employed to construct such sensors. Examples of estimators include reduced precision replica (RPR) [27], where the estimator is a reduced precision version of the main block. The prediction-based technique [28] employs a linear predictor to estimate in the main block output $y[n]$. The input subsampled replica (ISR) technique [29] employs polyphase decomposition to obtain an estimator for uncorrelated input signals. Other estimators can be designed similarly.

Next, we illustrate the use of the proposed SNC approach for designing a robust low-power PN-code acquisition system and study its power versus performance tradeoffs. In the following, we employ the notation SNC(median) and SNC(1-step) to denote SNC-based architectures employing the median and one-step Huber algorithms, respectively.

III. SNC-BASED PN-CODE ACQUISITION SYSTEM

Wireless communication standards such as the IEEE 802.11 and WCDMA employ PN-codes for coarse timing acquisition. PN codes are characterized by the property that the cross-correlation of two different PN codes is zero for all lags, the autocorrelation function has an impulse at lag zero, and it is nearly zero for all other values of lag. Code acquisition is performed by correlating the noisy received signal with a local PN code using a *matched filter*. The peaks in the output of the matched filters are used for detection and synchronization. Reducing power consumption of these matched filters is essential [20], [21] in portable wireless applications.

A. Conventional Architecture

Fig. 5(a) shows the traditional direct-form implementation of the matched filter with N taps. Multiply-accumulate (MAC) units are commonly employed to compute the correlation of the received signal with the PN-code [21]. In the conventional architecture, the MACs are designed and operated at a critical supply voltage $V_{\mathrm{dd-crit}}$, such that the worst-case critical path (with respect to process/voltage/temperature corner and input combination) satisfies the timing constraint determined by the clock period. This worst-case design, although free of timing violations, consumes high power compared to an average-case design. We push the limits of conventional power savings by employing an SNC-based average-case implementation.

B. SNC-Based Robust Low-Power Architecture

In the SNC-based design, statistically similar decomposition with $R = 1/M$ is achieved via polyphase decomposition, as shown in Fig. 5(b), where $\tilde{h}_j = \{h_j, h_{j+M}, \ldots, h_{j+(M-1)M} \mathrm{mod} N\}$ $(j = 1, \ldots, M-1)$. We choose this form of decomposition as it has the lowest hardware overhead, i.e., the overhead is due to the fusion block only. The M sensor outputs exhibit an estimation error $\eta_{e,i}[n] = y[n] - My_i[n]$ due to input subsampling under error-free hardware operation. Additionally, we expect the sensors to generate computational errors due to nominal case design, or by operating them at subcritical supply voltage also known as *voltage overscaling* (VOS) [30], where the supply voltage is given by

$$V_{\mathrm{dd}} = K_{\mathrm{vos}} V_{\mathrm{dd-crit}} \tag{5}$$

where $0 < K_{\mathrm{vos}} \leq 1$ and $V_{\mathrm{dd-crit}}$ is the minimum supply voltage needed to avoid timing violations.

Process variations are classified as die-to-die (D2D) and within-die (WID) variations. The standard deviation of the gate delay due to D2D process variations is denoted as σ_g. Process variations and VOS result in input-dependent timing violations when specific input patterns excite circuit paths with a delay greater than the clock period. The probability of timing errors depends on the path-delay distribution of the architecture and the probability distribution of the inputs. The probability of VOS errors $\epsilon(\mathrm{VOS})$, for a 16-bit RCA implemented in an IBM 130-nm process technology, is shown in Fig. 6(a), assuming uniformly distributed inputs. The simulation methodology for obtaining the plots in Fig. 6 is described in Section IV-A. We observe that the RCA exhibits $\epsilon(\mathrm{VOS}) = 0.01$ at $K_{\mathrm{vos}} = 0.7$. Similarly, the probability of error due to D2D process variations $\epsilon(\mathrm{process})$ for the RCA is shown in Fig. 6(b) in terms of σ_g, where the supply voltage is chosen to avoid errors at the

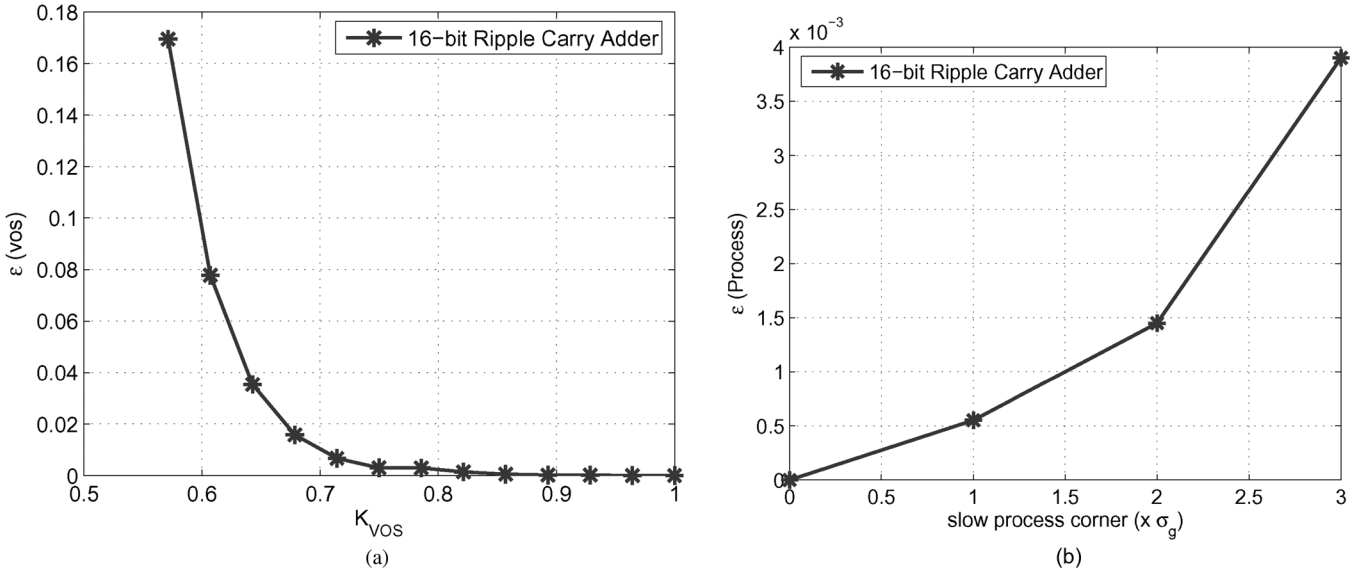


Fig. 6. Error probability of a 16-bit RCA due to: (a) VOS and (b) process variations.

	Process corner	Supply voltage, Body-bias voltage (V_{dd}, V_b)				
Process scenario	$3\sigma_g$ slow	1.35, 0.2	1.2, 0	1.15, 0.2	1.05, 0	1.0, 0.2
VOS scenario	Nominal	1.1, 0	1.05, 0	1.0, 0	0.9, 0	0.84 0

(a)

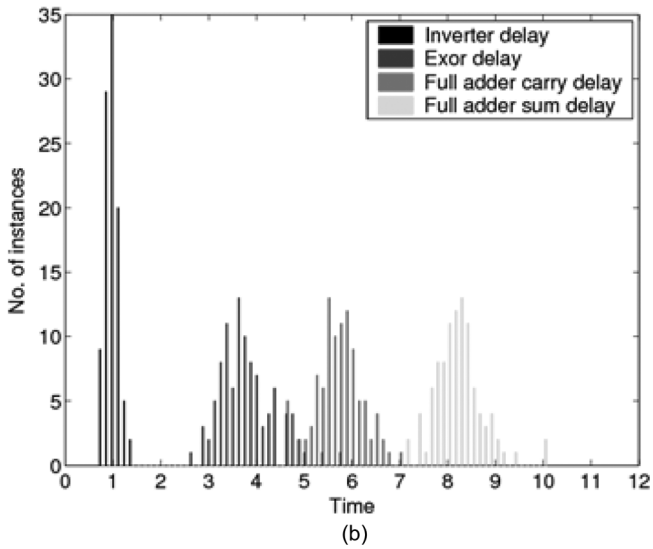


Fig. 7. Simulation setup showing: (a) various process and voltage conditions for process variations scenario, VOS scenario, and (b) normalized delay distributions of various gates for a $3\sigma_g$ slow die with WID variations at $V_{\rm dd} = 1.2$ V.

nominal process corner. We observe that the RCA exhibits an $\epsilon(\text{process}) = 0.004$ at the $3\sigma_g$ slow process corner. This is because the path-delay distribution of an RCA has a long tail.

Timing errors due to process variations or VOS deteriorates the algorithmic performance of the PN acquisition system if left uncorrected/uncompensated for. In this paper, we employ the SNC to restore system performance. The fusion block in SNC is implemented using $\hat{\theta}_1$ in the one-step Huber and the median $\hat{\theta}_0$ in the Huber algorithm. Note that, from Fig. 5(b), the sensors (MAC units) operate at a fast clock frequency of $f_{\rm fclk}$ while the fusion block operates at a slow clock frequency $f_{\rm sclk} = (M/N)f_{\rm fclk}$. This helps in reducing the power overhead due to the fusion block. In fact, in a 130-nm IBM CMOS process, we find that an M-operand adder and the median fusion block consume 1.1% of the total power, while the one-step Huber fusion block consumes 14% of the total power.

IV. Simulation Results

In this section, we compare the power, performance, and complexity tradeoffs between the conventional architecture and SNC-based architectures (see Fig. 5) using: 1) the one-step Huber algorithm and b) a median filter as the fusion block. For simplicity, we consider the scenario where $R = 1/M$, i.e., the only overhead in the SNC architecture is the fusion block.

A. Simulation Setup

1) Algorithmic Setup: The system-level throughput requirements for PN-code acquisition system was fixed at 12.5 Mchips/s. We chose a PN code of length $N = 256$ from a subset of the length 2^{15} PN sequence specified in the CDMA2000 standard. The received signal was assumed to be an 8-bit length-1000 subsequence of the same PN sequence corrupted

by additive white Gaussian noise (AWGN) channel noise to yield an $\mathrm{SNR} = -12$ dB. The fusion algorithms were implemented in MATLAB. We assume error-free operation by the fusion blocks and the threshold-based detector, and compute the receiver operating characteristic (ROC), the probability of detection (P_{Det}) versus probability of false alarm (P_F). A false-alarm event occurs when the threshold detector declares the presence of a PN code in the input incorrectly due to the AWGN channel noise or due to the computation errors originating from slow process/VOS. For power versus performance comparisons at different supply voltages, we calculated P_{Det} from the ROC for a typical fixed value of $P_F = 5\%$ [31]. For process variations scenario, we simulate 30 architectural instances with 100 independent data streams containing the PN code to compute 30 ROCs. Thus, P_{Det} is a random variable, and hence, it has a distribution, which is characterized in the simulation results. For VOS scenario, we simulated 1000 independent data streams to compute the ROC and P_{Det}.

2) Architectural Setup: For a fair comparison, the conventional matched filter in Fig. 5(a) is also polyphase decomposed [20], as in Fig. 5(b), but with an M-operand adder in place of the fusion block. Thus, the MAC units in the conventional and the SNC implementations have the same critical path delay.

Fig. 7(a) tabulates the process variation and VOS scenarios being considered for a 1.2-V, 130-nm CMOS process from IBM. For simulating process variations, we sample the gate-delay distributions, such as the one in Fig. 7(b), which are obtained from circuit simulations at: 1) the $3\sigma_g$ slow process corner operating at critical supply voltage $V_{\mathrm{dd-crit}} = 1.2$ V corresponding to nominal process corner and 2) the $3\sigma_g$ slow process corner operating under adaptive supply voltage (ASV) boost and adaptive body bias voltage (ABB). This process is repeated 30 times in order to obtain 30 instances of the conventional and SNC architectures. Each of the 30 instances of the conventional and the SNC architectures were simulated using a hardware description language (HDL) simulator, which operates at the gate level. The results are shown in Section IV-B. VOS simulations are performed at the nominal process corner, as shown in the second row of Fig. 7(a), at various subcritical supply voltages, and the results are shown in Section IV-C.

3) Circuit Setup: Monte Carlo simulations using statistical model files were employed to characterize the gate delays for basic gates such as the full adder, XOR, and others with respect to supply voltage and process corner using HSPICE for an IBM 130-nm CMOS process. Fig. 7(b) shows the typical example of the normalized delay distributions resulting from the presence of WID variations at $V_{\mathrm{dd}} = 1.2$ V. We simulate the transistor-level netlists of sensors (MACs) in the conventional and SNC architectures using HSPICE with a few random input vectors to obtain the power consumption of the sensors at different (V_{dd}, V_b), supply voltage, and body-bias voltage levels. The power consumption of the fusion blocks (conventional M-operand adder and median) are obtained using *Synopsys design analyzer* for a $f_{sclk} = 12.5$ MHz.

B. SNC Performance With Process Variations

The ROCs and the P_{Det} distributions of SNC with $M = 8$, due to WID variations at the $3\sigma_g$ slow process corner, supply, and body-bias voltages of (1.2, 0 V) are shown in Fig. 8(a). Similar ROCs and the P_{Det} distributions were obtained for the ABB/ASV cases enumerated in Fig. 7(a). We note from Fig. 8(a) that the mean P_{Det} drops by approximately three orders of magnitude at $P_F = 0.05$ for the conventional architecture at the $3\sigma_g$ slow process corner operating at supply voltage $V_{\mathrm{dd-crit}} = 1.2$ V. The SNC architecture improves P_{Det} by close to three orders of magnitude over the conventional architecture under identical process and voltage conditions. In addition, the SNC improves the variation in P_{Det} (σ/μ) by two orders of magnitude.

As we increase the number of sensors from $M = 8$ to $M = 32$ at a fixed throughput of 12.5 Mchips/s, the performance degradation due to process variations is reduced for all architectures. Since the performance of the SNC(median) is comparable to that of SNC(1-step), we choose SNC(median) in the following discussion of power-performance tradeoff.

Next, we show a plot of power consumption of the two architectures with $M = 8$ along with the mean P_{Det} in Fig. 9(a). The (V_{dd}, V_b) are adjusted in order to minimize power. At the $3\sigma_g$ slow process corner (second bar), the power consumption of the conventional architecture drops by 8%, but is accompanied by a three orders of reduction in P_{Det}, as indicated earlier. The application of ABB and ASV (third bar) to reduce the gate delays and correct the timing errors increases the power consumption of the conventional architecture by 33% at (1.35 V, 0.2 V) while achieving a mean $P_{\mathrm{Det}} = 0.83$. The SNC architecture (fourth bar) at (1.15 V, 0.2 V) achieves a comparable $P_{\mathrm{Det}} = 0.80$, but consumes 31% lower power than the conventional architecture does. Similar performance-power tradeoffs can be observed for the case when $M = 32$ in Fig. 9(b).

C. SNC Performance With VOS

In this section, we study the power saving and performance SNC via VOS only. The process corner is at nominal. The purpose of this study is to understand the impact of low-voltage operation on SNC architectures. We perform HDL simulations for $K_{\mathrm{vos}} = 0.92, 0.87, 0.83, 0.75, 0.7$, and 0.6 to obtain a set of ROCs. Representative ROC plots corresponding to $K_{\mathrm{vos}} = 0.75$ for $M = 8$, and $M = 32$ are shown in Fig. 10(a) and (b), respectively. With $M = 8$ [see Fig. 10(a)], the conventional architecture suffers about a two orders of magnitude loss in P_{Det} at $P_F = 0.05$. This is bit better than its performance in the presence of process variations. The SNC architecture recovers this loss completely. Similar results are observed for the case $M = 32$ [see Fig. 10(b)], with the only difference being that the loss in P_{Det} is not as severe for all the architectures.

Fig. 11 shows the power consumption versus performance tradeoff at $P_F = 5\%$. From Fig. 11(a), for $M = 8$, we note that P_{Det} for SNC(median) is three orders of magnitude better than that of the conventional architecture and saves 36% power. The power savings remain practically unchanged with $M = 32$, as shown in Fig. 11(b).

D. Soft Errors Results

We introduce random errors in SNC-based $(M = 8)$ gate-level MAC implementations by flipping the gate outputs randomly for sample error probabilities of 10^{-1} [see Fig. 12(a)] and 10^{-2} [see Fig. 12(b)], which the performance of the SNC-based

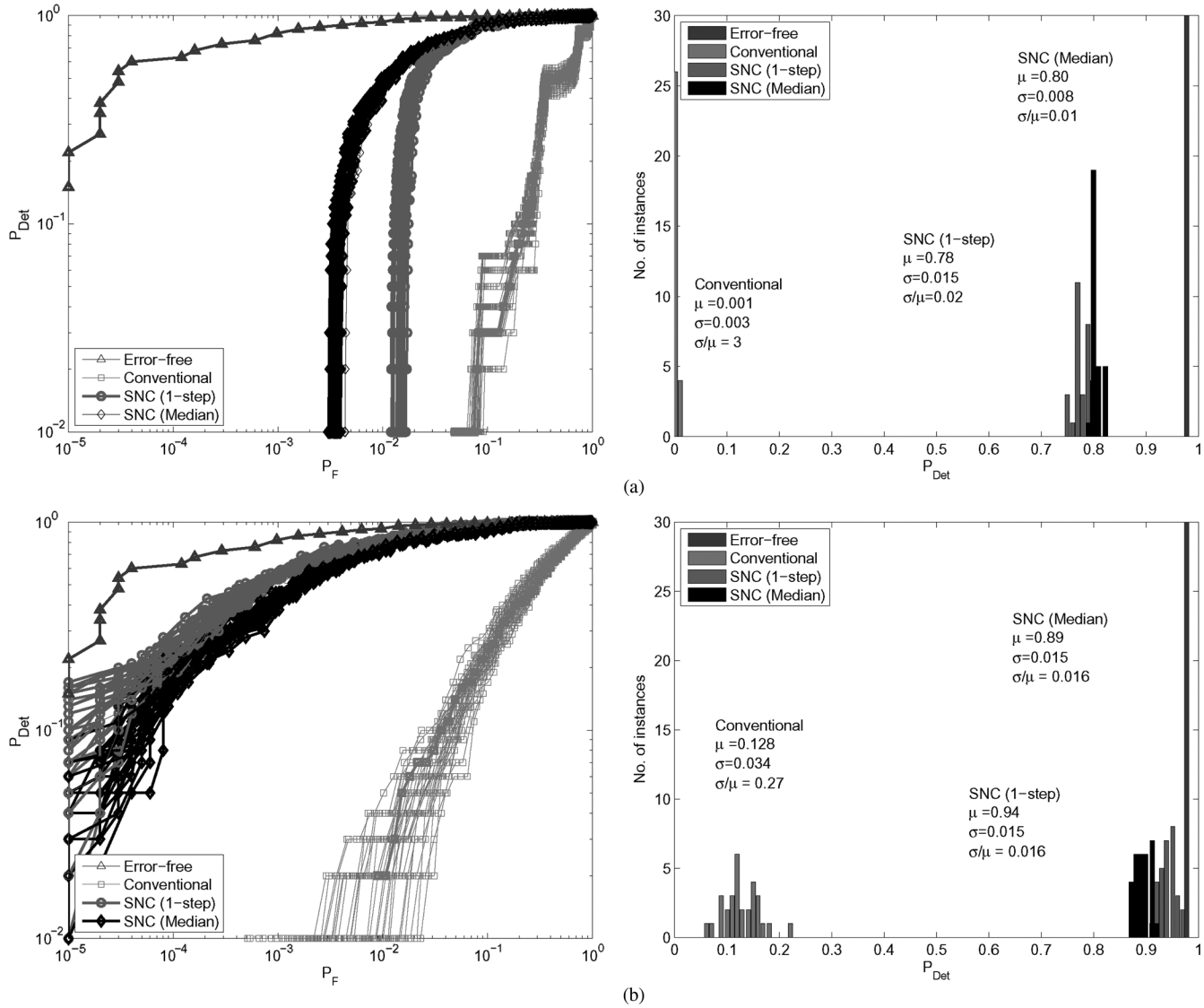


Fig. 8. ROCs and pdfs of $P_{\rm Det}$ (at $P_F = 5\%$) of the conventional and SNC-based architectures due to WID variations at the $3\sigma_g$ slow process corner, supply, and body-bias voltages of (1.2, 0 V) for: (a) $M = 8$ and (b) $M = 32$.

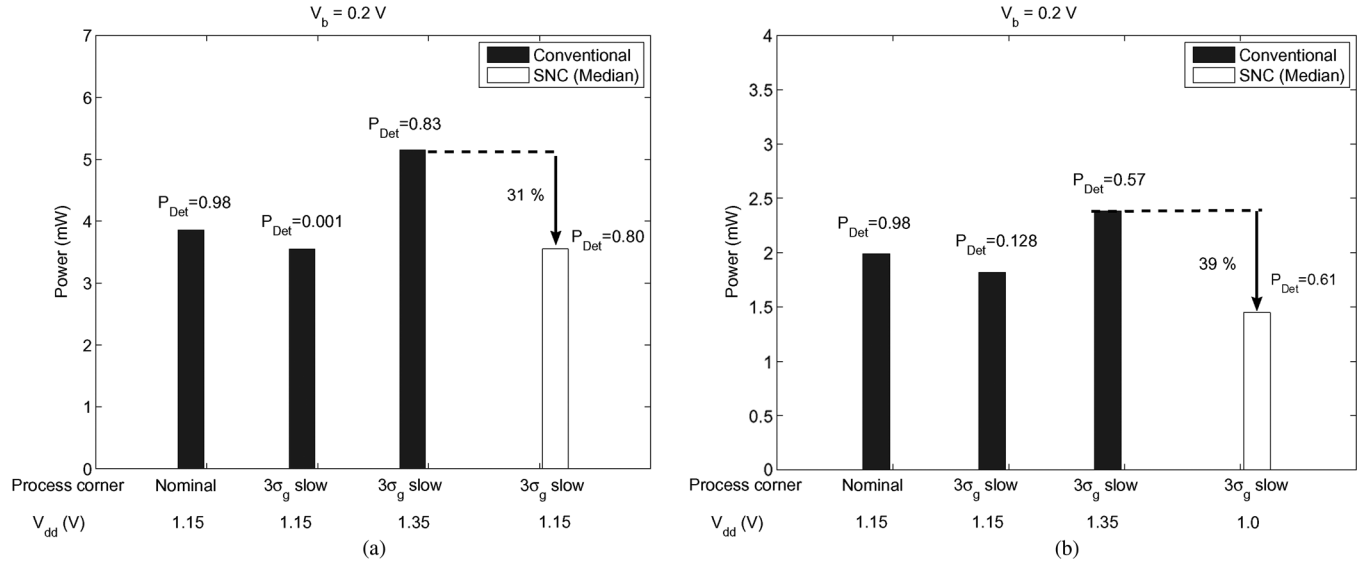


Fig. 9. Power performance tradeoff (at $P_F = 5\%$) for: (a) $M = 8$ and (b) $M = 32$.

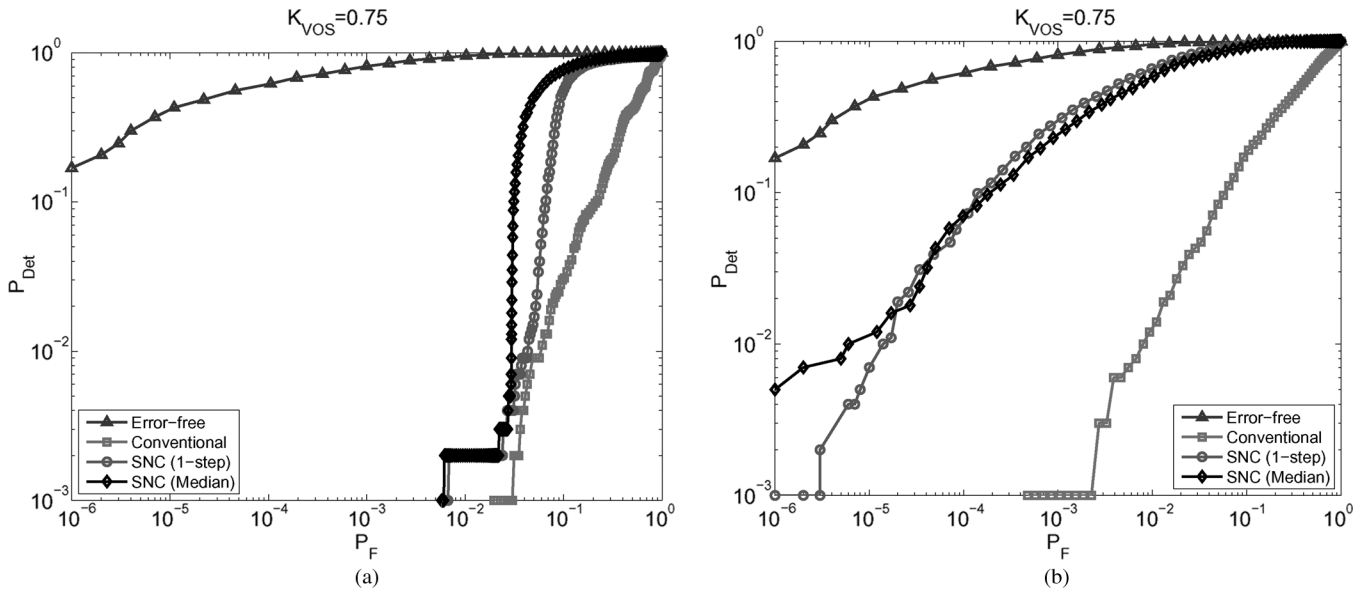


Fig. 10. ROC of the PN-code detector at $K_{\rm vos} = 0.75$ for: (a) $M = 8$ and (b) $M = 32$.

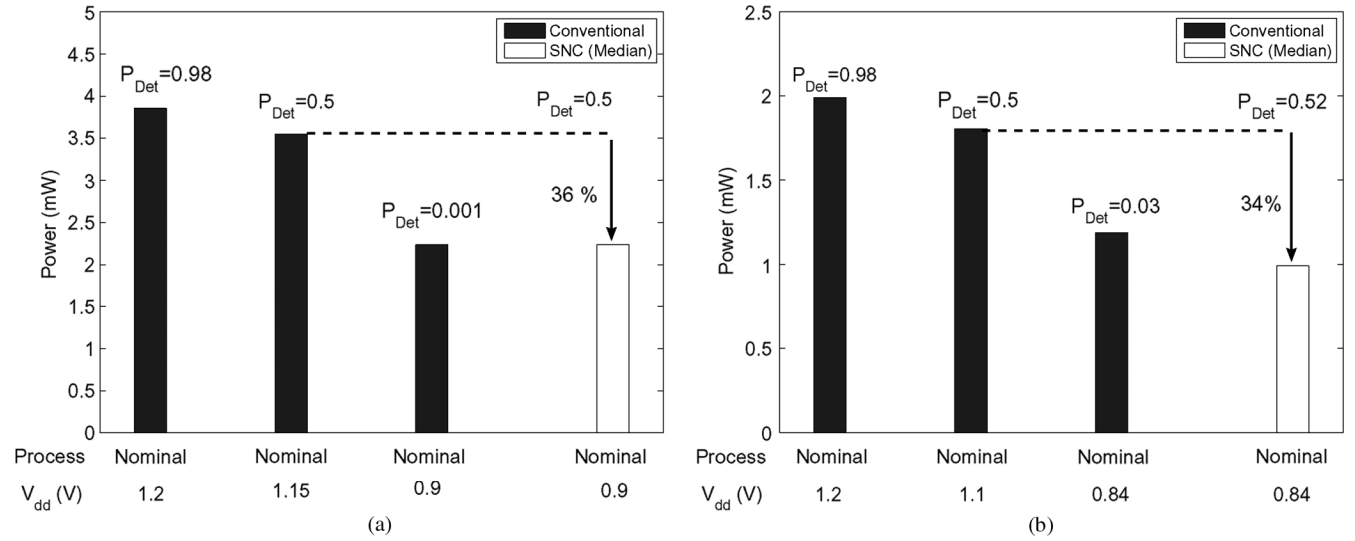


Fig. 11. Total power savings are plotted against probability of detection (at $P_F = 5\%$) for the conventional and the SNC-based detectors for: (a) $M = 8$ and (b) $M = 32$.

design is robust to random errors. Again, we find that SNC provides close to three orders-of-magnitude improvement over the conventional system. Similar results were obtained for $M = 32$.

E. Synthesis and Layout Results

We synthesized the layouts of the SNC and the conventional architectures in the 130-nm process technology. We designed the MACs and both the fusion block data paths in Verilog, and synthesized the layouts via *Synopsys design analyzer* and *Cadence silicon ensemble*. Fig. 13 shows the layouts for the conventional and the proposed SNC-based systems. The SNC(median) and SNC(one-step Huber) result in area overheads of 11% and 715%, respectively. Clearly, the SNC(median) provide extremely favorable tradeoffs between power, performance, and area.

F. SNC Performance Across Technology Nodes

We study the impact of technology scaling from the 90 nm node to 32 nm using predictive technology models (PTMs) [32] on the performance of the SNC architecture.[2] The results for the 130-nm IBM CMOS process are reproduced in Table II for the sake of continuity. We note that the mean $P_{\rm Det}$ drops severely by approximately three orders of magnitude for the conventional architecture in all technologies. The SNC architecture improves $P_{\rm Det}$ and the variation in $P_{\rm Det}$ by close to three orders of magnitude and two orders of magnitude, respectively, for the 130 and 90 nm processes.

However, at the 65 and 45 nm nodes, the SNC performance degrades to the level of the conventional architecture. This is because the frequency of timing violations exceeds the error-compensation capability of SNC. The SNC (median) architecture needs to be augmented with circuit-level techniques of ABB/ASV in order to improve $P_{\rm Det}$ by two orders of magnitude and reduce its variation (σ/μ) by one order of magnitude. The SNC with ABB/ASV cannot recover the performance loss

[2][Online]. Available: http://www.eas.asu.edu/~ptm

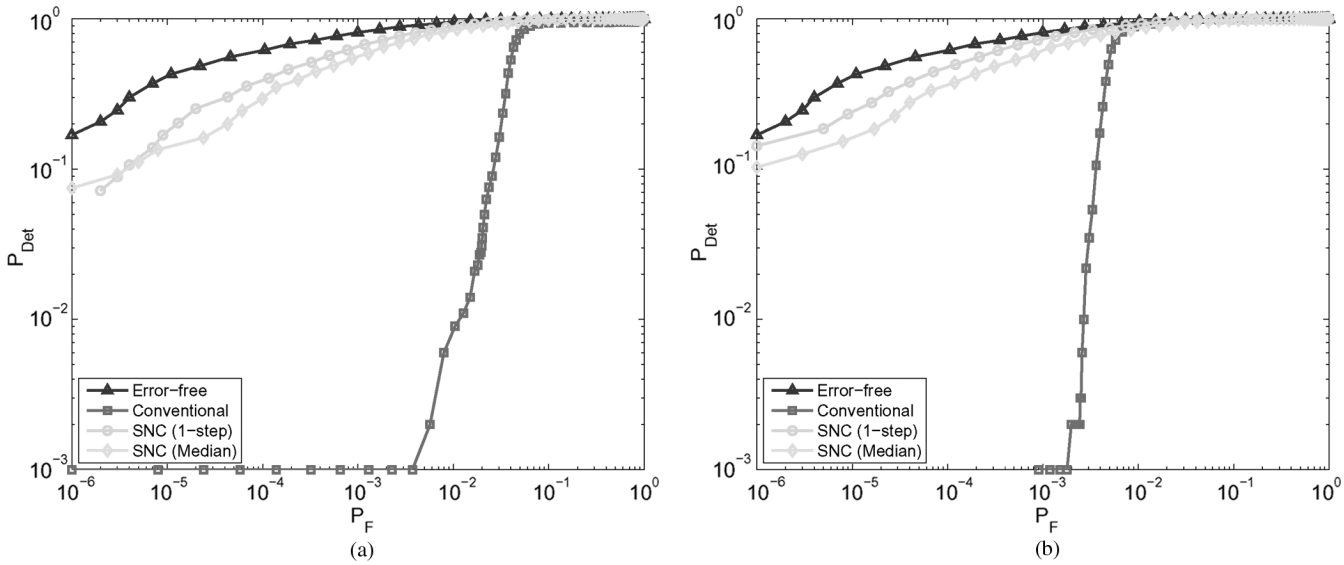


Fig. 12. ROC in the presence of soft errors with: (a) probability of output sample error $P_e = 10^{-1}$ and (b) $P_e = 10^{-2}$.

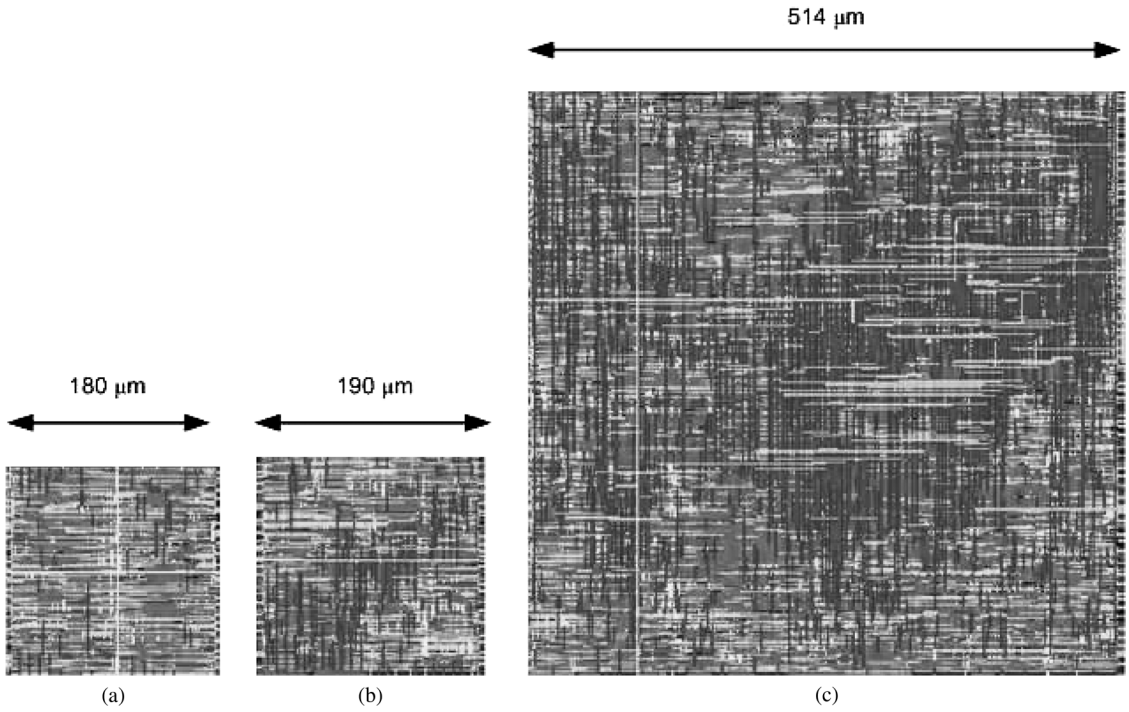


Fig. 13. Synthesized layouts of: (a) conventional, (b) SNC(median), and (c) SNC(one-step) architectures.

at the 32 nm node. This trend clearly indicates that increasingly powerful fusion techniques such as the one outlined in Fig. 1(c) and/or a higher value of M would be required as process technology scales, and that system-level reliability enhancement techniques such as SNC need to be combined with circuit-level techniques.

Table III shows that the achievable power savings using SNC reduces from 31% to 22% for the same detection probability $P_{\rm Det}$. Note that the power optimum $(V_{\rm dd}, V_b)$ combination shows the trend of changing from forward body bias (increased leakage) to zero body bias (decreased leakage) in smaller process technologies. The power for the 32 nm process node is not shown as SNC with ABB/ASV is unable to provide the requisite detection performance, as mentioned earlier.

V. Conclusion and Future Work

This paper has introduced a networked computational paradigm referred to as the SNC. SNC was shown to provide orders-of-magnitude improvement in key reliability metrics such as probability of detection in a PN-code acquisition system while achieving 30%–40% energy savings.

Natural extensions of this study include exploring the entire architectural space spanned by various choices of complexity ratio R and the decomposition factor M, investigating decentralized and iterative fusion algorithms, studying the tradeoff between communication and computational cost involved in SNC, outlining the class of algorithms that can be statistically decomposed into an SNC, developing fundamental theoretical under-

TABLE II
CHARACTERISTICS OF P_{Det} (AT $P_F = 5\%$) DISTRIBUTIONS OF THE THREE ARCHITECTURES DUE TO PROCESS VARIATIONS

Tech. Node		Process			Process and ABB/ASV		
		Conventional	SNC (1-step)	SNC (median)	Conventional	SNC (1-step)	SNC (median)
130 nm	(V_{dd}, V_b) (V)	1.2, 0			1.35, 0.2		
	μ_i	0.001	0.78	0.80	0.84	0.98	0.97
	σ_i	0.003	0.015	0.008	0.24	0.005	0.003
	σ_i/μ_i	3	0.02	0.01	0.3	0.005	0.003
90 nm	(V_{dd}, V_b) (V)	1.1, 0			1.2, 0.1		
	μ_i	0.009	0.02	0.51	0.002	0.79	0.82
	σ_i	0.009	0.016	0.06	0.004	0.03	0.02
	σ_i/μ_i	1.0	0.75	0.13	1.8	0.034	0.025
65 nm	(V_{dd}, V_b) (V)	1.0, 0			1.1, 0		
	μ_i	0.004	0.005	0.008	0.009	0.59	0.77
	σ_i	0.007	0.006	0.02	0.008	0.1	0.02
	σ_i/μ_i	1.6	1.2	2.4	0.89	0.17	0.03
45 nm	(V_{dd}, V_b) (V)	0.9, 0			1.0, 0		
	μ_i	0.001	0.0027	0.003	0.006	0.053	0.60
	σ_i	0.004	0.005	0.006	0.008	0.05	0.07
	σ_i/μ_i	2.2	1.9	1.9	1.3	0.9	0.12
32 nm	(V_{dd}, V_b) (V)	0.8, 0			0.9, 0.2		
	μ_i	0.008	0.01	0.01	0.007	0.006	0.006
	σ_i	0.01	0.008	0.008	0.007	0.007	0.007
	σ_i/μ_i	1.1	0.8	0.8	1	1.2	1.2

TABLE III
POWER SAVINGS USING SNC WITH ABB/ASV UNDER PROCESS ERRORS FOR EQUAL P_{Det} (AT $P_F = 5\%$)

	130nm IBM	90 nm PTM	65 nm PTM	45 nm PTM
Conventional (V_{dd}, V_b) (V)	1.35, 0.2	1.3, 0.1	1.2, 0	1.1, 0
SNC (median) (V_{dd}, V_b) (V)	1.15, 0.2	1.2, 0.1	1.1, 0	1.0, 0
Power savings (%)	31	25	23	22

pinnings for networked computation in general, and SNC in particular, and seeking new applications including high data-rate ones in advanced process technologies. It is clear that the proposed SNC concept represents a fertile area of research for years to come.

REFERENCES

[1] S. Borkar, T. Karnik, and V. De, "Design and reliability challenges in nanometer technologies," presented at the DAC, San Diego, CA, Jun. 2004.

[2] "International Technology Roadmap for Semiconductors," 2001. [Online]. Available: http://www.itrs.net

[3] K. L. Shepard and V. Narayanan, "Noise in deep submicron digital design," in *Proc. ICCAD*, Nov. 1996, pp. 524–531.

[4] K. Agarwal and S. Nassif, "Characterizing process variation in nanometer CMOS," in *Proc. Des. Autom. Conf.*, San Diego, CA, Jun. 4–8, 2007, pp. 396–399.

[5] R. Baumann, "The impact of technology scaling on soft-error rate performance and limits to the efficacy of error correction," in *Proc. IEEE Int. Electron Devices Meet. (IEDM)*, 2002, pp. 329–332.

[6] X. Tang, "Intrinsic mosfet parameter fluctuations due to random dopant placement," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 5, no. 4, pp. 369–376, Dec. 1997.

[7] A. P. Chandrakasan and R. W. Broderson, "Minimizing power consumption in digital CMOS circuits," *Proc. IEEE*, vol. 83, no. 4, pp. 498–523, Apr. 1995.

[8] R. Gonzalez, B. Gordon, and M. Horowitz, "Supply and threshold voltage scaling for low-power CMOS," *IEEE J. Solid-State Circuits*, vol. 31, no. 3, pp. 395–400, Mar. 1999.

[9] N. Banerjee, A. Raychowdhury, K. Roy, S. Bhunia, and H. Mahmoodi, "Novel low-overhead operand isolation techniques for low-power datapath synthesis," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 14, no. 9, pp. 1034–1039, Sep. 2006.

[10] J. Tschanz, J. Kao, S. Narendra, R. Nair, D. Antoniadis, A. Chandrakasan, and V. De, "Adaptive body-bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage," in *ISSCC Dig. Tech. Papers*, Feb. 2002, pp. 422–423.

[11] D. K. Pradhan, *Fault-Tolerant Computer System Design*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[12] D. P. Siewiorek and R. S. Swarz, *Reliable Computer Design*, 2nd ed. Burlington, MA: Digital Press, 1992.

[13] J.-Y. Jou and J. A. Abraham, "Fault-tolerant FFT networks," *IEEE Trans. Comput.*, vol. 37, no. 5, pp. 548–561, May 1988.

[14] N. R. Shanbhag, "Reliable and efficient system-on-a-chip design," *IEEE Comput. Mag.*, vol. 37, no. 3, pp. 42–50, Mar. 2004.

[15] J. A. Cortese and R. Goodman, "A statistical analysis of neural computation," in *Proc. IEEE Symp. Inf. Theory*, Jul. 1994, p. 215.

[16] K. Nepal, I. R. Bahar, J. Mundy, W. R. Patterson, and A. Zaslavsky, "Designing logic circuits for probabilistic computation in the presence of noise," in *Proc. 42nd Annu. Conf. Des. Autom.*, Anaheim, CA, Jun. 13–17, 2005, pp. 485–490.
[17] K. Palem, "Energy aware computing through probabilistic switching: A study of limits," *IEEE Trans. Comput.*, vol. 54, no. 9, pp. 1123–1137, Sep. 2005.
[18] G. Varatkar, S. Narayanan, N. Shanbhag, and D. L. Jones, "Sensor network-on-a-chip," presented at the IEEE Symp. Syst.-on-a-Chip Des., Tampere, Finland, Nov. 19–21, 2007.
[19] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, Aug. 2002.
[20] W. Namgoong and T. Meng, "Power consumption of parallel spread spectrum correlator architectures," in *Proc. ISLPED*, Aug. 1998, pp. 133–135.
[21] D. Senderowicz, "A 23 mW 256-tap 8 MSample/s QPSK matched filter for DS-CDMA cellular telephony using recycling integrator correlators," in *Proc. ISSCC*, Feb. 2000, pp. 354–355.
[22] T. Dumitras and R. Marculescu, "On-chip stochastic communication," in *Proc. Des., Automat. Test Eur. Conf. Exhib. (DATE)*, Munich, Germany, Mar. 2003, pp. 790–795.
[23] P. Bogdan, T. Dumitra, and R. Marculescu, "Stochastic communication: A new paradigm for fault-tolerant networks-on-chip," presented at the VLSI Des., 2007, 95348.
[24] S. Murali, D. Atienza, L. Benini, and G. De Micheli, "A method for routing packets across multiple paths in NoCs with in-order delivery and fault-tolerance guarantees," presented at the VLSI Des, 2007.
[25] P. Huber, *Robust Statistics*. New York: Wiley, 1981.
[26] C. L. Lee and C. W. Jen, "A bit-sliced median filter design based on majority gate," *Proc. Inst. Electr. Eng.*, vol. 139, no. 1, pp. 63–71, 1992.
[27] B. Shim and N. R. Shanbhag, "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.
[28] R. Hegde and N. R. Shanbhag, "Soft digital signal processing," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 9, pp. 813–823, Dec. 2001.
[29] G. Varatkar and N. R. Shanbhag, "Energy-efficient motion estimation using error-tolerance," in *Proc. ISLPED*, 2006, pp. 113–118.
[30] R. Hegde and N. R. Shanbhag, "A voltage overscaled low-power digital filter IC," *IEEE J. Solid-State Circuits*, vol. 39, no. 2, pp. 388 –391 , Feb. 2004.
[31] G. J. R. Povey, "Spread spectrum PN code acquisition using hybrid correlator architectures," *Wireless Pers. Commun., Int. J.*, vol. 8, no. 2, pp. 151 –164 , Sept. 1998.
[32] W. Zhao and Y. Cao, "New generation of predictive technology model for sub-45 nm early design exploration," *IEEE Trans. Electron Devices*, vol. 53, no. 11, pp. 2816–2823, Nov. 2006.

Girish Vishnu Varatkar received the Bachelor of Technology degree from the Indian Institute of Technology, Bombay, India, in 2001, the Master of Science degree from the Carnegie Mellon University, Pittsburgh, PA, in 2003, and the Ph.D. degree in electrical and computer engineering (ECE) from the University of Illinois at Urbana-Champaign (UIUC), Urbana.

Currently, he is working at Qualcomm Flarion Technologies, NJ. He has interned at Qualcomm Inc., San Diego, CA, in the area of SOC verification as an interim engineering intern, and at the Circuits Research Laboratory, Intel, Hillsboro, OH, in the area of high speed digital design as Engineering Research Intern. His research interests encompass a wide range of topics, including error tolerant and energy efficient VLSI architectures, digital signal processing for communication systems, and distributed estimation and detection. He is particularly interested in error-tolerant multimedia and communication systems design.

Dr. Varatkar was a recipient of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS Best Paper Award from the IEEE Circuits and Systems Society in the year 2005 and the C. R. Allen Outstanding International Student Award in 2008 from UIUC for demonstrating academic excellence and research potential.

Sriram Narayanan received the B.S. degree in computer engineering from Iowa State University, Ames, in 2001 and the M.S. degree in electrical and computer engineering from University of Illinois at Urbana-Champaign, Urbana (UIUC), where his thesis research was focused on asynchronous sensor networks. He is currently pursuing the Ph.D. degree from UIUC.

He was a Software Engineer with National Instruments from 2001 to 2003, where he developed software for PC-based data-acquisition systems. His general interests are in designing SoCs that are low power and robust to process, voltage and temperature variations.

Naresh R. Shanbhag (F'06) received the Ph.D. degree in electrical engineering from the University of Minnesota, Twin Cities, in 1993.

From 1993 to 1995, he worked at AT&T Bell Laboratories, Murray Hill, NJ, where he was the lead chip architect for AT&T's 51.84 Mb/s transceiver chips over twisted-pair wiring for Asynchronous Transfer Mode (ATM)-LAN and very high-speed digital subscriber line (VDSL) chip-sets. Since August 1995, he is with the Department of Electrical and Computer Engineering, and the Coordinated Science Laboratory where he is presently a Professor. He was on a sabbatical leave of absence at the National Taiwan University in Fall 2007. His research interests include the design of integrated circuits and systems for communications including low-power/high-performance VLSI architectures for error-control coding, equalization, as well as integrated circuit design. He has more than 150 publications in this area and holds four US patents. He is also a co-author of the research monograph *Pipelined Adaptive Digital Filters* (Kluwer, 1994). He is leading forward-looking research themes in DOD and Semiconductor Research Corporation (SRC) sponsored Microelectronics Advanced Research Corporation (MARCO) centers under their Focus Center Research Program (FCRP) since 2006.

Dr. Shanbhag was a recipient of the 2006 IEEE Journal of Solid-State Circuits Best Paper Award, the 2001 IEEE Transactions on VLSI Best Paper Award, the 1999 IEEE Leon K. Kirchmayer Best Paper Award, the 1999 Xerox Faculty Award, the Distinguished Lecturership from the IEEE Circuits and Systems Society in 1997, the National Science Foundation CAREER Award in 1996, and the 1994 Darlington Best Paper Award from the IEEE Circuits and Systems Society. He served as an Associate Editor for the IEEE TRANSACTION ON CIRCUITS AND SYSTEMS: PART II (1997-1999) and the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS (1999-2002 and 2009-present), respectively. He is currently serving on the technical program committees of major international conferences. In 2000, he co-founded and served as the chief technology officer of Intersymbol Communications, Inc., a venture-funded fabless semiconductor start-up that provides mixed-signal ICs for electronic dispersion compensation of OC-192 optical links. In 2007, Intersymbol Communications, Inc., was acquired by Finisar Corporation, Inc.

Douglas L. Jones (F'02) received the B.S.E.E., M.S.E.E., and Ph.D. degrees from Rice University, Houston, TX, in 1983, 1986, and 1987, respectively.

During the 1987-1988 academic year, he was at the University of Erlangen-Nuremberg in Germany on a Fulbright postdoctoral fellowship. Since 1988, he has been with the University of Illinois at Urbana-Champaign, where he is currently a Professor in Electrical and Computer Engineering, the Coordinated Science Laboratory, and the Beckman Institute. He was on sabbatical leave at the University of Washington in Spring 1995 and at the University of California at Berkeley in Spring 2002. In the Spring semester of 1999 he served as the Texas Instruments Visiting Professor at Rice University. He is an author of two DSP laboratory textbooks. His research interests include digital signal processing and communications, including nonstationary signal analysis, adaptive processing, multisensor data processing, OFDM, and various applications such as low-power implementations and advanced hearing aids.

Dr. Jones was selected as the 2003 Connexions Author of the Year. He served on the Board of Governors of the IEEE Signal Processing Society from 2002-2004.