Reliable Low-Power Digital Signal Processing via Reduced Precision Redundancy

Byonghyo Shim, *Member, IEEE*, Srinivasa R. Sridhara, *Student Member, IEEE*, and
Naresh R. Shanbhag, *Senior Member, IEEE*

Abstract—In this paper, we present a novel algorithmic noise-tolerance (ANT) technique referred to as reduced precision redundancy (RPR). RPR requires a reduced precision replica whose output can be employed as the corrected output in case the original system computes erroneously. When combined with voltage overscaling (VOS), the resulting soft digital signal processing system achieves up to 60% and 44% energy savings with no loss in the signal-to-noise ratio (SNR) for receive filtering in a QPSK system and the butterfly of fast Fourier transform (FFT) in a WLAN OFDM system, respectively. These energy savings are with respect to optimally scaled (i.e., the supply voltage equals the critical voltage $V_{dd-crit}$) present day systems. Further, we show that the RPR technique is able to maintain the output SNR for error rates of up to 0.09/sample and 0.06/sample in an finite impulse response filter and a FFT block, respectively.

Index Terms—Digital signal processing, low-power, noise-tolerance, reliability, supply voltage.

I. INTRODUCTION

THE rapid growth in demand for portable and wireless computing systems is driving the need for ultra low-power systems. Supply voltage scaling is widely acknowledged as an effective low-power technique [1]–[4]. However, in deep-submicron (DSM) process technologies, noise and process variations have emerged as formidable problems that circuit and system designers need to address [5], [6]. These problems have raised serious questions regarding our ability to design reliable and efficient (hence, affordable) microsystems and hence the ability to extend Moore's law [7] well into the DSM regime.

Our past research [8]–[12] on energy-efficiency bounds of DSM VLSI systems in the presence of noise strongly suggests that design techniques based on *noise-tolerance* need to be developed if energy-efficiency and reliability are to be jointly addressed. Indeed, the 2001 International Technology Roadmap for Semiconductors [14] refers to *error-tolerance* as a design challenge for the next decade. We have developed noise-tolerance at the algorithmic [9] as well as circuit [13] levels of the design hierarchy. In [9], we proposed algorithmic noise-tolerance (ANT) as a technique for combating system level errors in digital signal processing systems. An aggressive low-

power technique, referred to as *voltage overscaling* (VOS) was proposed in [9] and [11]. Voltage overscaling refers to reduction of supply voltage beyond $V_{dd-crit}$, without sacrificing the throughput, where $V_{dd-crit}$ is the supply voltage below which timing violations start to appear, i.e.,

$$V_{dd} = K_{vos} V_{dd-crit}, \quad 0 \leq K_{vos} < 1 \tag{1}$$

where K_{vos} is referred to as the VOS factor (VOSF). Since, under VOS, the critical path delay T_{cp} of the system becomes greater than the sampling period T_{samp}, input-dependent intermittent or *soft* errors occur whenever paths with delays longer than T_{samp} are excited. This leads to severe degradation in the signal-to-noise ratio (SNR).

ANT combined with VOS enables the design of low-power signal processing systems that operate at energy-efficiencies beyond those achieved by present-day systems. The overall approach of employing VOS in combination with ANT for low-power is referred to as *soft digital signal processing*. Soft DSP systems address energy-efficiency and reliability issues jointly.

Since the effect of increased clock frequency beyond a critical frequency f_{crit} is the same as VOS, ANT can also be used to design high-throughput systems using frequency overscaling. Further, ANT can also be employed to mitigate the effects of deep submicron (DSM) noise consisting of cosmic rays, ground bounce, crosstalk, or process variations [15]–[17] resulting in error-tolerant digital signal processing systems.

In this paper, we propose a novel ANT technique referred to as reduced precision redundancy (RPR) which combats soft-errors effectively while achieving significant energy savings. RPR employs a reduced precision replica of a DSP system [referred to as main DSP (MDSP)] to detect and correct the errors occurring at the output of the MDSP system. The proposed RPR-based ANT technique is distinct from previously proposed *prediction-based error-control* (PEC) [11] or *adaptive error-cancellation* (AEC) scheme [12]. While the PEC and AEC are effective for narrowband and broadband systems, respectively, the RPR technique can be applied to both. Reduction of precision has been employed in the past for power reduction [18]–[20]. These techniques trade off precision with SNR. However, since replica results are rarely used due to the infrequent VOS error, the proposed RPR technique can maintain the SNR. In fact, it is shown that RPR achieves better performance than a low-precision original MDSP.

The rest of the paper is organized as follows. Section II introduces the concept of the proposed RPR technique and its analysis. In Section III, we describe the architecture of an RPR digital signal processor in the context of the digital filtering and fast
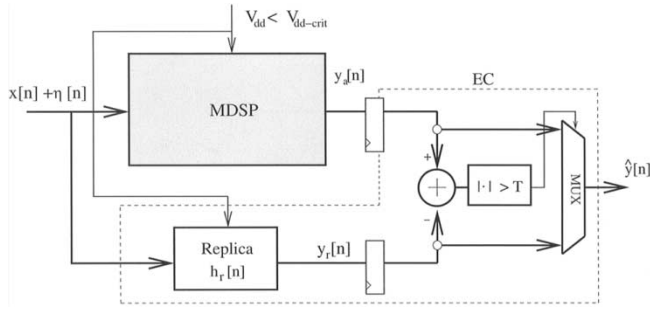
Fig. 1.   A DSP system employing RPR and VOS.

Fourier transform (FFT). In Section IV, we present simulation results that demonstrate the power savings.

II. REDUCED PRECISION REDUNDANCY

A DSP system incorporating RPR is shown in Fig. 1. The MDSP block is subject to VOS, which results in soft errors in its output $y_a[n]$. When a soft error in MDSP is detected using an error control (EC) block, the RPR output $y_r[n]$ is used as an output estimate $\hat{y}[n]$. Next, we describe the error characteristics of a system under VOS and then present the proposed error control algorithm.

A. Soft Error Characteristics

Voltage overscaling introduces input-dependent soft errors whenever a path with delay greater than the sample period T_{samp} is excited. Since the arithmetic units employed in DSP systems are based on least significant bit (LSB) first computation, soft errors appear first in the most significant bits (MSBs), resulting in errors of large magnitude. These errors severely degrade the performance but are desirable because they are easy to detect.

In general, a small fraction of input combinations excite longer paths. This fraction depends upon the delay distribution of a system, which in turn depends on the architecture. The path delay distribution for all possible input combinations of an 8×8 Baugh-Wooley multiplier is shown in Fig. 2. We observe that only 14% of the input combinations excite paths with delays greater than 75% of the critical path delay.

B. The RPR Technique

RPR utilizes a replica of the MDSP but with reduced precision operands. If the critical path delay of the replica is smaller than the sample period T_{samp} under VOS, the replica output $y_r[n]$ will not suffer from soft errors. Note that this condition is easily satisfied for array-based arithmetic units (e.g., ripple-carry adders and Baugh-Wooley multiplier) where the critical path delay decreases linearly with decrease in precision. The output $y_a[n]$ of MDSP can be written as

$$y_a[n] = y_o[n] + \gamma[n] = (s[n] + \eta[n]) + \gamma[n] \quad (2)$$

where $y_o[n]$ is the error free output composed of a desired signal $s[n]$ and channel noise $\eta[n]$, and $\gamma[n]$ is the soft error. The output SNR of the MDSP under VOS, SNR_{vos} is given by

$$\text{SNR}_{\text{vos}} = 10 \log_{10} \frac{\sigma_s^2}{\sigma_\eta^2 + \sigma_\gamma^2} \quad (3)$$

where σ_s^2, σ_η^2, and σ_γ^2 are the power of $s[n]$, $\eta[n]$, and $\gamma[n]$, respectively.

The replica output $y_r[n]$ is not equal to the $y_o[n]$ (the output of MDSP when $V_{\text{dd}} = V_{\text{dd-crit}}$) due to the LSB truncation noise. However, since VOS induces errors of large magnitude, we can employ $y_r[n]$ to detect errors in the MDSP output $y_a[n]$. Error detection is accomplished by comparing the difference $|y_a[n] - y_r[n]|$ against a threshold T_h. Error correction involves setting the final output $\hat{y}[n]$ to $y_r[n]$. Therefore, the decision rule for choosing the final output $\hat{y}[n]$ is given by

$$\hat{y}[n] = \begin{cases} y_a[n], & \text{if } |y_a[n] - y_r[n]| \leq T_h \\ y_r[n], & \text{if } |y_a[n] - y_r[n]| > T_h. \end{cases} \quad (4)$$

In order to guarantee that $\hat{y}[n] = y_o[n]$ when $y_a[n] = y_o[n]$ (i.e., the MDSP output equals the final output in the absence of errors), the threshold T_h is chosen as

$$T_h = \max_{\forall \text{ input}} |y_o[n] - y_r[n]|. \quad (5)$$

The SNR of an RPR-based scheme SNR_{rpr}, is given by

$$\text{SNR}_{\text{rpr}} = 10 \log_{10} \frac{\sigma_s^2}{\sigma_\eta^2 + \sigma_{\gamma_r}^2} \quad (6)$$

where $\sigma_{\gamma_r}^2 = E\left[|y_o[n] - \hat{y}[n]|^2\right]$ is the power of residual soft error in the corrected output $\hat{y}[n]$. In order to meet a specific desired SNR (SNR_{des}) imposed by the application at hand, we need to satisfy the following inequality

$$\text{SNR}_{\text{rpr}} \geq \text{SNR}_{\text{des}} = 10 \log_{10} \frac{\sigma_s^2}{\sigma_{\text{des}}^2} \quad (7)$$

which directly implies that energy savings without performance loss can be achieved if

$$\underbrace{\sigma_{\text{des}}^2 - \sigma_\eta^2}_{\text{system margin}} \geq \sigma_{\gamma_r}^2. \quad (8)$$

The system margin can be computed once a conventional system satisfying (7) is designed. Next, we describe the design of a proper replica satisfying the performance constraints in (8).

C. Quantization Noise Analysis

In this subsection, we first present the quantization noise analysis of replica. In what follows, we assume that the operand precision of the MDSP block to be equal to $B + 1$ bits and that of the replica to be $B_r + 1$ bits, where $B > B_r$. In addition, we
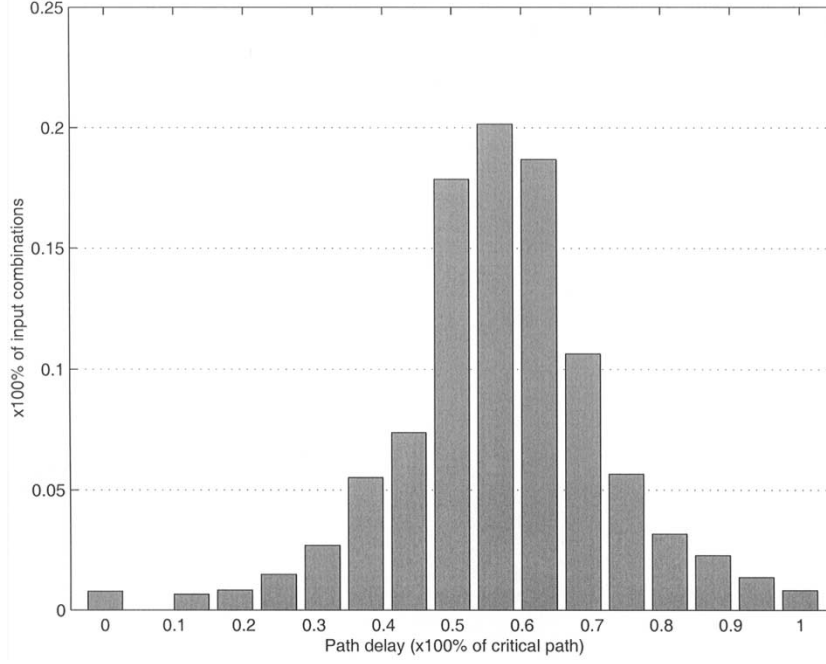
Fig. 2. Path delay distribution of an 8×8 Baugh-Wooley multiplier.

assume that all the quantization noise is due to truncation, and that both signal and quantization noise are uncorrelated.

A $B+1$ bit representation of the number x in two's complement representation is given by

$$x = -b_0 + \sum_{i=1}^{B} b_i 2^{-i}. \tag{9}$$

Similarly, x_r, the representation of x in the replica is

$$x_r = -b_0 + \sum_{i=1}^{B_r} b_i 2^{-i}. \tag{10}$$

Then the quantization noise q_x between original value x and x_r is defined as follows:

$$q_x = x - x_r = \sum_{i=B_r+1}^{B} b_i \cdot 2^{-i}. \tag{11}$$

Note that the maximum value of q_x, $q_{x,\text{max}}$ is

$$q_{x,\text{max}} = \sum_{i=B_r+1}^{B} 2^{-i} = 2^{-B}(2^{B-B_r} - 1) \tag{12}$$

and minimum value of q_x is clearly 0 (when all $b_i = 0$). Note that q_x is always nonnegative.

Let us denote the quantization step size of MDSP and replica to be $\Delta_o = 1/2^B$ and $\Delta_r = 1/2^{B_r}$, respectively. First, we compute the mean (μ_q) and the power (σ_q^2) of the quantization noise q_x for uniformly distributed input x.

Lemma 1: If discrete input sample x is uniformly distributed in $[-1, 1)$, the mean and the power of the quantization noise q_x are respectively

$$\mu_q = \frac{1}{2}(\Delta_r - \Delta_o) \tag{13}$$

$$\sigma_q^2 = \frac{1}{6}\left(2\Delta_r^2 - 3\Delta_r\Delta_o + \Delta_o^2\right) \tag{14}$$

where Δ_r and Δ_o are the quantization step sizes of x in the replica and the MDSP block, respectively.

Proof: Since x is uniformly distributed, q_x takes discrete nonnegative values $n_i = i \cdot \Delta_o$, $i = 0, 1, \ldots, 2^{B-B_r} - 1$ with probability mass function $\Pr(q_x = n_i) = 1/2^{B-B_r} = \Delta_o/\Delta_r$. Then, the mean can be calculated as

$$\mu_q = E[q_x] = \sum_{i=0}^{2^{B-B_r}-1} n_i \Pr(q_x = n_i)$$
$$= \sum_{i=0}^{(\Delta_r/\Delta_o)-1} i\Delta_o \frac{\Delta_o}{\Delta_r}$$
$$= \frac{1}{2}(\Delta_r - \Delta_o). \tag{15}$$

The quantization noise power $\sigma_q^2 = E[|x - x_r|^2]$ is

$$\sigma_q^2 = \sum_{i=0}^{(\Delta_r/\Delta_o)-1} (i\Delta_o)^2 \frac{\Delta_o}{\Delta_r}$$
$$= \frac{1}{6}\left(2\Delta_r^2 - 3\Delta_r\Delta_o + \Delta_o^2\right). \tag{16}$$

∎

Next, we investigate the quantization noise at the output of a reduced precision multiplier $x_r \times h_r$, with reference to a full precision multiplier with operands $x = x_r + q_x$ and $h = h_r + q_h$. In most DSP applications, one operand is the signal and the other operand is the coefficient. Therefore, we regard the signal x as a uniformly distributed random variable and the coefficient h as a constant.
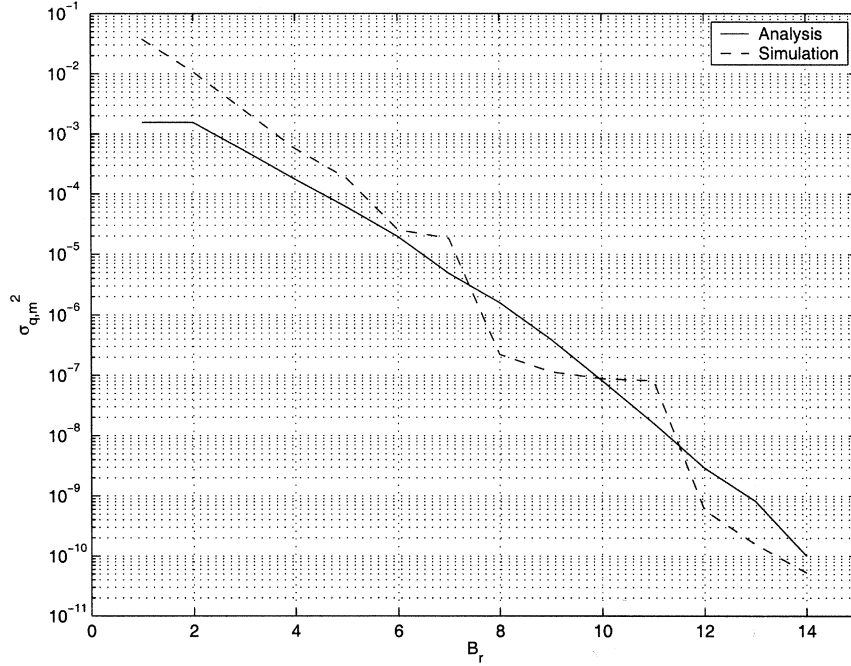
Fig. 3. Quantization noise power of multiplication $x \times h$ for the reference precision $B_1 = 15$.

Lemma 2: The mean $\left(\mu_{q_m}\right)$ and the noise power $\left(\sigma_{q_m}^2\right)$ of the quantization noise q_m at the output of a multiplier $y_r = x_r \times h_r$ with respect to $y = x \times h$ is given by

$$\mu_{q_m} = \frac{h_r}{2}(\Delta_r - \Delta_o) \qquad (17)$$

$$\sigma_{q_m}^2 = q_h^2 \sigma_x^2 + \frac{h_r^2}{6}\left(2\Delta_r^2 - 3\Delta_r\Delta_o + \Delta_o^2\right) \qquad (18)$$

where Δ_r and Δ_o are the quantization step sizes of x in the replica and the MDSP block, respectively.

Proof: See Appendix A. ∎

Fig. 3 compares $\sigma_{q_m}^2$ for a multiplier obtained from simulations with that computed from (18) indicating that (18) is accurate.

D. Residual Noise Power Analysis

In Section II-C, we obtained the quantization noise power $\sigma_{q_m}^2$ of replica. By combining VOS error probability $P_e(K_{\text{vos}})$ with $\sigma_{q_m}^2$, we obtain the residual noise power $\sigma_{\gamma_r}^2$ of RPR scheme in this subsection.

First, we compute the error probability due to VOS. The propagation delay τ of a logic gate in CMOS process technology [4] is given by

$$\tau = \frac{C_L V_{\text{dd}}}{\beta(V_{\text{dd}} - V_t)^\alpha} \qquad (19)$$

where C_L is the load capacitance, V_{dd} is the supply voltage, V_t is the device threshold voltage, β is the device transconductance, and α is the velocity saturation index. As the supply voltage is reduced, gate delay increases thereby increasing the error probability.

Definition 1: The cumulative distribution function (CDF) of a path delay random variable T, $F_T(t_p)$ is defined as

$$F_T(t_p) = P(T \le t_p) \qquad (20)$$

where $0 \le t_p \le T_{\text{cp}}$ and T_{cp} is the critical path delay. $F_T(t_p)$ denotes the probability that the path delay is less than or equal to a specified value t_p.

Lemma 3: For the given path delay CDF of system, $F_T(t_p)$, the error probability due to VOS P_e is given by

$$\begin{aligned} P_e &= \Pr(y_a[n] \neq y_o[n]) \\ &= 1 - F_T\left(\frac{T_{\text{samp}}}{K_{\text{vos}}}\left(\frac{K_{\text{vos}}V_{\text{dd}} - V_t}{V_{\text{dd}} - V_t}\right)^\alpha\right) \\ &= P_e(K_{\text{vos}}) \end{aligned} \qquad (21)$$

Proof: See Appendix B. ∎

Note that Fig. 4 shows a plot of error probability function in (21) for a 8×8 Baugh-Wooley multiplier whose delay distribution was shown in Fig. 2. We observe that the error probability does not increase significantly until K_{vos} approaches 0.7. In addition, only about 5% of inputs result in output errors even at $K_{\text{vos}} = 0.6$.

We now seek an upper and a lower bound on the noise power $\sigma_{\gamma_r}^2$ in the RPR system by combining $P_e(K_{\text{vos}})$ and $\sigma_{q_m}^2$,

Theorem 1: For a given VOSF K_{vos}, the residual noise power $\sigma_{\gamma_r}^2$ at the output of a multiplier in an RPR-based system is bounded by

$$P_e \sigma_{q_m}^2 \le \sigma_{\gamma_r}^2 \le P_e\left(2\sigma_{q_m}^2 + T_h^2 + 2T_h\mu_{q_m}\right). \qquad (22)$$
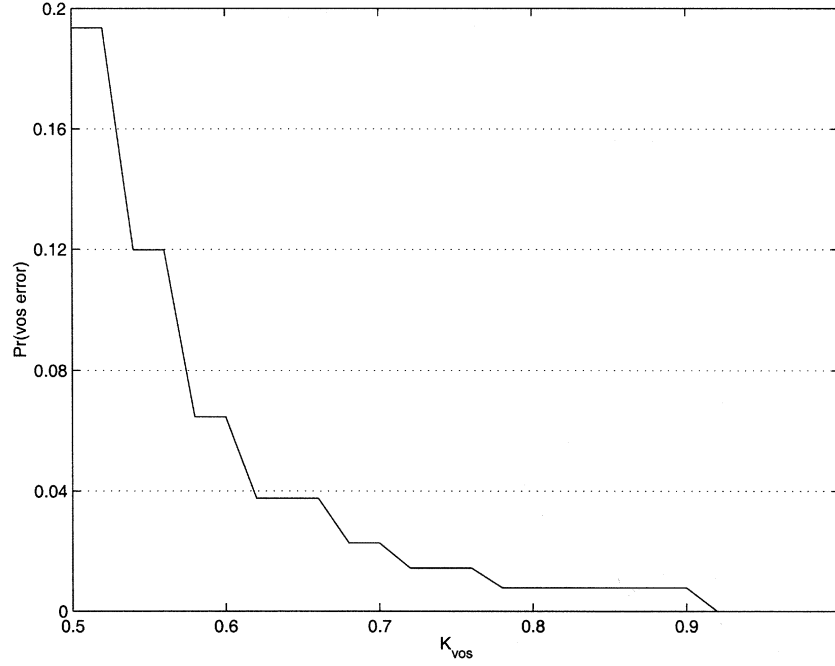
Proof: See Appendix C. ∎

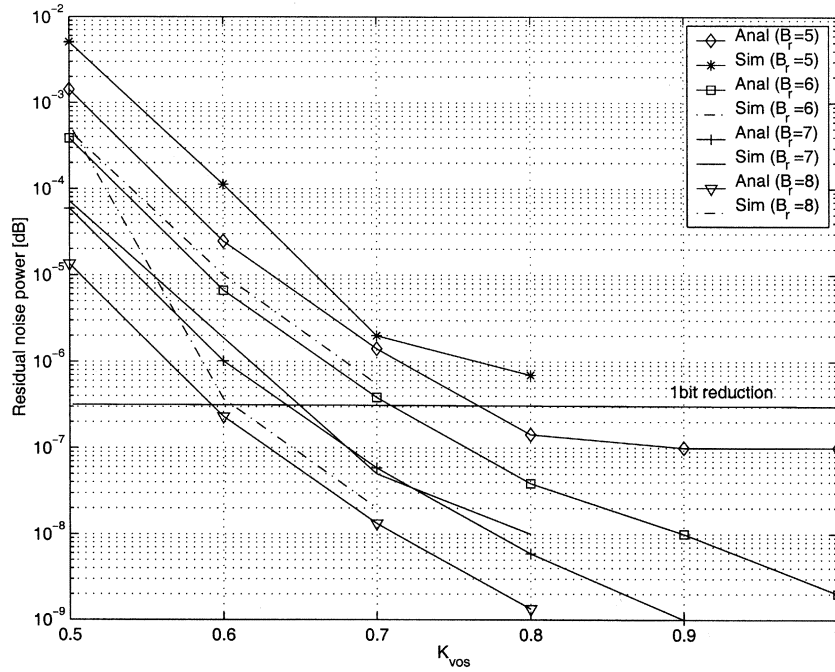Fig. 4. $P_e(K_{\text{vos}})$ versus K_{vos} ($\alpha = 1.2$).

Fig. 5. Performance analysis and simulation results of residual noise power $\sigma^2_{\gamma_r}$.

The bound in (22) depends on the soft error probability P_e and the precision of replica. From (22), we can obtain a bound on the signal-to-quantization noise (SQNR) in an RPR-based system.

Corollary 1: The SQNR at the output of a multiplier $y = x \times h$ employing RPR is bounded by

$$\frac{h^2 \sigma_x^2}{P_e \left(2\sigma_{q_m}^2 + T_h^2 + 2T_h \mu_{q_m}\right)} \leq SQNR_{\text{rpr}} \leq \frac{h^2 \sigma_x^2}{P_e \sigma_{q_m}^2} \quad (23)$$

where μ_{q_m} and $\sigma^2_{q_m}$ are given by (17) and (18) and $P_e(K_{\text{vos}})$ by (21).

Fig. 5 compares the results of analysis (22) and simulation results for RPR with $B = 12$ as B_r and K_{vos} vary. We observe that the achievable values of K_{vos} reduces as B_r increases. However, (22) is no longer valid at low K_{vos} region where the replica also begins to generate errors (e.g., $B_r = 8$, $K_{\text{vos}} < 0.6$ in Fig. 5).

Fig. 5 also shows that $\sigma^2_{\gamma_r}$ is lower than the noise power of an MDSP block whose precision is reduced by 1-bit over a wide range of K_{vos}. This implies that simple 1-bit reduction in the

TABLE I
ALGORITHM TO DETERMINE THE OPTIMUM REPLICA PRECISION B_r

Step	Procedure
1.	Assign initial value $K_{vos} = 1$, $B_r = 0$, and $B_{opt} = 0$.
	Compute the initial power consumption of RPR DSP and save it as $P_{RPR,\min}$.
2.	Reduce K_{vos} by specified amount Δ_K, i.e., $K'_{vos} = K_{vos} - \Delta_K$.
3.	Increase B_r by 1. If (29) is violated, go to step 2.
	If $B_r = B$, stop the process and exit.
4.	Compute the noise power $\sigma_{q_r}^2$ in (22). If it does not satisfy (28), go to step 3.
	Otherwise, compute the power consumption and save it as $P_{RPR,\min}(K_{vos})$.
5.	If $P_{RPR,\min} > P_{RPR,\min}(K_{vos})$, then $P_{RPR,\min} = P_{RPR,\min}(K_{vos})$ and $B_{opt} = B_r$.
6.	Go to step 2.

MDSP precision at $V_{dd-crit}$ will give a lower SNR than RPR indicating that RPR provides nontrivial benefits.

E. Optimum Precision Selection

The dynamic power dissipation of an original MDSP system at $V_{dd-crit}$ is given by

$$P_{org} = C_L V_{dd-crit}^2 f_{clk} \qquad (24)$$

where C_L is the effective capacitance of the MDSP and f_{clk} is the system clock frequency. The power dissipation of an RPR-based system is given by

$$P_{RPR} = (C_L + C_{EC})(K_{vos} \cdot V_{dd-crit})^2 f_{clk} \qquad (25)$$

where C_{EC} is the effective switching capacitance of the error control block. In order to guarantee power savings, i.e., $P_{RPR} \leq P_{org}$, we can show from (24) and (25) that

$$C_{EC} \frac{K_{vos}^2}{1 - K_{vos}^2} \leq C_L. \qquad (26)$$

As mentioned, noise power $\sigma_{\gamma_r}^2$ in the RPR scheme depends on soft error probability P_e and the replica precision B_r. Note that $\sigma_{\gamma_r}^2$ will be reduced when B_r is high at the expense of power. Likewise, P_e trades off performance and power as a function of K_{vos} [see (21)]. Therefore, our goal is to determine values for K_{vos} and B_r that results in minimum power dissipation while satisfying the SNR requirements. The problem is expressed as follows:

$$\text{minimize } P_{RPR} = (C_L + C_{EC})(K_{vos} \cdot V_{dd-crit})^2 f_{clk} \qquad (27)$$

subject to

$$\sigma_{des}^2 - \sigma_n^2 \geq \sigma_{\gamma_r}^2 \qquad (28)$$

$$C_{EC} \frac{K_{vos}^2}{1 - K_{vos}^2} \leq C_L. \qquad (29)$$

The solution of problem stated in (27)–(29) can be found using a two dimensional search method which can be implemented easily. The key idea is to search for a boundary point of two dimensional regions consisting of B_r and K_{vos} axis which satisfies noise and power constraints. Specifically, for a given K_{vos},
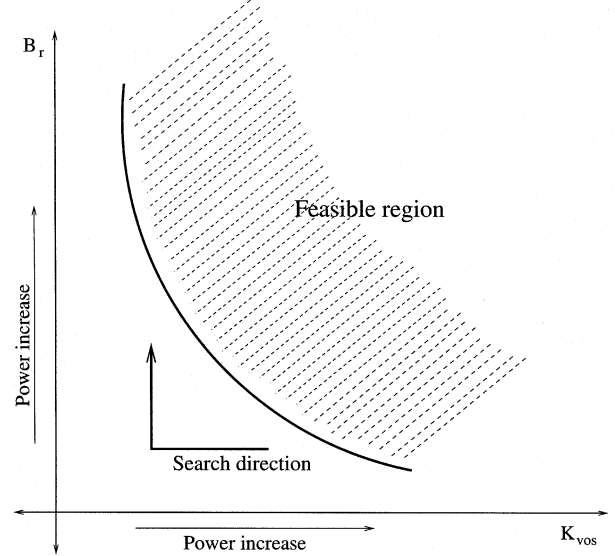

Fig. 6. Illustration of optimum replica precision search.

we increase B_r (vertical axis in Fig. 6) until it satisfies (28) and then repeat this step for the next K_{vos}. Notice that we do not need to increase B_r beyond the boundary point since the power consumption only increases. This algorithm is described in Table I.

The feasible region and power consumption for 12×12 bit multiplier-and-accumulator (MAC), found by the proposed algorithm, is shown in Fig. 7(a), where we used a 32 tap low-pass filter with cutoff frequency $\omega = 0.2\pi$ and the constraint of noise power $\sigma_{\gamma_r}^2 \leq -30$ dB is assumed. We also assumed the lumped capacitance of $b \times b$ MAC to be proportional to $(b^2 + b)$ [22]. It can be seen from Fig. 7(b) that at each K_{vos}, the point of maximum power savings occurs at the boundary. By following the search procedure described above, we obtained $K_{vos} = 0.52$, and the optimum power savings of 63.25% at a replica precision of $B_r = 7$. Note that if K_{vos} is decreased beyond this point, soft error probability P_e will increase abruptly and noise power constraints would not be satisfied.

III. RPR DIGITAL SIGNAL PROCESSOR

In this section, we describe DSP architectures for RPR based systems. Two of the most common DSP applications used in communication systems are digital filtering and FFT. Thus, we develop the RPR architectures for the MAC in digital filters and multipliers in FFT processors.
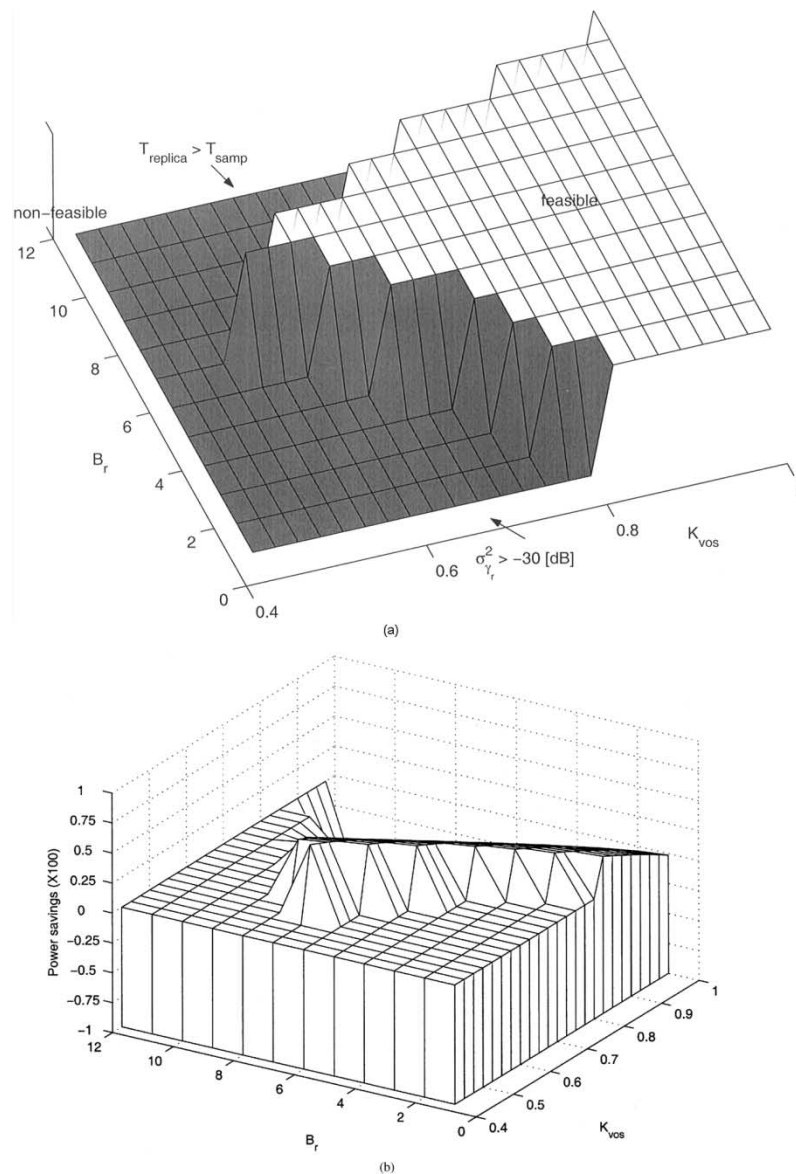
Fig. 7.   Precision optimization for the replica in a 12×12 bit MAC MDSP system. (a) Feasible region. (b) Power savings.

A. RPR for Digital Filtering

Fig. 8(a) shows the proposed folded RPR FIR architecture. Along with the main MAC, a replica MAC is employed for generating estimates of main MAC. The operands to the replica MAC are the same as the main MAC but have a smaller precision, which makes it immune to VOS errors. After executing an N-tap multiply-accumulate operation, the results of the main MAC and replica MAC are compared for error detection. If an error is detected then the result of the replica MAC is chosen, otherwise, the main MAC output is selected as the final output.

Fig. 8(b) shows an unfolded RPR digital filter. The unfolded RPR filtering has a one cycle latency. However, it does not suffers any loss in throughput. The overhead of the proposed scheme includes the replica filter and the decision block (subtractor and comparator). Employing the fact that soft errors are mostly of large magnitude, the complexity of the error control block can be reduced significantly, as will be discussed in Section III-C.

B. RPR for FFT

Here, we consider a radix-2 decimation-in-time (DIT) based FFT processor. The processor's datapath computes one complex radix-2 DIT butterfly per cycle [23]. As shown in Fig. 9(a), the DIT butterfly calculates two outputs, $Y_1 = X_1 + WX_2$ and $Y_2 = X_1 - WX_2$, from two inputs X_1 and X_2, and a twiddle factor W. It is assumed that appropriate pipelining is employed to route data between the memory and the functional units in order to maximize throughput. Apart from the main memory, multipliers are the largest functional units in a VLSI implementation of such a processor [24].

In this paper, we consider a 64-point FFT processor with 16-bit precision operating on 10-bit fixed-point complex inputs, which are typical parameters for an FFT in a wireless local area network (WLAN) orthogonal frequency division multiplexing (OFDM) modem [25]. Four multipliers along with two adders multiply W and X_2 and four additional adders generate the stage output Y_1 and Y_2. Since the fixed-point data
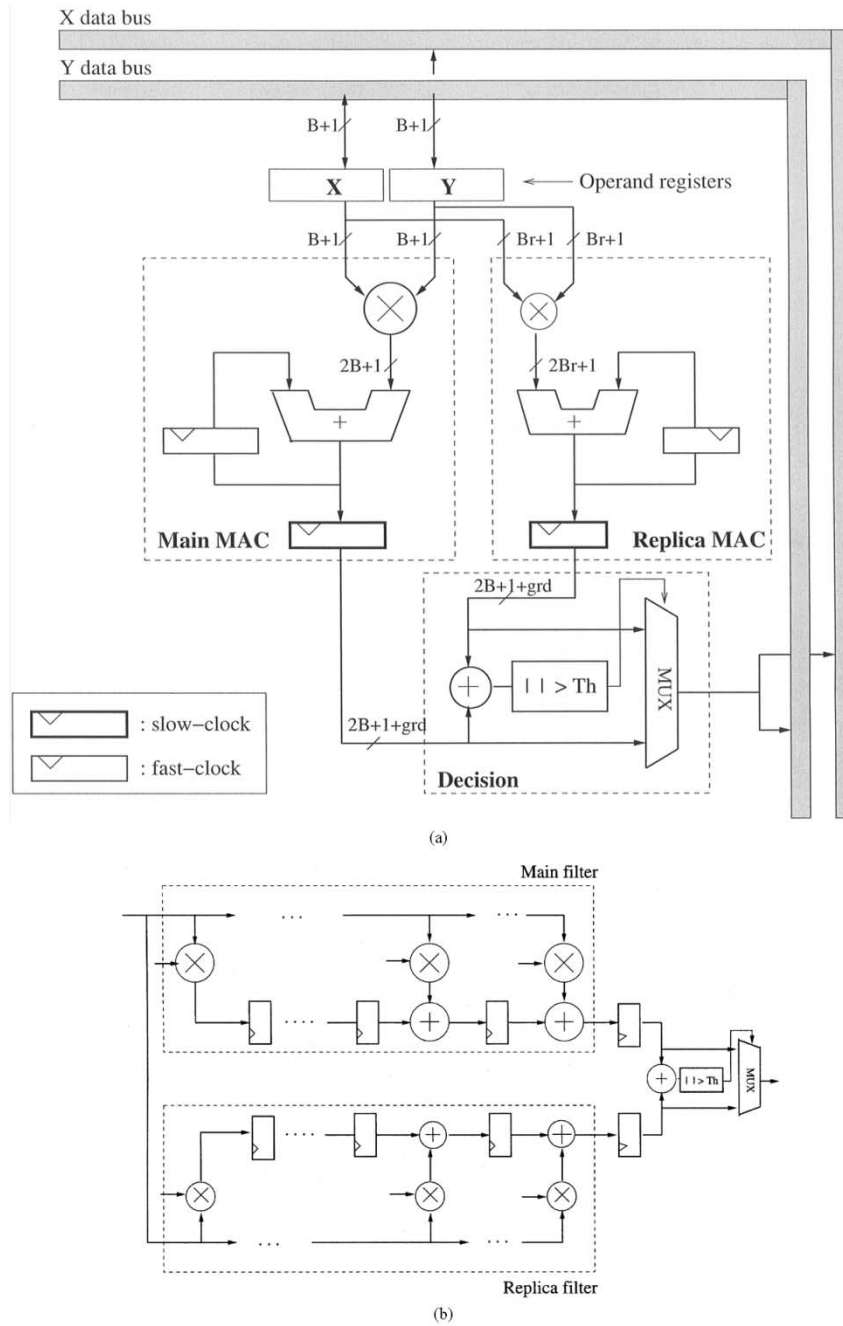
Fig. 8.   Proposed RPR based digital filtering: (a) DSP architecture and (b) its application in an FIR filter.
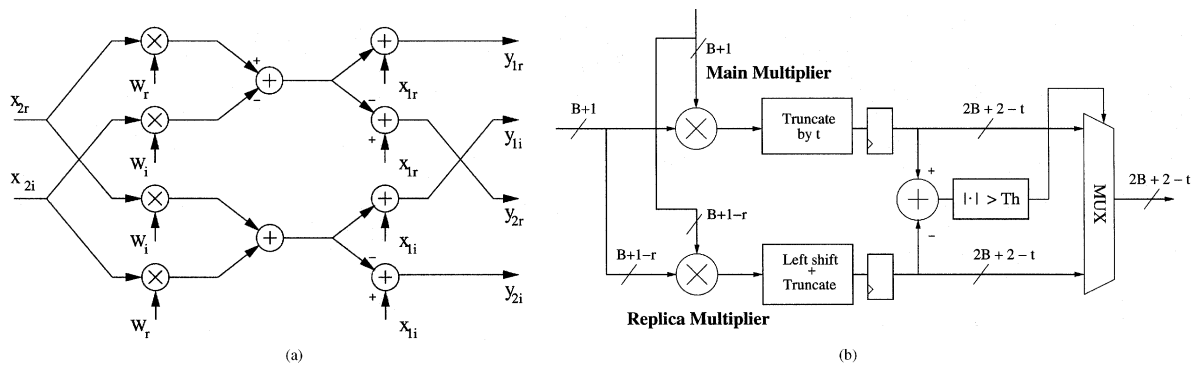

Fig. 9.   Proposed RPR based MAC architecture for FFT: (a) butterfly in DIT FFT, and (b) architecture of multiplier employing RPR.
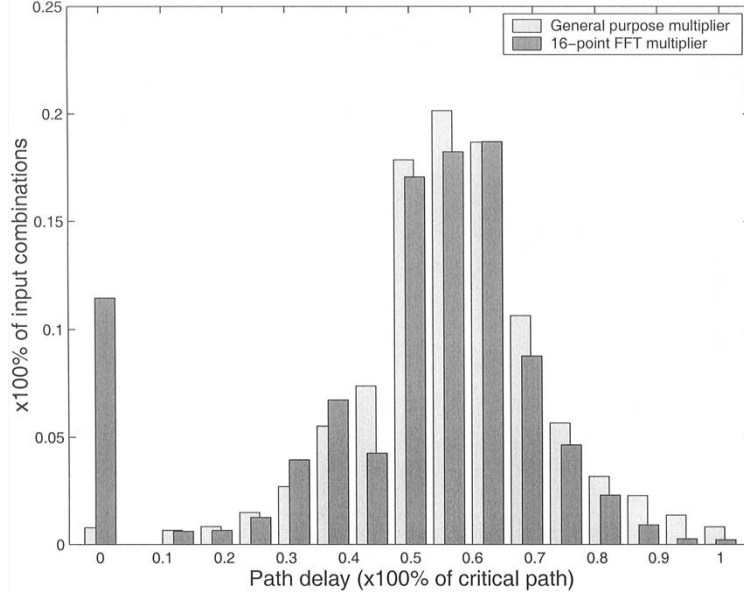
Fig. 10.   Path-delay distribution of an 8×8 Baugh-Wooley multiplier.

format requires the eventual truncation of the butterfly outputs when writing the outputs back to memory, computation of all the multiplier output bits is unnecessary [23]. Therefore, the outputs of the multipliers are truncated by t bits. The multiplier structure of RPR FFT is shown in Fig. 9(b), where a $B_r = B + 1 - r$ bit reduced precision multiplier is employed.

The multiplier used in FFT processor is more amenable to VOS than a general purpose array multiplier due to the fact that the real and imaginary parts of the twiddle factor take only $N/2 + 1$ distinct values in an N-point FFT. For example, in a 16-point FFT, the twiddle factor components take only $16/2 + 1 = 9$ distinct values among all the possible $2^8 = 256$ values. Fig. 10 compares the path delay histograms of the general purpose multiplier and the FFT multiplier in which one of the operands takes only the 9 possible twiddle factor values. We can easily observe the significant reduction in the percentage of longer paths. In particular, only 8% of the input combinations excite paths with delays longer than 75% of the critical path delay.

C. Error Control Block

The error control block described in Sections III-A and B requires a subtractor and a comparator followed by a 2-to-1 multiplexer. As discussed in Section II-B, the maximum difference between the replica output and MDSP output at $V_{\mathrm{dd-crit}}$ is used as the decision threshold in the comparator.

The input to the RPR MAC (or multiplier) suffers maximum quantization noise when all the truncated bits in the operand are 1. We denote the number of truncated bits as $r = B - B_r$ and use integer representation for notational convenience. Since the largest number in magnitude of B bits 2's complement representation is -2^B, the maximum difference occurs when both input operands are $-2^B + 2^r - 1$. In this case, the decision threshold T_h is

$$
\begin{aligned}
T_h &= |(-2^B + 2^r - 1)(-2^B + 2^r - 1) - (-2^B)(-2^B)| \\
&= (2^{B+r+1} - 2^{B+1} - 2^{2r} + 2^{r+1} - 1).
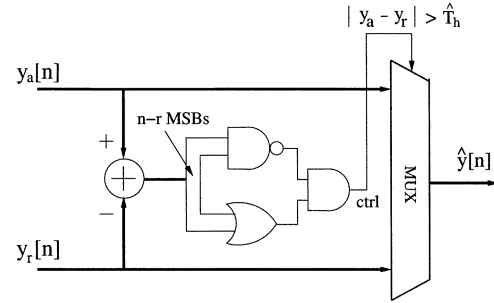\end{aligned} \tag{30}
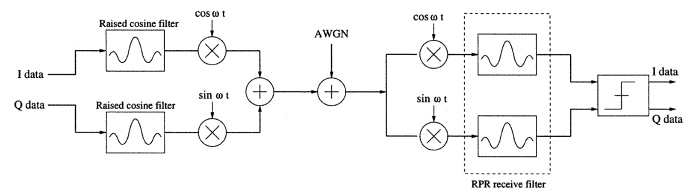$$

Fig. 11.   Decision block structure.

Fig. 12.   Simulation setup for RPR receive filter in QPSK.

Typically, for large B and $r \approx B/2$, $2^{B+r+1} \gg (2^{B+1} + 2^{2r} - 2^{r+1} + 1)$. Thus, T_h becomes approximately

$$
\hat{T}_h = 2^{B+r+1}. \tag{31}
$$

By choosing the threshold $\hat{T}_h$ in (31), the comparator can be realized with the simple circuit consisting of AND, NAND, and OR shown in Fig. 11. Notice that the outputs of NAND and OR gate need to be logic 1 to enable the ctrl signal, which corresponds to the condition $|y_a - y_r| > \hat{T}_h$. This is much simpler than a full-blown implementation requiring $(B+1)$-bit full adder. Even in the case where one operand is fixed, a similar structure can be employed by recognizing the fact that performance is insensitive to small changes in the threshold.

IV. SIMULATION RESULTS AND DISCUSSION

In this section, we discuss the performance of the proposed RPR ANT technique. First, we define the measure for power
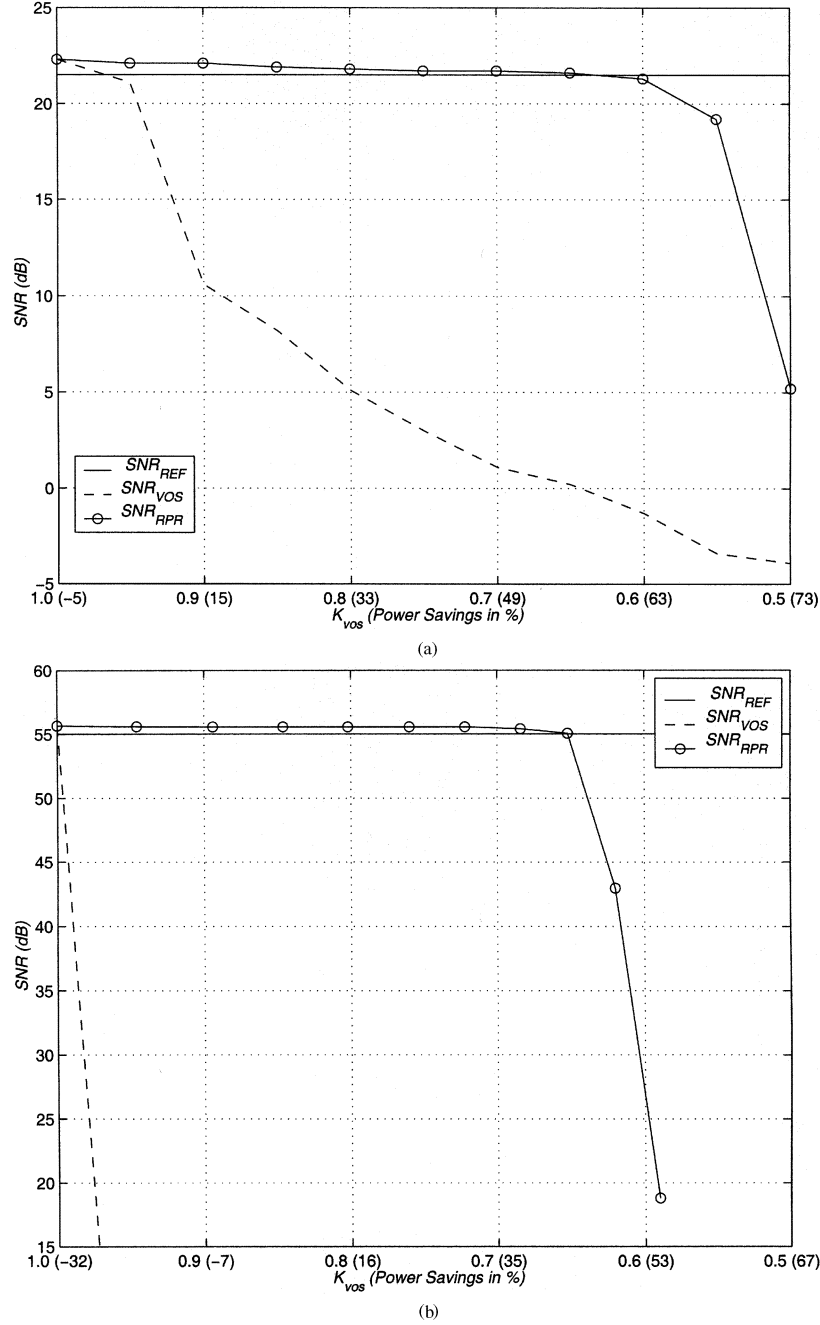
Fig. 13. Power savings via RPR based ANT technique: (a) 16-tap QPSK receiver filtering and (b) a 64-point FFT.

savings and then discuss the power savings of a proposed scheme.

A. Measure for Power Savings

The power savings P_{sav} of the proposed RPR scheme is given by

$$P_{\mathrm{sav}} = \frac{P_{\mathrm{ref}} - P_{\mathrm{RPR}}}{P_{\mathrm{ref}}} \times 100\% \qquad (32)$$

where P_{ref} and P_{RPR} are the power dissipation of MDSP at $V_{\mathrm{dd-crit}}$ and RPR DSP with VOS, respectively. Because the dynamic power dissipation depends on the square of a supply voltage, $P_{\mathrm{RPR}} = K_{\mathrm{vos}}^2(P_{\mathrm{ref}} + P_{\mathrm{ovh}})$, where P_{ovh} is the power

overhead due to error canceller including replica and error decision block. Then the percent power saving is given by

$$P_{\mathrm{sav}} = \left[1 - K_{\mathrm{vos}}^2\left(1 + \frac{P_{\mathrm{ovh}}}{P_{\mathrm{ref}}}\right)\right] \times 100\%. \qquad (33)$$

B. Simulation Results

In this section, the performance of the digital filter and the FFT butterfly are studied in the context of communication systems. Simulations are performed assuming a 0.25 μm, 2.5 V CMOS process technology with velocity saturation index $\alpha = 1.2$. Once T_{samp} is set to the critical path delay of the system at $V_{\mathrm{dd-crit}} = 2.5$ V, the delay of system at each $V_{\mathrm{dd}}(< V_{\mathrm{dd-crit}})$
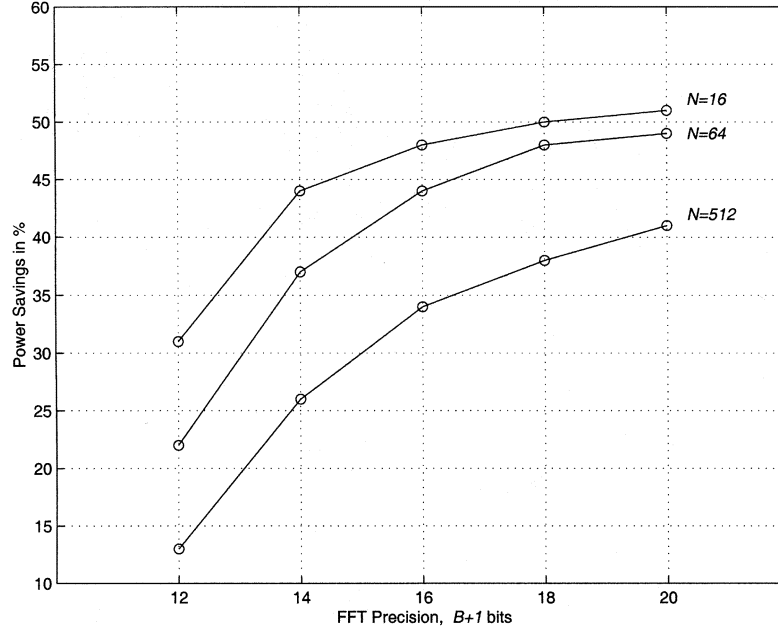
Fig. 14.   Power savings for various FFT sizes.

is obtained by scaling using (19). At each V_{dd}, the gate-level power simulator MED [26] is employed to estimate the energy savings obtained via voltage reduction as proposed.

In the simulations for digital filter, receive filtering of a QPSK communication system in the presence of AWGN noise is considered. To achieve bit error rate of 10^{-7}, decoder slicer input SNR should be 21.5 dB including a 6 dB margin [27]. Conventional filters (optimized at $V_{dd-crit}$) have been designed to meet this performance specification, where 16 taps with $B = 12$ bit MAC satisfies the requirements with minimal complexity. The replica MAC precision $B_r = 7$ bit and, in both cases, a Baugh-Wooley signed multiplier is employed [28]. The plot of K_{vos} versus SNR for receiver equalization employing the proposed RPR scheme is shown in Fig. 13(a). While the conventional filter suffers sharp SNR drop when K_{vos} is reduced, the proposed RPR technique maintains desired performance near $K_{vos} = 0.6$. In this case, achievable energy savings over a conventional MAC that operates at $V_{dd-crit}$ is 60%. Note that the error rate at this value of K_{vos} is 0.09/sample. Beyond this point, VOS affects the replica MAC and therefore the assumption in Section II-B is violated. As a result, reliable error control is no longer possible.

For the FFT butterfly simulations, we consider an FFT processor that has typical WLAN OFDM parameters [25] with FFT length $N = 64$, FFT precision $B = 15$ bits and input precision of 10 bits. The inputs of replica multipliers are truncated by $r = 7$ and the internal truncation t is set to 12 bits for the reference multiplier and 9 bits for the replica. Fig. 13(b) plots SNR versus K_{vos} and the corresponding power savings. The SNR_{des} when $V_{dd} = V_{dd-crit}$ equals 55 dB. We can observe that $SNR_{rpr} \geq SNR_{des}$ is satisfied until $K_{vos} = 0.65$ (approximately 44% power savings) while SNR_{vos} falls off rapidly even when K_{vos} decreases slightly. The error rate at this value of K_{vos} is 0.06/sample. Note that about 27% power savings can be

achieved by directly reducing the FFT precision by 1-bit without VOS. This, however, results in an SNR loss of 3 dB which fails to meet the SNR requirements.

The achievable power savings for the proposed FFT scheme depends on the length and precision of the FFT due to variation in the path delay distribution. Fig. 14 shows the variation of power savings with FFT precision of $B + 1$ and FFT length N. When the precision is fixed, power savings decrease with increase in FFT length since the frequency of longer paths that fail at a given K_{vos} increases. We also observe that power savings increase with the FFT precision.

Finally, in order to compare the area overhead of RPR system over the conventional MDSP, we synthesized layouts for both systems in 0.25 μm process technology. We designed the receive filter for QPSK system described earlier using VHDL and synthesized the layouts via *Synopsys Design Analyzer* and *Cadence Silicon Ensemble*. We also estimated the power consumption via *Nanosim*. Fig. 15 shows the layouts for MDSP and the proposed RPR system, respectively. We define the area overhead ρ of RPR over the MDSP as

$$\rho = \left(\frac{A_{RPR}}{A_{mdsp}} - 1 \right) \times 100\%. \tag{34}$$

Substituting $A_{mdsp} = (457.2 \, \mu m)^2$ and $A_{rpr} = (511.125 \, \mu m)^2$ in (34), we obtain the area overhead $\rho \sim 20\%$. In addition, the power consumption of MDSP layout at $V_{dd-crit} = 2.5$ V is 229 μW @ 100 MHz and that of RPR at 0.65 $V_{dd-crit}$ is 96 μW @ 100 MHz resulting in power savings of 58.1%. This example clearly demonstrates the power savings achievable through RPR. These power savings can be improved significantly by employing separate supplies for the replica and the MDSP blocks and sizing the replica transistors differently from those in the MDSP block.
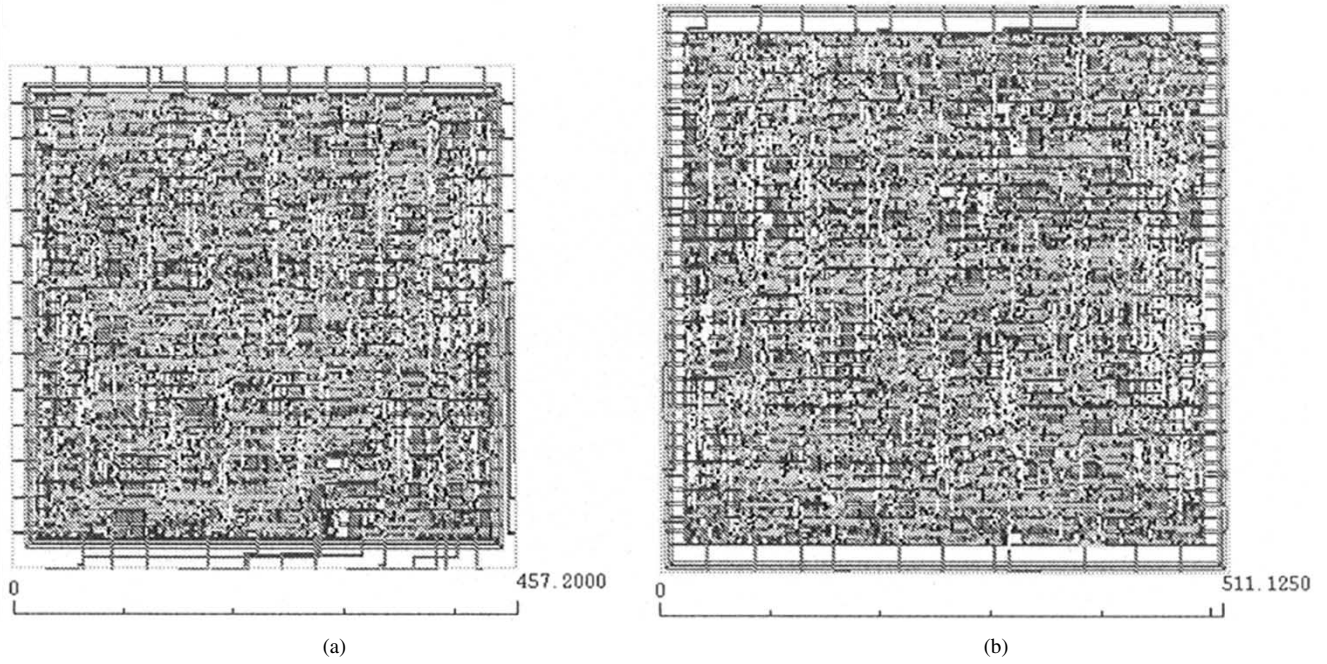
Fig. 15.   Synthesized layouts of: (a) MDSP and (b) RPR.

V. CONCLUSIONS

In this paper, we have proposed a novel algorithmic noise-tolerance technique referred to as reduced precision redundancy (RPR) to combat errors in hardware. Combining ANT and VOS results in soft digial signal processing systems that consume much less power than systems operating error-free at critical supply voltages. The RPR scheme was shown to be very effective in mitigating system-level errors via analysis, and simulations on circuit layouts in 0.25-μm CMOS for a QPSK receive filter and an FFT block.

Soft digital signal processing systems can reduce leakage power and provide robustness to errors caused by leakage currents. This topic requires future study. Noise sources such as cosmic rays and alpha particles can impact the error-control blocks as well. Future research needs to be directed toward the problem of efficient error-control in the presence of errors in the error-correction blocks. ANT is an elegant approach for trading-off reliability and energy-efficiency in deep submicron systems.

APPENDIX A
PROOF OF LEMMA 2

The quantization noise in a multiplication is defined as $q_m = xh - x_r h_r$. The mean is

$$
\begin{aligned}
\mu_{q_m} &= E[q_m] \\
&= E[(x_r + q_x)(h_r + q_h) - x_r h_r] \\
&= E[xq_h + q_x h_r].
\end{aligned} \tag{A-1}
$$

Since h is deterministic, we get

$$
\mu_{q_m} = q_h \mu_x + h_r \mu_q \tag{A-2}
$$

where μ_q has been computed in Lemma 1. Substituting μ_q from (13) and using $\mu_x = 0$ as x is uniformly distributed in $[-1, 1)$, we get

$$
\mu_{q_m} = \frac{h_r}{2}(\Delta_r - \Delta_o). \tag{A-3}
$$

The power $\sigma_{q_m}^2 = E\left[q_m^2\right]$ is

$$
\begin{aligned}
\sigma_{q_m}^2 &= E\left[(xq_h + q_x h_r)^2\right] \\
&= E\left[x^2 q_h^2 + 2xq_h q_x h_r + h_r^2 q_x^2\right] \\
&= q_h^2 E[x^2] + 2q_h h_r E[xq_x] + h_r^2 E\left[q_x^2\right] \\
&= q_h^2 \left(\sigma_x^2 + \mu_x^2\right) + q_h h_r \mu_x (\Delta_r - \Delta_o) \\
&\quad + \frac{h_r^2}{6}\left(2\Delta_r^2 - 3\Delta_r \Delta_o + \Delta_o^2\right).
\end{aligned} \tag{A-4}
$$

Using $\mu_x = 0$, we get

$$
\sigma_{q_m}^2 = q_h^2 \sigma_x^2 + \frac{h_r^2}{6}\left(2\Delta_r^2 - 3\Delta_r \Delta_o + \Delta_o^2\right). \tag{A-5}
$$

APPENDIX B
PROOF OF LEMMA 3

Consider a path with M logic gates. When the supply voltage is reduced from V_{dd} to $K_{\mathrm{vos}} V_{\mathrm{dd}}$, the delay τ for a single gate [see (19)] increases to

$$
\tau' = \frac{C_L K_{\mathrm{vos}} V_{\mathrm{dd}}}{\beta(K_{\mathrm{vos}} V_{\mathrm{dd}} - V_t)^\alpha}. \tag{B-1}
$$

Clearly the original path delay $t_{\mathrm{org}} = M\tau$ becomes $t_{\mathrm{vos}} = M\tau'$. Therefore, as shown in Fig. 16, shaded region in an original delay distribution becomes VOS error region
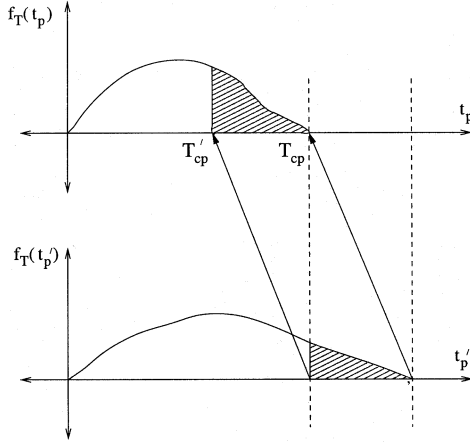
Fig. 16. VOS error region for the path delay distribution variation.

and the error probability is obtained by integrating this region, i.e., $\int_{T'_{cp}}^{T_{cp}} f_T(t_p) dt_p$. By using (19) and (B-1), one can show that the point T'_{cp} which starts to generate an error when VOS is applied is

$$T'_{cp} = \frac{T_{cp}}{K_{vos}} \left(\frac{K_{vos} V_{dd} - V_t}{V_{dd} - V_t} \right)^{\alpha}. \quad (B-2)$$

Thus, VOS error probability is now given by

$$P_e(K_{vos}) = \int_{T'_{cp}}^{T_{cp}} f_T(t_p) dt_p$$
$$= 1 - F_T \left(\frac{T_{cp}}{K_{vos}} \left(\frac{K_{vos} V_{dd} - V_t}{V_{dd} - V_t} \right)^{\alpha} \right).$$

Note that $F_T(T_{cp}) = 1$.

APPENDIX C
PROOF OF THEOREM 1

In this proof, we drop the time index n for notational convenience. Recall that the quantization noise power $\sigma_{\gamma_r}^2$ employing RPR is

$$\sigma_{\gamma_r}^2 = E \left[a|y_o - \hat{y}|^2 \right]. \quad (C-1)$$

If T_h is chosen according to (5), no false alarm event can occur during error detection and hence three possible scenarios exist: (i) no error (C-2), (ii) undetected error (C-3), and (iii) detected error (C-4).

$$y_a = y_o (\text{thus} |y_a - y_r| \leq T_h) \Rightarrow \hat{y} = y_a = y_o \quad (C-2)$$
$$y_a \neq y_o \text{ and} |y_a - y_r| \leq T_h \Rightarrow \hat{y} = y_a \quad (C-3)$$
$$y_a \neq y_o \text{ and} |y_a - y_r| > T_h \Rightarrow \hat{y} = y_r. \quad (C-4)$$

Thus, (C-1) can be rewritten as

$$\sigma_{\gamma_r}^2 = \Pr(|y_a - y_r| > T_h) \cdot E \left[|y_o - y_r|^2 \right] \text{ (DER)}$$
$$+ \Pr(|y_a - y_r| \leq T_h) \cdot E \left[|y_o - y_a|^2 \right] \text{ (UER)}$$
$$+ \Pr(|y_a - y_r| \leq T_h) \cdot E \left[|y_o - y_a|^2 \right]. \text{ (NER)}$$
$$(C-5)$$

Note that there is no information loss in scenario (i) and from scenarios (ii) and (iii)

$$P_e = \Pr(y_a \neq y_o, |y_a - y_r| > T_h)$$
$$+ \Pr(y_a \neq y_o, |y_a - y_r| \leq T_h). \quad (C-6)$$

Thus, (C-5) becomes

$$\sigma_{\gamma_r}^2 = \Pr(y_a \neq y_o, |y_a - y_r| \leq T_h) \cdot E \left[|y_o - y_a|^2 \right]$$
$$+ P_e(y_a \neq y_o, |y_a - y_r| > T_h) \cdot E \left[|y_o - y_r|^2 \right]. \quad (C-7)$$

In most cases, the magnitude of VOS error is large and hence detected, i.e., $P_e \sim \Pr(y_a \neq y_o || y_a - y_r| > T_h)$. Thus

$$\sigma_{\gamma_r}^2 \geq P_e(K_{vos}) \sigma_{q_m}^2 \quad (C-8)$$

where $\sigma_{q_m}^2 = E \left[|y_o - y_r|^2 \right]$.
By using the auxiliary variable y_r, we have

$$E \left[(y_o - y_a)^2 \right] = E \left[(y_o - y_r + y_r - y_a)^2 \right]$$
$$\leq E \left[|y_o - y_r|^2 \right] + E \left[|y_r - y_a|^2 \right]$$
$$+ 2E[y_o - y_r] \cdot E[y_r - y_a]. \quad (C-9)$$

In addition, as $|y_a - y_r| < T_h$ from scenario (ii) (see (C-3)), (C-9) can be simplified to

$$E \left[|y_o - y_a|^2 \right] \leq \sigma_{q_m}^2 + T_h^2 + 2T_h \mu_{q_m}. \quad (C-10)$$

By substituting (C-10) into (C-7), we obtain

$$\sigma_{q,rpr}^2 \leq P_e \left(\sigma_{q_m}^2 + T_h^2 + 2T_h \mu_{q_m} \right) + P_e \sigma_{q_m}^2 \quad (C-11)$$
$$= P_e \left(2\sigma_{q_m}^2 + T_h^2 + 2T_h \mu_{q_m} \right). \quad (C-12)$$

REFERENCES

[1] B. Davari, R. H. Dennard, and G. G. Shahidi, "CMOS scaling for high-performance and low power—The next ten years," *Proc. IEEE*, vol. 83, pp. 595–606, Apr. 1995.
[2] A. P. Chandrakasan and R. W. Brodersen, "Minimizing power consumption in digital CMOS circuits," *Proc. IEEE*, vol. 83, pp. 498–523, Apr. 1995.
[3] R. Gonzalez, B. Gordon, and M. Horowitz, "Supply and threshold voltage scaling for low-power CMOS," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1210–1216, Aug. 1997.
[4] J. M. Rabaey, *Digital Integrated Circuits: A Design Perspective*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[5] K. L. Shepard and V. Narayanan, "Noise in deep submicron digital design," in *Proc. Int. Conf. CAD*, Nov. 1996, pp. 524–531.
[6] P. Larsson and C. Svensson, "Noise in digital dynamic CMOS circuits," *IEEE J. Solid-State Circuits*, vol. 29, pp. 655–662, Jun. 1994.
[7] G. E. Moore, "Cramming more components onto integrated circuits," *Proc. IEEE*, vol. 86, pp. 82–85, Jan. 1998.
[8] N. R. Shanbhag, "A mathematical basis for power-reduction in digital VLSI systems," *IEEE Trans. CAS Part II*, vol. 44, no. 11, pp. 935–951, Nov. 1997.
[9] R. Hedge and N. R. Shanbhag, "Energy-efficient signal processing via algorithmic noise-tolerance," in *Proc. Int. Symp. Low-Power Electronics and Design*, Aug. 1999, pp. 30–35.
[10] ——, "Soft digital signal processing," *IEEE Trans. on VLSI*, vol. 9, no. 6, pp. 813–823, Dec. 2001.
[11] ——, "A low-power digital filter IC via soft DSP," in *Proc. of CICC*, May 2001, pp. 309–312.
[12] L. Wang and N. R. Shanbhag, "Low-power filtering via adaptive error-cancellation," *IEEE Trans. Signal Processing*, vol. 51, pp. 575–583, Feb. 2003.

[13] G. Balamurugan and N. R. Shanbhag, "The twin-transistor noise-tolerant dynamic circuit technique," *IEEE J. Solid-State Circuits*, vol. 36, pp. 273–280, Feb. 2001.

[14] The 2001 International Technology Roadmap for Semiconductors . [Online]. Available: http://public.itrs.net/Files/2001ITRS/Home.htm.

[15] P. Hazucha and C. Svensson, "Impact of CMOS technology scaling on the atmospheric neutron soft error rate," *IEEE Trans. on Nuclear Science*, vol. 47, pp. 2586–2594, Dec. 2000.

[16] N. Shanbhag, K. Soumyanath, and S. Martin, "Reliable low-power design in the presence of deep submicron noise," in *Proc. of Intl. Symp. on Low-Power Electronics and Design*, 2000, pp. 295–302.

[17] P. Shivakumar *et al.*, "Modeling the effect of technology trend on the soft error rate of combinational logic," in *Proc. Int. Conf. Dependable Systems Networks*, 2002, pp. 389–398.

[18] C. Nicol *et al.*, "A low-power 128-tap digital adaptive equalizer for broadband modems," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1777–1789, Nov. 1997.

[19] P. Larsson and C. Nicol, "Self-adjusting bit-precision for low power digital filters," in *Symp. VLSI Circuits*, 1997, pp. 123–124.

[20] R. Amirtharajah, T. Xanthopoulos, and A. Chandrakasan, "Power scalable processing using distributed arithmetic," in *Proc. Int. Symp. Low-Power Electronics Design*, 1999, pp. 170–175.

[21] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Norwood, NJ: Prentice-Hall, 1989, pp. 637–641.

[22] J. M. Rabaey and M. Pedram, *Low Power Design Methodologies*. Norwell, MA: Kluwer, 1996, pp. 160–200.

[23] B. M. Baas, "A low-power, high-performance 1024-point FFT processor," *IEEE J. Solid-State Circuits*, vol. 34, pp. 380–387, Mar. 1999.

[24] S. Hong, S. Kim, M. C. Papaefthymiou, and W. E. Stark, "Power-complexity analysis of pipelined VLSI FFT architectures for low energy wireless communication applications," in *Proc. 42nd Midwest Symp. Circuits Systems*, 2000, pp. 313–316.

[25] N. Weste and D. J. Skellern, "VLSI for OFDM," *IEEE Commun. Mag.*, vol. 36, pp. 127–131, Oct. 1998.

[26] F. Najm, "A survey of power estimation techniques in VLSI circuits," *IEEE Trans. VLSI Syst.*, vol. 2, pp. 446–455, Dec. 1994.

[27] E. A. Lee and D. G. Messerschmitt, *Digital Communication*. Norwell, MA: Kluwer, 1994.

[28] C. R. Baugh and B. A. Wooley, "A two's complement parallel array multiplication algorithm," *IEEE Trans. Comput.*, vol. C-22, pp. 1045–1047, Dec. 1973.

Byonghyo Shim (S'96–M'97) received the B.S. and M.S. degrees in control and instrumentation engineering from Seoul National University, Seoul, Korea, in 1995 and 1997 respectively. He is currently pursuing the Ph.D. degree in electrical and computer engineering at University of Illinois at Urbana-Champaign.

From 1997 to 2000, he was a full time instructor at the electronics engineering in Korean AirForce Academy, Cheongju, Korea. His research interests include signal processing for communication, VLSI signal processing, and low-power communication transceiver design.

Srinivasa R. Sridhara (S'01) received the B.Tech. degree in electronics and electrical communications engineering from Indian Institute of Technology, Kharagpur in 1999. He is currently pursuing the Ph.D. degree in electrical engineering at University of Illinois at Urbana-Champaign.

From 1999 to 2000, he was with Synopsys, Bangalore, India. Between 2000 and 2003, he has interned at Lucent, Whippany, NJ and Intel, Hillsboro, OR. His research interests include VLSI design of low-power signal processing systems, on-chip bus coding, and high-speed interconnects. Mr. Sridhara received the Outstanding Student Designer award from Analog Devices Inc. in 2001.

Naresh R. Shanbhag (S'87–M'93–SM'98) received the B.Tech. degree from the Indian Institute of Technology, New Delhi, India, (1988), the M.S. degree from the Wright State University, Dayton, OH, in 1990, and the Ph.D. degree from the University of Minnesota, Minneapolis, in 1993, all in electrical engineering.

From July 1993 to August 1995, he worked at AT&T Bell Laboratories at Murray Hill, NJ, where he was responsible for the development of VLSI algorithms, architectures and implementation of broadband data communications transceivers. In particular, 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 broadband access chip-sets. Since August 1995, he has been with the Department of Electrical and Computer Engineering, and the Coordinated Science Laboratory where he is presently an Associate Professor and the Director of the Illinois Center for Integrated Microsystems, University of Illinois. At the University of Illinois, he founded the VLSI Information Processing Systems (ViPS) Group, whose charter is to explore issues related to low-power, high-performance, and reliable integrated circuit implementations of broadband communications and digital signal processing systems. He has published numerous journal articles/book chapters/conference publications in this area and holds three US patents. He is also a co-author of the research monograph *Pipelined Adaptive Digital Filters* (Norwell, MA: Kluwer, 1994).

Dr. Shanbhag received the 2001 IEEE Transactions on VLSI Systems Best Paper Award, the 1999 IEEE Leon K. Kirchmayer Best Paper Award, the 1999 Xerox Faculty Award, the National Science Foundation CAREER Award in 1996, and the 1994 Darlington best paper award from the IEEE Circuits and Systems society. From July 1997–2001, he was a Distinguished Lecturer for the IEEE Circuits and Systems Society. From 1997 to 1999, he served as an Associate Editor for the IEEE Transaction on Circuits and Systems: Part II. He is currently the Associate Editor for the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS. He was the Technical Program Chair of the 2002 IEEE Workshop on Signal Processing Systems.