Error-Resilient Motion Estimation Architecture

Girish Vishnu Varatkar and Naresh R. Shanbhag, *Fellow, IEEE*

***Abstract*—In this paper, we propose an energy-efficient motion estimation architecture. The proposed architecture employs the principle of error-resiliency to combat logic level timing errors that may arise in average-case designs in presence of process variations and/or due to overscaling of the supply voltage [voltage overscaling (VOS)] and thereby achieves power reduction. Error-resiliency is incorporated via algorithmic noise-tolerance (ANT). Referred to as input subsampled replica ANT (ISR-ANT), the proposed technique incorporates an input subsampled replica of the main sum-of-absolute-difference (MSAD) block for detecting and correcting errors in the MSAD block. Simulations show that the proposed technique can save up to 60% power over an optimal error-free system in a 130-nm CMOS technology. These power savings increase to 78% in a 45-nm predictive process technology. Performance of the ISR-ANT architecture in the presence of process variations indicates that average peak signal-to-noise ratio (PSNR) of the ISR-ANT architecture increases by up to 1.8 dB over that of the conventional architecture in 130-nm IBM process technology. Furthermore, the PSNR variation (σ/μ) is also reduced by 7× over that of the conventional architecture at the slow corner while achieving a power reduction of 33%.**

***Index Terms*—Algorithmic noise-tolerance, error-tolerant design, low power design, motion estimation, process variation-tolerance.**

I. Introduction

RAPID growth in demand for portable and wireless computing systems is driving the need for ultra low-power systems. Next generation wireless multimedia communications standards such as digital video broadcast (DVB),[1] fourth-generation (4G) mobile systems[2] need to provide services such as video transmission on handheld units. These units need to be energy-efficient while providing a high quality of service. The MPEG-4 encoder is the most computationally intensive block in a video processor. The motion estimation (ME) kernel consumes 66%–94% of the encoder computational complexity [1]. A typical motion estimation accelerator consists of a RAM for storing the search area and the current block, an address generation unit, a datapath consisting of a processor element and a control unit. The datapath power consumption is found to be 75% of the total ME power consumption for full search motion estimation algorithm and 60% of the total ME power consumption for the three-step search algorithm [2]. Therefore, low-power motion-estimation architectures and implementations are of great interest.

A. Previous Work

Low-power motion estimation is a well-studied subject [1]–[10]. Algorithmic low-power ME techniques focus on heuristics to reduce the number of macro-blocks processed per motion vector [1]. These include employing a simpler distance criterion [6], fixed or adaptive search area [7], adaptive pel decimation [8], temporal or spatial prediction [9], and feature matching [10]. Several VLSI architectures have been proposed for ME with various tradeoffs between gate-count, input/output (I/O) bandwidth and throughput [1], [11], [12], [14]. Scaling of supply $(V_{\rm dd})$ and threshold voltage (V_t) has been commonly employed to reduce the total power consumption in digital circuits [15], [16]. The benefits of conventional voltage scaling are limited by the $(V_{\rm dd}, V_t)$ combination at which the worst case critical path delay is equal to the clock period [16].

All the previously mentioned approaches assume error-free computation. This assumption has been justified in the past by designing for the worst-case scenario. However, worst-case designs tend to be power hungry and hard to design. The reason being nanometer process technologies suffer from non-idealities such as process variations, voltage, or temperature induced noise and soft errors [17], [18]. One source of process variations is the random fluctuations in the number of dopant atoms in the MOS channel [19] which affects the device threshold voltage V_t of the transistor. The usage of sub-wavelength lithography for patterning transistors results in width and gate-length variations. These device parameter variations impact the functionality of the circuit by generating variations in the transistor drive current and gate capacitance. This creates delay variations which result in uncertainty in the data arrival time at the registers or memory elements causing them to latch incorrect data leading to logic errors.

An example of a variation-driven logic error event is shown in Fig. 1(a) using HSPICE simulation of a latched full adder designed in an IBM 130-nm process technology. We can see that the circuit which operates correctly at the nominal process corner produces an erroneous output at the slow process corner. Previous schemes to *avoid* errors due to timing violations have relied on adaptive body biasing (ABB) to modulate the transistor threshold voltage V_t [20] and adaptive supply voltage (ASV) [21]. These techniques are applicable to all digital circuits within the voltage levels specified by the process technology. However, the effectiveness of ABB is known to decrease as the channel length shrinks while ASV requires accurate, power-hungry circuitry. Process variations cause 30%

Manuscript received February 2, 2007; revised June 14, 2007. Current version published September 19, 2008. This work was supported in part by the Giga scale System Research Center (GSRC), one of five research centers funded under the Focus Center Research Program (FCRP), by a Semiconductor Research Corporation Program, and by Texas Instruments, Inc.

The authors are with the Coordinated Science Laboratory/Electrical and Computer Engineering Department, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: varatkar@uiuc.edu; shanbhag@uiuc.edu).

Digital Object Identifier 10.1109/TVLSI.2008.2000675

[1][Online]. Available: http://www.dvb.org

[2][Online]. Available: http://www.4gmf.org

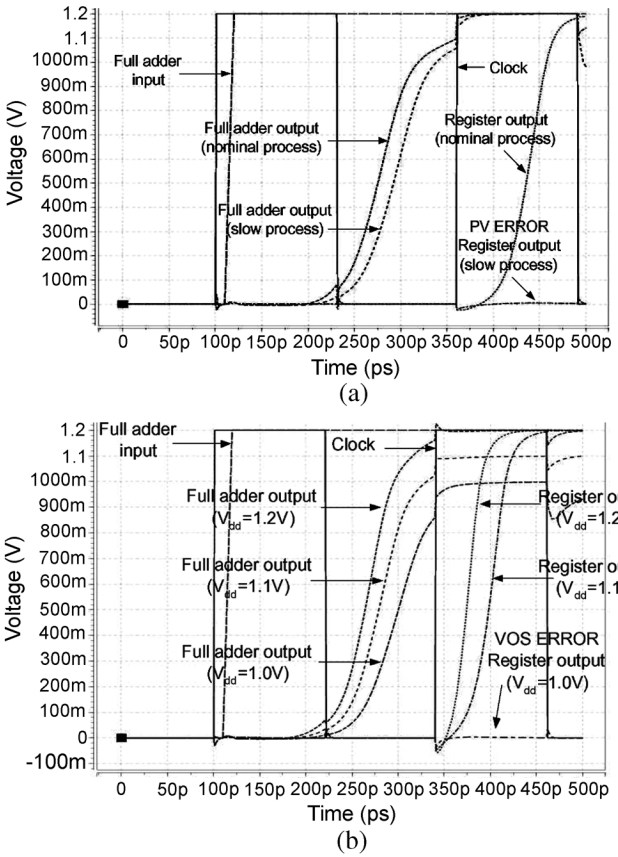


Fig. 1. Typical timing violation induced error due to: (a) process variations and (b) VOS.

variability in operating frequency in current process technology [22] and this variability is expected to increase to 60% within the next 10 years [23]. In the presence of such increased variations, a worst-case design has high power consumption while the nominal design, even though it is energy-efficient, will exhibit intermittent errors. Therefore, error-resilient architectures and implementations which tradeoff power with reliability are of great interest [24].

In the past, we have proposed the concept of voltage overscaling (VOS) [25] in order to push the limits of conventional voltage scaling for power savings. In VOS, the supply voltage is reduced beyond $V_{\text{dd-crit}}$, i.e.,

$$V_{\text{dd}} = K_{\text{vos}} V_{\text{dd-crit}} \tag{1}$$

where $0 < K_{\text{vos}} \leq 1$ and $V_{\text{dd-crit}}$ is the supply voltage below which timing violations occur. These violations are referred to as VOS errors. An example of such a VOS error is shown in Fig. 1(b). We can see that the circuits, which operates correctly at a nominal $V_{\text{dd}} = 1.2$ V, produces an erroneous output at $V_{\text{dd}} = 1.0$ V. With increased variations in process and voltage, researchers have proposed solutions at circuit [26], architectural [27], and system [28] level to *tolerate* nanometer non-idealities and tradeoff power with reliability. In the past, we have also proposed algorithmic techniques known as algorithmic noise-tolerance (ANT) to correct VOS errors efficiently [29]. In ANT (see Fig. 2), a main block is assumed to make intermittent errors due

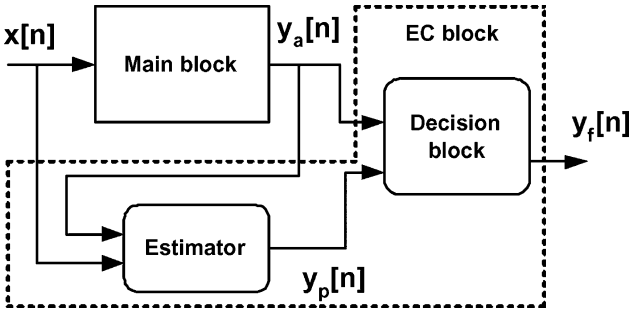


Fig. 2. ANT-based system.

to timing violations which are then corrected by an error-control block ($\mathbf{EC}$). The $\mathbf{EC}$ block includes an estimator and a decision block.

Many ANT techniques exist. These include the reduced precision replica (RPR) ANT [30] technique where the estimator is a reduced precision replica of the main block. In prediction-based ANT [25], the estimator is a predictor that exploits the correlation in the output of the main block. As VOS errors are input-dependent, the adaptive error-cancellation technique [31] employs an error estimator to estimate and cancel VOS errors at the main block output. Our past work has focused on ANT-based finite impulse response (FIR) filtering, where we have shown that the energy savings up to 7× can be achieved. ANT techniques are applicable to DSP/communications applications where the algorithmic performance is specified in terms of average metrics such as signal-to-noise ratio (SNR), peak-signal-to-noise ratio (PSNR), or bit error rate (BER). Logic and microarchitectural techniques such as RAZOR [27] correct VOS errors in microprocessors via detection and recovery mechanisms. Such techniques would be applicable to general purpose and data-centric computations where metrics such as BER and SNR may not apply.

B. Contribution

In this paper, we present a novel error-resilient low-power ME architecture that is based on the previously proposed concept of ANT. Preliminary results showing the benefits of the proposed architecture in correcting VOS errors were presented in [32]. In this paper, in addition to VOS, we study the effectiveness of the proposed architecture in correcting process variation-induced errors and determine achievable power savings. Simulations using an IBM 130-nm CMOS process show that up to 60% power savings can be achieved over an optimal error-free architecture. Power savings increase to 78% in a 45-nm predictive process technology[3] [35]. Thus, the combination of VOS and ANT can reduce power beyond that achievable by conventional voltage scaling alone. Simulations with a statistical process model of the 130-nm process technology show that the proposed technique increases the mean PSNR by up to 1.8 dB when compared to the PSNR of the conventional architecture on a slow die. It also reduces the PSNR variation (σ/μ) by 7× when compared to that of the conventional architecture in the presence of within-die (WID) variations at the slow process

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

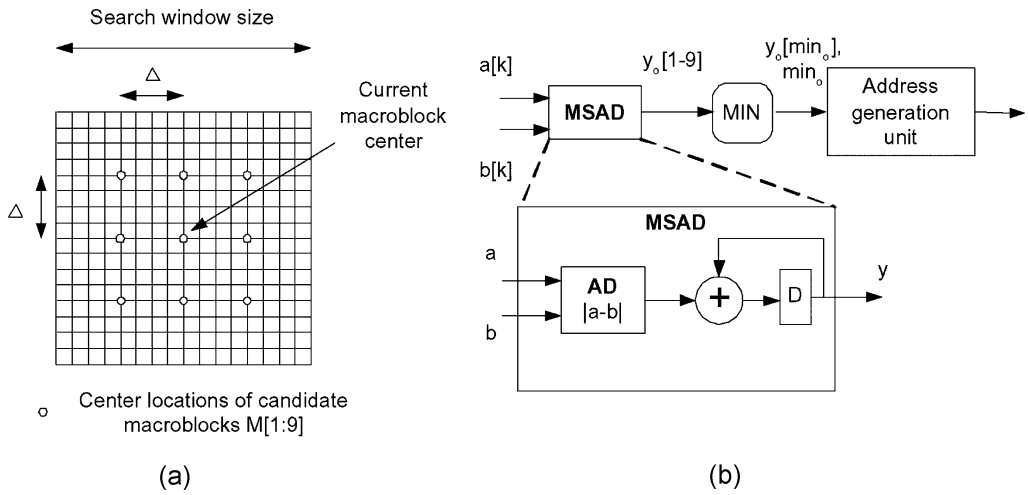


Fig. 3. TSS algorithm: (a) the search window and (b) a block level implementation.

corner. ISR-ANT architecture achieves up to 33% power savings for nearly equal values of PSNR when compared to the conventional architecture using adaptive supply voltage.

Section II describes the ME algorithm and a straightforward application of ANT to ME referred to as motion-vector replica ANT (MVR-ANT). In Section III, we present our main contribution: a new technique referred to as input sub-sampled replica ANT (ISR-ANT) for energy-efficient motion estimation. In Section IV, we present results in the presence of VOS for the ISR-ANT datapath designed using an IBM 130-nm process technology and a 45-nm predictive technology models. In Section V, we present simulation results showing the impact of process variations and combination of both process variations and VOS on the PSNR using the conventional and the ISR-ANT architectures.

II. Preliminaries

In this section, we present preliminaries of ME and ANT. We first introduce the ME algorithm, characterize the error probabilities for the arithmetic units employed in ME implementation and then demonstrate a straightforward but ineffective application of ANT resulting in the MVR-ANT. The latter will then be modified to form the proposed ISR-ANT in Section III.

A. Three-Step Search (TSS) Algorithm

An ME algorithm reduces temporal redundancy between consecutive video frames. In block matching ME algorithms, the current video frame is partitioned into non-overlapping macroblocks of size N pixels by N pixels. For each macroblock in the current frame, the ME algorithm efficiently searches for the best matching macroblock in the previous frame.

There are numerous algorithms for efficient search [1]–[14] since the ME algorithm is not standardized. For energy-efficiency purposes, we select an algorithm that is suitable for VLSI implementation. The full-search block matching (FSBM) algorithm is optimal in terms of visual quality and suitable for VLSI implementation due to its regularity. However, FSBM algorithm is computationally intensive.

The TSS algorithm [13] is a commonly employed sub-optimal block matching algorithm because of the simplicity of its implementation, robustness and near optimal performance. In this paper, we choose the TSS algorithm to demonstrate the effectiveness of the proposed ANT technique. Note that the proposed ANT technique can be applied to any other block matching algorithm.

In the TSS algorithm [see Fig. 3(a)], an initial step size Δ, typically equal to half of the search window size is chosen. Next, nine candidate macroblocks $\mathbf{M}[1:9]$ with their center locations as shown in Fig. 3(a), are chosen from the previous frame for comparison. Eight of these candidate macroblocks have their centers at a distance of $\pm\Delta$ in the x and y direction from the current macroblock. The ninth macroblock is at the same location as the current macroblock.

The sum of absolute differences (SAD) for each of the nine macroblocks are calculated by the main SAD ($\mathbf{MSAD}$) block [see Fig. 3(b)] by summing up the absolute difference between the corresponding pixels in the candidate macroblocks and the current macroblock. Thus, the $\mathbf{MSAD}$ block accepts input pixels $a_i[k]$ and $b_i[k]$ from the current and the candidate macroblocks, respectively. The output of the $\mathbf{MSAD}$ block are the nine candidate SAD values denoted by $y_o[i](1 \leq i \leq 9)$, where

$$y_o[i] = \sum_{k=1}^{N \times N} |a_i[k] - b_i[k]|, \quad \text{for } 1 \leq i \leq 9 \tag{2}$$

where we assume that a macroblock has $N \times N$ pixels. The index corresponding to the best match is obtained as

$$\begin{aligned} y_o[\min_o] &= \min\{y_o[1], y_o[2], \ldots, y_o[9]\} \\ \min_o &= \arg\min\{y_o[1], y_o[2], \ldots, y_o[9]\}. \end{aligned} \tag{3}$$

The motion vector is the vector difference between $\mathbf{M}[\min_o]$ and the current block. Next, Δ is halved and the center of the search window is moved to coincide with that of $\mathbf{M}[\min_o]$. Previous steps are repeated till the Δ becomes less than 1. In the block level implementation of TSS in Fig. 3(b), the $\mathbf{MSAD}$ block calculates the SAD in (2) while the $\mathbf{MIN}$ block determines $\min_o$ using (3).

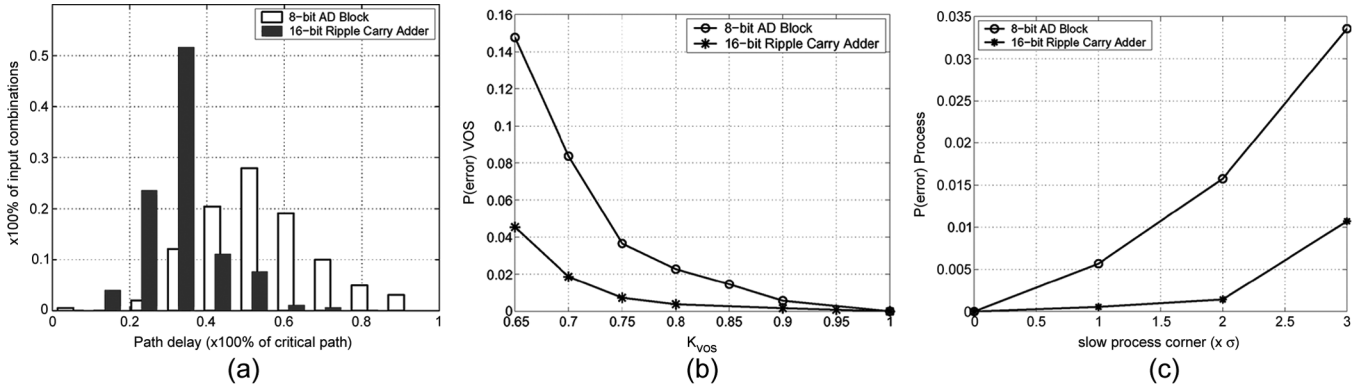


Fig. 4. Error characterization of an $\mathbf{AD}$ block and a ripple carry adder: (a) path delay distribution, (b) probability of VOS error, and (c) probability of process variation error.

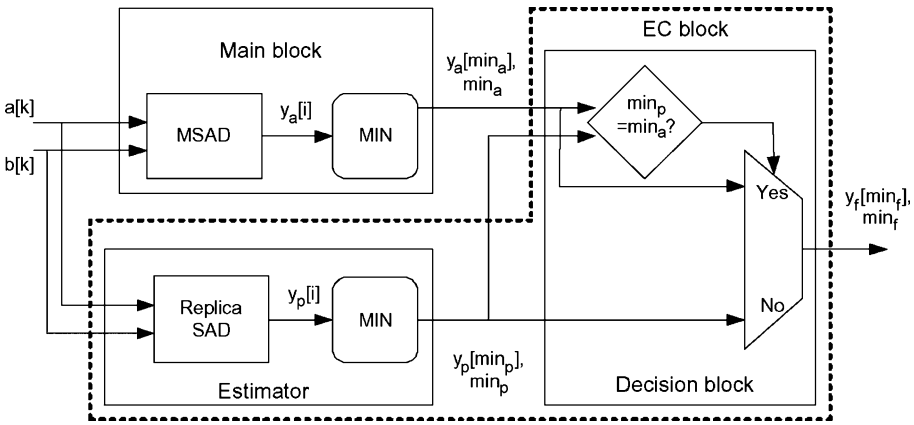


Fig. 5. MVR ANT architecture.

B. Error Characterization of Arithmetic Units

The $\mathbf{MSAD}$ block employs an absolute difference $(\mathbf{AD})$ block followed by an accumulator [see Fig. 3(b)]. These arithmetic units are based on least significant bit (LSB) first computation. Therefore, critical path timing violations due to VOS or process variation will result in errors in the most significant bits (MSBs). These errors are large in magnitude and hence severely degrade the performance in terms of PSNR.

The probability of timing errors depends on the path delay distribution of the architecture and the probability distribution of the inputs. The delay distributions of an 8-bit $\mathbf{AD}$ block and a 16-bit ripple carry adder are shown in Fig. 4(a). The probability of VOS errors for the $\mathbf{AD}$ block and the ripple-carry adder are shown in Fig. 4(b) for uniformly distributed inputs. We observe that the $\mathbf{AD}$ block and the ripple carry adder exhibit errors for 8% and 2% of the inputs, respectively, at $K_{\rm vos} = 0.7$. This is because the path delay distribution of a ripple carry adder has a longer tail than that of an $\mathbf{AD}$ block as shown in Fig. 4(a). The probability of process variations induced errors for the $\mathbf{AD}$ block and the ripple-carry adder are shown in Fig. 4(c) for uniformly distributed inputs at different process corners. The supply voltage is kept constant such that there are no errors at the nominal process corner. We observe that the $\mathbf{AD}$ block and the ripple carry adder exhibit errors for 3% and 1% of the inputs, respectively, at 3σ slow process corner. This is because the $\mathbf{AD}$ block has a greater number of paths with delays close to the critical path delay than the ripple carry adder as shown in Fig. 4(a). Therefore, the probability of error due to process variations is higher for the $\mathbf{AD}$ block than for the ripple carry adder. This motivates us to design arithmetic unit architectures whose path delay distribution has a long tail.

C. Motion Vector Replica (MVR) ANT

A straightforward application of the ANT framework (see Fig. 2) results in the MVR ANT shown in Fig. 5. An MVR ANT-based ME has a main block and an error control block $(\mathbf{EC})$. The address generation block is not shown. The main block is a complete ME engine that includes the $\mathbf{MSAD}$ and the $\mathbf{MIN}$ blocks. The main block is made energy-efficient via VOS or by designing at nominal process corner but makes intermittent errors. These timing violation errors degrade the output PSNR if left uncorrected. Let the error-free $\mathbf{MSAD}$ block outputs be denoted as $y_o[i]$ for $i = 1, \ldots, 9$. Under VOS or process variations error, the $\mathbf{MSAD}$ output denoted as $y_a[i]$ is given by

$$y_a[i] = y_o[i] + \eta[i] \tag{4}$$

where $\eta[i]$ is the error. Next, we define the main block output $\min_a$ as follows:

$$\begin{aligned} y_a[\min_a] &= \min\{y_a[1], y_a[2], \ldots, y_a[9]\} \\ \min_a &= \arg\min\{y_a[1], y_a[2], \ldots, y_a[9]\}. \end{aligned} \tag{5}$$

The **EC** block has an estimator and a decision block. The estimator estimates the correct motion vector and is designed to have a low complexity and hence an error-free operation. This means that the estimator output will not be as accurate as the error-free main block output. A simple **EC** block is one whose estimator is a reduced precision replica of the main block [30]. For example, the **MSAD** block can have 8-bit pixels as input while the replica **SAD** block can employ reduced input bit precision. If the main block and the estimator outputs differ, then the decision block employs the estimator output as the final corrected output $y_f[\min_f]$ as shown in Fig. 5. Thus, the final motion vector $\min_f = \min_a$ when $\min_a = \min_p$, otherwise $\min_f = \min_p$. As a result, $\min_f = \min_p$ for all cases, i.e., the PSNR of MVR ANT equals that of a low-complexity ME, which is what the estimator really is. As the probability of VOS error is very low (see Fig. 4) for $K_{\text{vos}} \geq 0.7$, the **MSAD** outputs are error-free most of the time. An error-free **MSAD** output is more accurate than the estimator output. Therefore, one should be able to conceive of a much better error correction scheme than the MVR ANT. The proposed ISR-ANT architecture presented in Section III represents such an improvement. We prove the superiority of ISR-ANT over MVR ANT in Appendix I.

III. Input Subsampled Replica (ISR) ANT

In this section, we describe the main contribution of this paper referred to as the ISR-ANT architecture.

A. ISR-ANT Architecture

We make the following modifications to the MVR ANT **EC** block of Fig. 5 to generate ISR-ANT as shown in Fig. 6.

1) We employ an estimator based on input sub-sampling, where an estimate of the **MSAD** output is calculated by employing an **ISR – SAD** block which sub-samples the input streams $a[k]$ and $b[k]$ by a factor of m as shown in the following:

$$y_p[i] = m \times \sum_{k=1}^{\lfloor N^2/m \rfloor} |a_i[mk] - b_i[mk]|. \tag{6}$$

Let $e_p[i]$ denote the SAD estimation error defined as follows:

$$e_p[i] = y_p[i] - y_o[i]. \tag{7}$$

Note that $e_p[i]$ and $y_o[n]$ is not required in implementing ISR-ANT as will be shown later. Also, note that the **ISR – SAD** block will consume lower power than the **MSAD** block and can be made to operate error-free because it can operate with a lower clock frequency and performs fewer computations.

2) We modify the decision block as follows. We detect and correct errors at the output of the **MSAD** block. Thus, error correction takes place before the **MIN** blocks instead of after as was the case with MVR ANT.

Note, the **ISR – SAD** output $y_p[i]$ is an estimate of the error-free sum $y_o[i]$ for $1 \leq i \leq 9$. Hence, a threshold T_h can be chosen in such a way that $\max(|e_p[i]|) < T_h$. This was done empirically by observing the probability density function (pdf) of estimation error e_p, for various frames. Let $\gamma[i]$ denote the difference between the actual (potentially erroneous) **MSAD** output $y_a[i]$ and **ISR – SAD** output $y_p[i]$, i.e.,

$$\gamma[i] = y_a[i] - y_p[i]. \tag{8}$$

An error is declared if $|\gamma[i]| > T_h$. The decision block employs the **ISR – SAD** output $y_p[i]$ as input to the **MIN** block if an error is detected. If there is no error, the **MSAD** output $y_a[i]$ is employed as input to the **MIN** block.

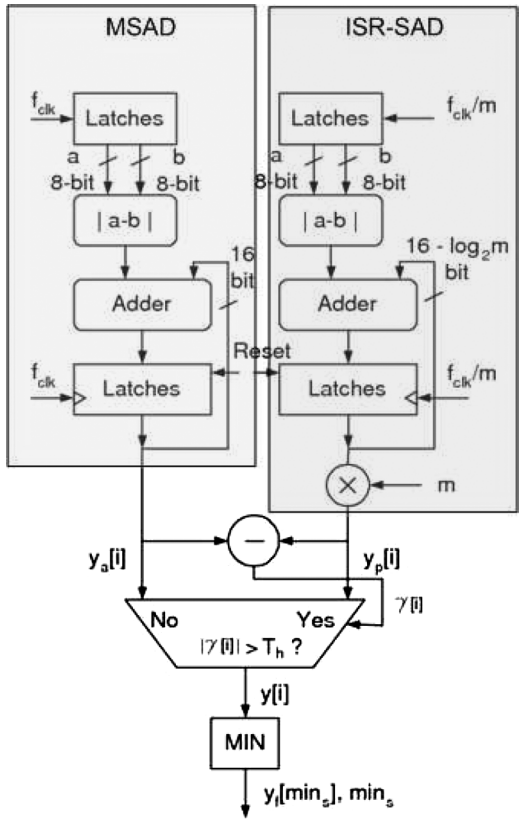


Fig. 6. Finite-precision ISR-ANT architecture.

The ISR-ANT algorithm is described as follows.

1) An initial step size Δ is chosen. Eight blocks at a distance of $\pm\Delta$ from the center (around the center block) are chosen for SAD computation and comparison. The **MSAD** and the **ISR – SAD** blocks calculate nine candidate SADs each.
2) Error detection: an error is declared if $|\gamma[i]| > T_h$, where $T_h = \max|e_p[i]|$.
3) Error correction: if an error is declared then $y[i] = y_p[i]$ else $y[i] = y_a[i]$.
4) The output is given by

$$y_f[\min_s] = \min\{y[1], y[2], \ldots, y[9]\}$$
$$\min_s = \arg\min\{y[1], y[2], \ldots, y[9]\}. \tag{9}$$

5) The step size Δ is halved. The center of the search window is moved to the center of $\mathbf{M}[\min_s]$. Go to Step 1). Repeat till the step size is less than 1.

ISR-ANT works well under the following assumptions.

1) The magnitude of error in **MSAD** block output is large. This makes it easy to detect errors.
2) The $\mathbf{ISR-SAD}$ and the decision blocks are error-free.

Both assumptions are easily met in practice. This is because the errors due to timing violations occur in the MSBs due to LSB first nature of computation in **MSAD**. As a result, the magnitude of the error in **MSAD** block output is large. The $\mathbf{ISR-SAD}$ block has only N/m inputs to process as compared to N inputs for the **MSAD** block. Hence, it is able to operate in an error-free manner.

The finite-precision ISR-ANT architecture is shown in Fig. 6, where the **MSAD** block consists of a modulus block computing the absolute difference between 8-bit luminance values. These absolute differences are accumulated by a 16-bit adder whose outputs are latched at a frequency $f_{\rm clk}$. The $\mathbf{ISR-SAD}$ block employs pixel luminance values which are subsampled by a factor of m. Hence, the adder in the $\mathbf{ISR-SAD}$ block has $16-\log_2 m$ bits and its outputs are latched at $f_{\rm clk}/m$. The absolute value of $\gamma[i]$ is compared with the threshold T_h after N^2 cycles of the clock with frequency $f_{\rm clk}$. The multiplexer output $y[i]$ is fed to the **MIN** block.

B. Power Savings via VOS

The total power consumption for the conventional architecture operating at $V_{\rm dd\text{-}crit}$ is given by

$$P_{\rm orig} = C_{\rm orig}V_{\rm dd\text{-}crit}^2\alpha_1 f_{\rm clk} + V_{\rm dd\text{-}crit}I_{l1} \tag{10}$$

where $C_{\rm orig}$ is the total capacitance of the **MSAD** block, $f_{\rm clk}$ is operating frequency, I_{l1} is the leakage current of the **MSAD** block, and α_1 is the activity factor. Assuming that the leakage power is typically G times the active switching power ($G \approx 0.4$ for current generation of microprocessors [34])

$$P_{\rm orig} = (1+G)C_{\rm orig}V_{\rm dd\text{-}crit}^2\alpha_1 f_{\rm clk}. \tag{11}$$

Similarly, the power dissipation of ISR-ANT is given by

$$\begin{aligned} P_{\rm ANT} =& C_{\rm orig}K_{\rm vos}^2V_{\rm dd\text{-}crit}^2\alpha_1 f_{\rm clk} + C_{\rm EC}V_{\rm dd\text{-}EC}^2\alpha_2\frac{f_{\rm clk}}{m} \\ &+ K_{\rm vos}V_{\rm dd\text{-}crit}I_{l2} + V_{\rm dd\text{-}EC}I_{l3} \\ =& (1+G)C_{\rm orig}K_{\rm vos}^2V_{\rm dd\text{-}crit}^2\alpha_1 f_{\rm clk} \\ &+ (1+G)C_{\rm EC}V_{\rm dd\text{-}EC}^2\alpha_2\frac{f_{\rm clk}}{m} \end{aligned} \tag{12}$$

where $C_{\rm EC}$ is the total capacitance of the $\mathbf{ISR-SAD}$ block, $V_{\rm dd\text{-}EC}$ is the $\mathbf{ISR-SAD}$ block supply voltage, m is the subsampling ratio, I_{l2} and I_{l3} are the **MSAD** and $\mathbf{ISR-SAD}$ block leakage currents, respectively, and α_2 is the activity factor of the $\mathbf{ISR-SAD}$ block. The **MIN** block and the multiplexer operate at a frequency of $f_{\rm clk}/N^2$ and hence their power consumption will be negligible compared to that of the **MSAD** block for large values of N, e.g., $N = 16$.

From (11) and (12), assuming that $\alpha_1 = \alpha_2 = \alpha$, it can be shown that $P_{\rm ANT} < P_{\rm orig}$ if

$$C_{\rm EC}V_{\rm dd\text{-}EC}^2/m < C_{\rm orig}V_{\rm dd\text{-}crit}^2\left(1-K_{\rm vos}^2\right). \tag{13}$$

The previous condition can be satisfied in practice by reducing $C_{\rm EC}$ and $K_{\rm vos}$, and increasing m. The power savings obtained by ISR-ANT can then be calculated as

$$\%\text{Power Savings} = \frac{P_{\rm orig}-P_{\rm ANT}}{P_{\rm orig}} \times 100. \tag{14}$$

In ISR-ANT ME architecture with $m = 4$, using (11), (12), and (14) and typical values of $C_{\rm EC} \approx 0.8C_{\rm orig}$, $K_{\rm vos} \approx 0.7$ and $V_{\rm dd\text{-}EC} \approx 0.5V_{\rm dd\text{-}crit}$, we find that power savings of 45% can be achieved.

IV. VOS Simulation Results

In this section, we compare the power versus performance tradeoffs between the conventional error-free architecture and ISR-ANT in an IBM 130-nm CMOS process and 45-nm predictive technology models,[3] [35]. Three different video clips are evaluated: flower garden (low motion), mobile calendar (medium motion), and football (high motion).

A. Simulation Set-Up

The system level throughput requirements for motion estimation in real time encoding of MPEG-II main profile at main level [33] is a CIF frame size of 288 by 352 pixels at the rate of 30 frames/s. We chose the macroblock size to be $N = 16$. We simulate the conventional and the ISR-ANT architectures using an HDL simulator which operates at the granularity of a 1-bit full adder (FA). The HDL simulation enables us to determine the maximum $T_{\rm FA}$ (delay of a 1-bit FA), necessary to support the required system level throughput. The conventional architecture is found to require a $T_{\rm FA} \leq 150$ ps for error-free operation.

We simulate the conventional TSS and ISR-ANT-based TSS architectures using the HDL simulator for various values of $T_{\rm FA}$ in order to determine the output motion vectors for three different clips. We predict the current frame from these motion vectors and the previous frame. The prediction error is calculated as the difference between the predicted frame and the actual current frame luminance values. The PSNR is calculated as

$$\text{PSNR(dB)} = 20 \times \log_{10}\frac{255}{\sigma_r} \tag{15}$$

where σ_r^2 is the prediction noise power. We set the desired PSNR requirement to be 0.5 dB less than the PSNR of the error-free conventional architecture.

For the conventional architecture, we observe that as $T_{\rm FA}$ is increased beyond 150 ps, VOS errors degrade the PSNR drastically. For the ISR-ANT architecture with $m = 4$, we determine that $T_{\rm FA}$ needs to be less than 225 ps in order for the **EC** block to correct the resulting errors effectively and maintain the system PSNR requirement. Similarly for $m = 3$ and $m = 5$, we find the delay requirement to be $T_{\rm FA} \leq 225$ ps and $T_{\rm FA} \leq 180$ ps, respectively, for the required PSNR.

B. Delay, Power Versus Supply and Body Bias Voltage

Next, we characterize a full adder with the mirror structure [37] in terms of its delay versus supply voltage and body bias voltage. The adder outputs are loaded with identical mirror full adders to determine the worst case output delay. Isodelay curves

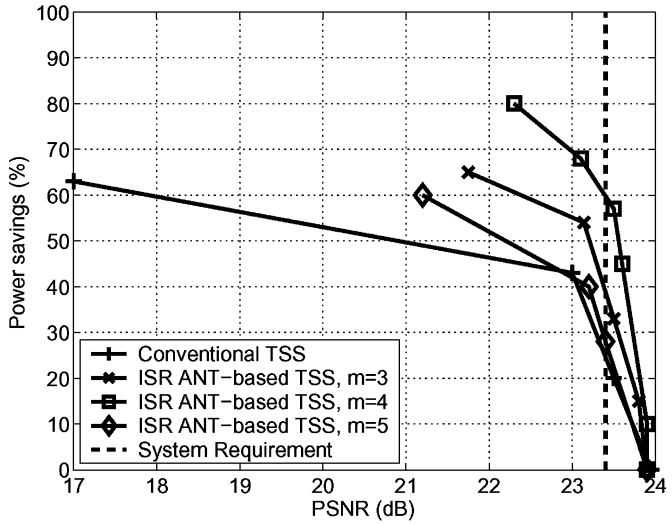


Fig. 8. Power savings versus PSNR plot for various sub-sampling ratios for mobile calendar clip in 130-nm process technology.

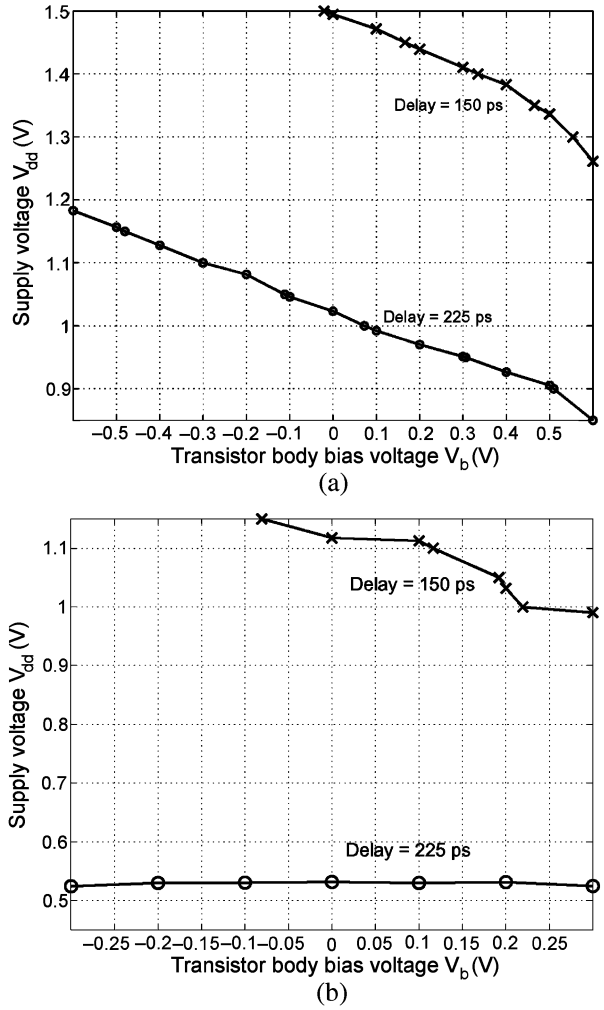


Fig. 7. Supply voltage and body bias for full adder delay = 150 and 225 ps for: (a) 130- and (b) 45-nm process technology.

are obtained by varying the supply and body bias voltage combinations $(V_{\rm dd}, V_b)$ via HSPICE in 130-nm IBM process technology and 45-nm predictive technology. Here, V_b represents the nMOS body voltage while the pMOS body voltage is $V_{\rm dd} - V_b$.

Fig. 7(a) shows the $(V_{\rm dd}, V_b)$ combinations that result in a constant full-adder delay $T_{\rm FA} = 150$ ps and $T_{\rm FA} = 225$ ps in 130-nm IBM technology. Similarly, Fig. 7(b) shows the isodelay plots in the 45-nm process technology. Similar isodelay curves are derived for intermediate delay values. We simulate the schematic netlists of conventional architecture and the ISR-ANT architecture with m =3, 4, and 5 using HSPICE for a few random input vectors to obtain the power consumption for both the architectures at the $(V_{\rm dd}, V_b)$ combinations obtained from the isodelay curves. We determine the power-optimum $(V_{\rm dd}, V_b)$ combination for the conventional and ISR-ANT architectures from these simulations.

In the 130-nm IBM process technology, the conventional architecture operates at $V_{\text{dd-crit}} = 1.35$ V, $V_b = 0.45$ V), while the ISR-ANT architecture operates at ($V_{\rm dd} = 0.95$ V, $V_b = 0.3$ V). The $\mathbf{EC}$ block operates at lower voltage, $V_{\text{dd-EC}} = 0.6$ V. In the 45-nm predictive process technology, the conventional architecture operates at ($V_{\text{dd-crit}} = 1.0$ V, $V_b = -0.3$ V), whereas ISR-ANT operates at ($V_{\rm dd} = 0.53$ V, $V_b = -0.3$ V). Here, the $\mathbf{EC}$ block operates at a voltage of $V_{\text{dd-EC}} = 0.4$ V.

C. Power Versus Performance Tradeoff

In this subsection, we compare the power savings obtained using ISR-ANT as a function of the PSNR. First, we vary the sub-sampling ratio m (see Fig. 8) in order to determine the value of m for which we obtain maximum power savings under the desired PSNR constraint for the mobile calendar clip. Note that the PSNR of the conventional architecture drops severely as the supply voltage is reduced from $V_{\text{dd-crit}}$. The ISR-ANT architecture is seen to be robust to VOS errors. As m increases from 3 to 4, the maximum achievable power savings increases from 41% to 60%. Note that the frequency of operation of the $\mathbf{ISR-SAD}$ block is $f_{\rm clk}/m$. As m increases from 3 to 4, the $\mathbf{ISR-SAD}$ block operates at lower frequency and lower voltage. As a result, the power consumption of the $\mathbf{ISR-SAD}$ block reduces significantly (from 20% to 6% of the error-free $\mathbf{MSAD}$ power) resulting in higher overall power savings. As we increase m from 4 to 5, the total power savings of the ISR-ANT decreases to 28%. This is because the estimation accuracy of the $\mathbf{ISR-SAD}$ output decreases significantly. Therefore, the minimum value of $K_{\rm vos}$ satisfying the system requirement increases resulting in reduced power savings. In the following subsection, results are presented and discussed for $m = 4$.

Next, we show a plot of power savings versus PSNR for the three different clips in Figs. 9 and 10 corresponding to 130- and 45-nm process nodes, respectively. All the clips are of CIF frame size of 288 by 352 pixels at the rate of 30 frames/s. The flower garden clip has slowly moving garden background. The mobile calendar clip shows objects moving with medium speed. The football clip has fast player movement. Plots are obtained by determining the power optimal $(V_{\rm dd}, V_b)$ combination from the isodelay curves similar to those in Fig. 7. From the plots, we note that the power savings are in the range of 41% to 60% when compared at the PSNR given by the system requirement. We note that as technology scales from 130 to 45 nm, the power

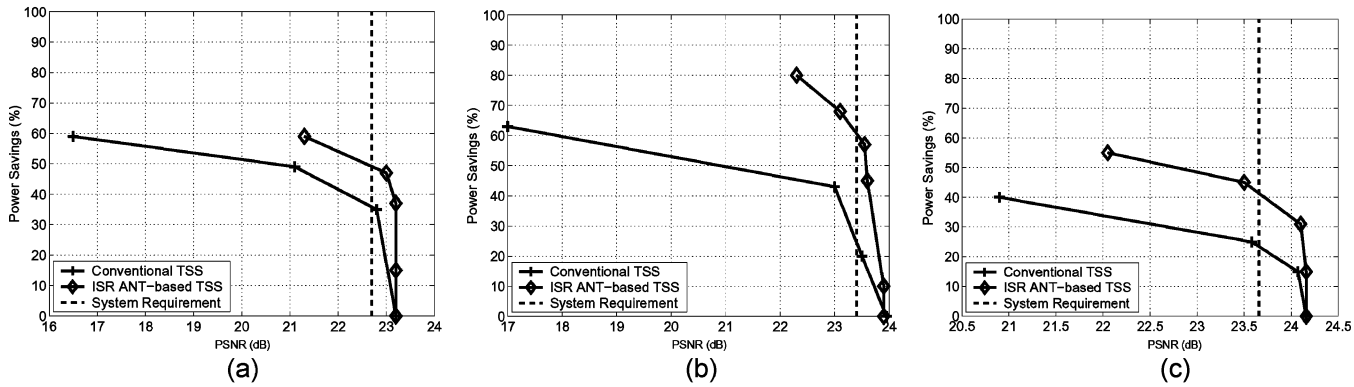


Fig. 9. Power savings versus PSNR plot in a 130-nm process technology for: (a) flower garden, (b) mobile calendar, and (c) football clip.

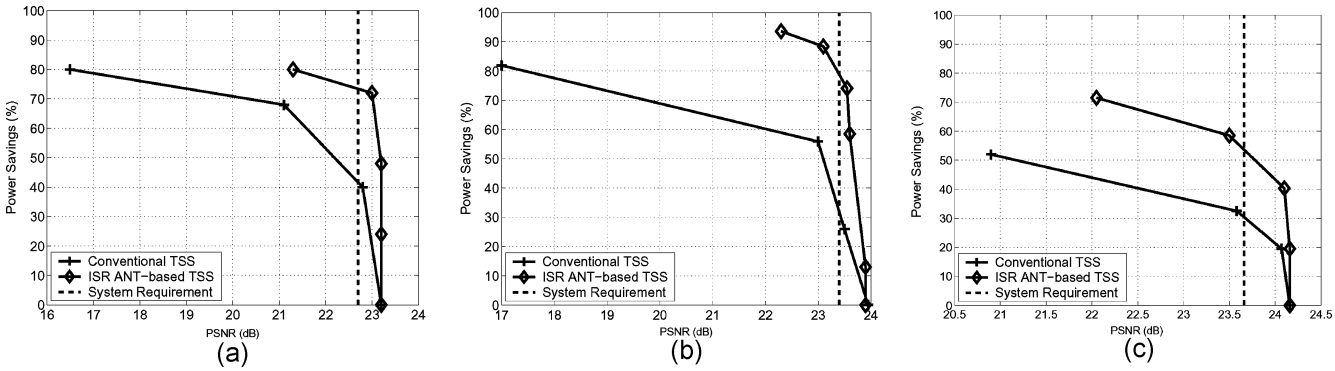


Fig. 10. Power savings versus PSNR plot in a 45-nm process technology for: (a) flower garden, (b) mobile calendar, and (c) football clip.

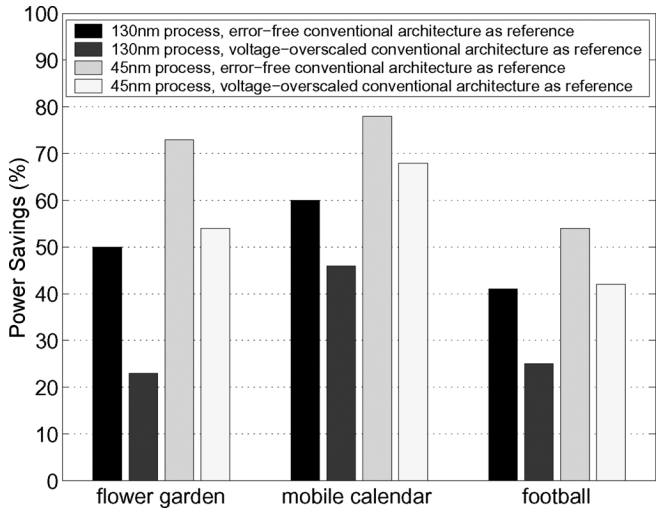


Fig. 11. Maximum power savings using ISR-ANT architecture over conventional architecture for flower garden ($\mathrm{PSNR} = 22.6$ dB), mobile calendar ($\mathrm{PSNR} = 23.5$ dB), and football ($\mathrm{PSNR} = 23.7$ dB) clips.

savings increase to the range of 54% to 78%. This is because the relative increase in delay due to voltage reduction is lower for the 45-nm process technology than for the 130-nm process due to reduced velocity saturation index with technology scaling.

The maximum achievable power savings obtained using ISR-ANT architecture for each of the clips is shown in Fig. 11. If we consider the power consumed by the conventional architecture operating under error-free conditions as reference, we can achieve 41% to 60% power savings in the 130-nm process technology. The maximum power savings increase to the range of 54% to 78% in the 45-nm predictive process technology. If we consider the power consumed by an overdesigned but voltage-overscaled conventional architecture with a PSNR value equal to the system requirement as a reference, the power savings are in the range of 23% to 46% in the 130-nm process technology and 42% to 68% in the 45-nm process technology. The reduction in power savings is because of the voltage overscaling of the conventional architecture that exploits the 0.5-dB slack in PSNR.

We note that if we use MVR ANT with $\mathbf{ISR-SAD}$ block as the estimator, then the PSNR of the mobile calendar clip degrades from PSNR $= 23.9$ dB to PSNR$= 23.23$, 22.8, 22.1 dB for $m = 2$, 4, 6, respectively. Similar degradation in performance is found in flower garden and football clips. This PSNR loss is unacceptable and it indicates that ISR-ANT is indeed a unique approach to power reduction.

Finally, in order to compare the area overhead of ISR-ANT architecture over the conventional architecture, we synthesized the layouts for both in the 130-nm process technology. We designed both the datapaths in Verilog and synthesized the layouts via *Synopsys Design Analyzer* and *Cadence Silicon Ensemble*. The conventional datapath required 250 standard cells while the ISR-ANT datapath needed 432 standard cells. Fig. 12 shows the layouts for the conventional and the proposed ISR-ANT system. We define the area overhead ρ of ISR-ANT as

$$\rho = \left(\frac{A_{\text{ISR-ANT}}}{A_{\text{Conventional}}} - 1 \right) \times 100\%. \tag{16}$$

Substituting $A_{\text{Conventional}} = (80.4\ \mu\text{m})^2$ and $A_{\text{ISR-ANT}} = (90.2\ \mu\text{m})^2$ in (16), we obtain the area overhead $\rho \approx 26\%$.

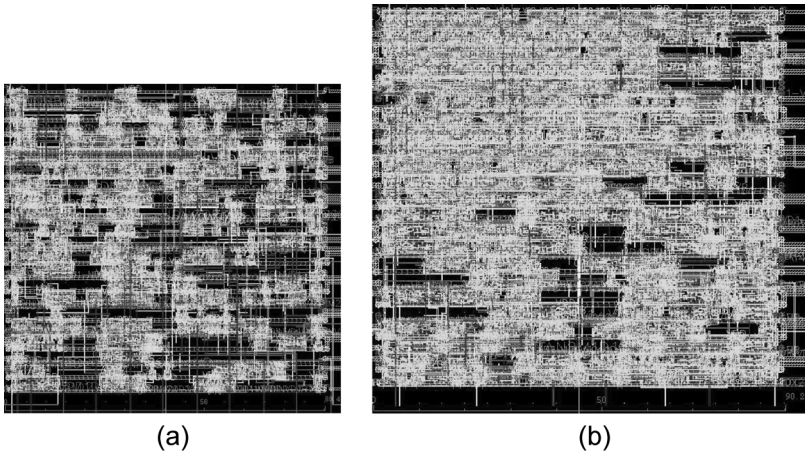


Fig. 12. Synthesized layouts of: (a) conventional datapath and (b) ISR-ANT datapath.

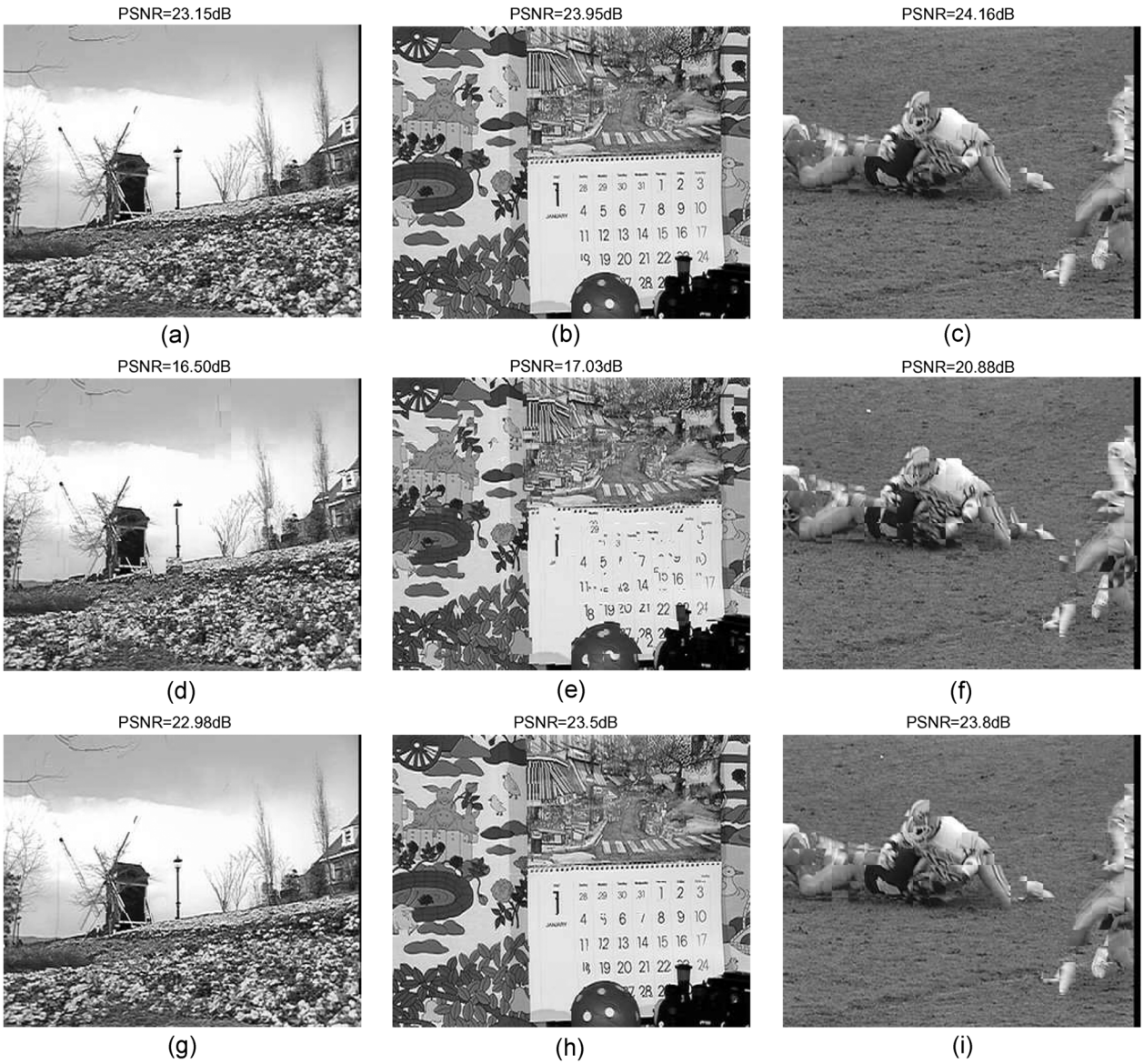


Fig. 13. Visual quality of the frames of three different clips predicted using: (a)–(c) error-free conventional architecture; (d)–(f) voltage-overscaled conventional architecture; and (g)–(i) voltage-overscaled ISR-ANT architecture.

The visual image quality of the frames predicted from the motion vectors are shown in Fig. 13. The first row of frames [see Fig. 13(a)–(c)] are obtained using an error-free conventional architecture which has a high power consumption. The second row

TABLE I
CHARACTERISTICS OF DELAY DISTRIBUTIONS OF VARIOUS GATES AT THE $3\sigma_g$ SLOW CORNER DUE TO WID VARIATIONS

Gate	V_b	-0.4V				0 V				0.4V			
	V_{dd}	1.35 V	1.2 V	1.05 V	0.9 V	1.35 V	1.2 V	1.05 V	0.9 V	1.35 V	1.2 V	1.05 V	0.9 V
Inverter	μ_l	39.23	43.19	49.48	61.33	36.18	39.04	43.43	49.59	34.19	36.33	39.44	43.40
	σ_l	5.4	6.0	7.3	10.7	4.9	5.4	6.1	8.4	4.7	5.1	5.7	6.8
	σ_l/μ_l	0.13	0.14	0.15	0.17	0.13	0.14	0.14	0.17	0.14	0.14	0.15	0.16
Exor	μ_l	136.4	162.6	204.3	276.8	122.4	142.9	172.7	223.1	110.8	126.8	150.0	185.21
	σ_l	15.21	18.95	25.87	38.48	13.61	16.30	21.32	29.66	11.80	14.02	17.37	22.92
	σ_l/μ_l	0.11	0.12	0.13	0.14	0.11	0.11	0.12	0.13	0.11	0.11	0.12	0.12
Full adder Carry	μ_l	215.6	253.2	310.9	422.3	199.7	223.4	265.5	337.7	186.1	207.2	235.9	287.8
	σ_l	17.06	22.97	29.47	47.37	16.27	18.69	25.02	34.04	14.20	16.81	20.05	27.90
	σ_l/μ_l	0.08	0.09	0.09	0.11	0.08	0.08	0.09	0.10	0.08	0.08	0.09	0.10
Full adder Sum	μ_l	305.1	352.9	485.8	805.9	288.2	320.2	411.5	640.5	262.5	292.9	363.6	548.8
	σ_l	19.4	22.7	35.2	50.8	18.03	20.03	30.07	41.45	16.8	18.6	26.9	35.6
	σ_l/μ_l	0.06	0.06	0.07	0.07	0.06	0.06	0.07	0.07	0.06	0.06	0.07	0.07

of frames [see Fig. 13(d)–(f)] correspond to the voltage-overscaled conventional architecture which has lower power consumption but degraded performance. We observe that the distortion in the images is very significant which is reflected in the low PSNR of the predicted frames. The third row of frames [see Fig. 13(g)–(i)] correspond to ISR-ANT architecture which has low power consumption comparable to the voltage overscaled conventional architecture but the image quality is comparable to the error-free output. The PSNR of the predicted frames can be seen to be within the system requirement.

V. SIMULATION RESULTS: PROCESS VARIATIONS

In this section, we first discuss the impact of process variations on circuit delay using a statistical process model. Next, we present simulation results showing the impact of delay variations on the PSNR of ME using conventional and ISR-ANT architectures.

Process variations are classified as die-to-die (D2D) and WID variations. D2D variations are caused by differences in process conditions (resist thickness, aberrations in the stepper lens and others) experienced by chips on different wafers in different lots. They modify the device properties (V_t, oxide thickness, conductance, and others) for all the devices on the chip in the same way. The standard deviation of the gate delay due to D2D process variations is denoted as σ_g. WID variations result in differences in device parameters for two instances of the same device on the same chip. WID variations are caused by geometric variation due to different layout conditions (nested versus isolated, vertical versus horizontal) and mismatch due to the placement of dopant atoms in the device channel. The mean and the standard deviation of the gate delay due to WID process variations are denoted as μ_l and σ_l, respectively. The impact of process variations are captured through measurements [36], which are then employed to generate statistical process models [39].

A. Simulation Setup

We characterized the delay distribution of basic gates such as an inverter, exor, and a full adder due to WID variations in a 130-nm, 1.2-V, IBM process. Monte Carlo simulations using statistical model files were employed for this purpose. Fig. 14 shows the delay distributions resulting from the presence of WID variations at the $3\sigma_g$ slow corner with $V_{\rm dd} = 1.2$ V, $V_b = 0$ V. Table I shows the impact of adaptive body bias (ABB) and

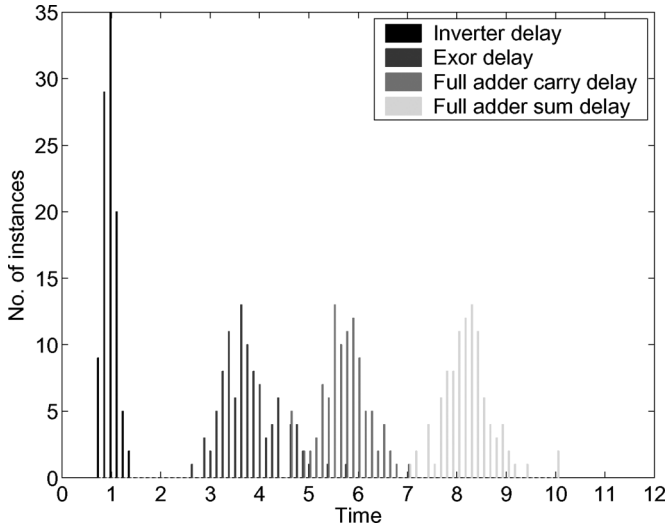


Fig. 14. Normalized delay distributions of various gates for a $3\sigma_g$ slow die with WID variations to the mean inverter delay.

TABLE II
CHARACTERISTICS OF PSNR DISTRIBUTIONS OF THE THREE CLIPS FOR CONVENTIONAL ARCHITECTURE ON A $3\sigma_g$ SLOW DIE DUE TO WID VARIATIONS

Clip	flower garden	mobile calendar	football
μ_c (dB)	21.31	21.95	23.17
σ_c (dB)	0.49	0.59	0.39
σ_c/μ_c	0.022	0.026	0.017

adaptive supply voltage (ASV) on the mean and the standard deviation of the delay for 3 representative values of nMOS body bias voltage V_b and four different supply voltages. From Table I, we observe that the relative delay variations (σ_l/μ_l) decreases as we move from the simplest gate (inverter) to a complex gate (full adder). The relative delay variations was also found to decrease with an increase in the supply voltage $V_{\rm dd}$.

Next, we sample the distribution in Fig. 14 to obtain the gate delays of a gate level implementation of the conventional and the ISR-ANT architectures at the $3\sigma_g$ slow process corner. This process is repeated 30 times in order to obtain 30 instances of the two architectures. We simulate the conventional and the ISR-ANT architectures using an HDL simulator which operates at the gate-level to determine the output motion vectors for the three clips. We predicted the current frame from these motion vectors and the previous frame to obtain the PSNR.

TABLE III
CHARACTERISTICS OF PSNR DISTRIBUTIONS OF THE THREE CLIPS FOR ISR-ANT ARCHITECTURE ON A $3\sigma_g$ SLOW DIE DUE TO WID VARIATIONS

		flower garden			mobile calendar			football		
		b=8	b=6	b=5	b=8	b=6	b=5	b=8	b=6	b=5
m=5	μ_i (dB)	22.84	22.75	21.75	22.29	22.06	21.85	22.92	22.90	22.76
	σ_i (dB)	0.27	0.28	0.67	0.24	0.25	0.51	0.21	0.21	0.31
	σ_i/μ_i	0.011	0.012	0.030	0.010	0.011	0.023	0.009	0.009	0.014
m=4	μ_i (dB)	23.12	22.80	22.19	23.42	23.33	22.07	23.58	23.48	23.19
	σ_i (dB)	0.08	0.14	0.36	0.11	0.19	0.49	0.26	0.29	0.38
	σ_i/μ_i	0.003	0.006	0.016	0.005	0.008	0.022	0.011	0.012	0.016
m=3	μ_i (dB)	23.18	23.10	22.29	23.70	23.51	23.05	24.14	24.08	23.51
	σ_i (dB)	0.01	0.16	0.4	0.02	0.2	0.3	0.01	0.04	0.14
	σ_i/μ_i	0.0004	0.007	0.018	0.0008	0.0085	0.013	0.0004	0.0017	0.006

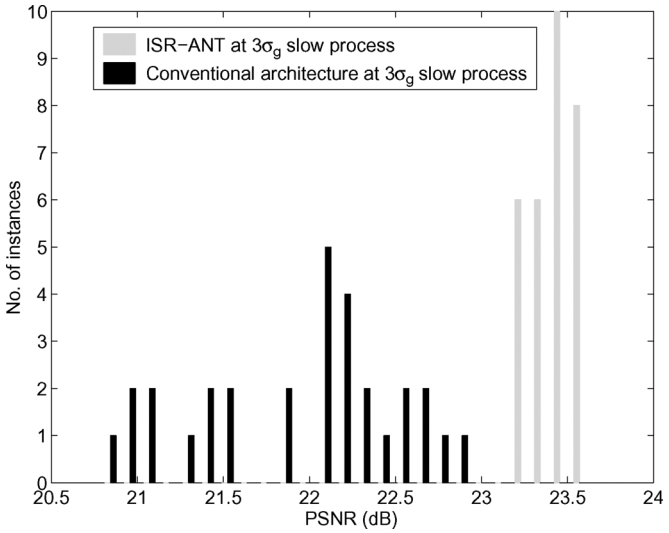


Fig. 15. PSNR distribution using conventional and ISR-ANT architecture on a slow die for mobile calendar clip.

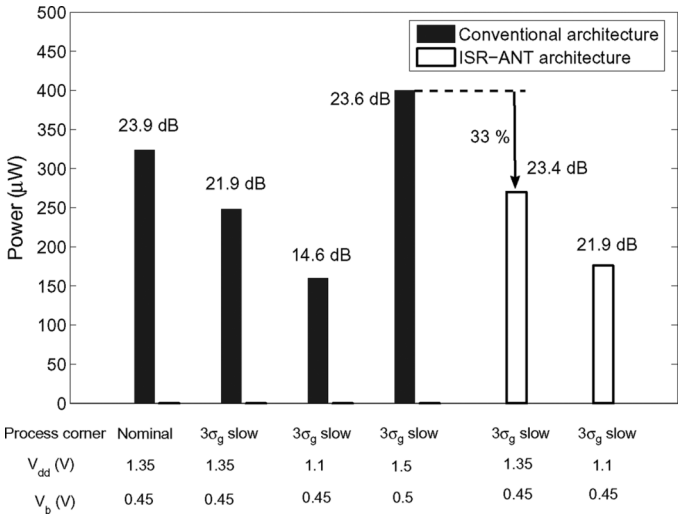


Fig. 16. Power performance tradeoff for mobile calendar clip.

B. Impact of Process Variations on PSNR

Each of the 30 instances of either the conventional or ISR-ANT architecture will result in a different PSNR. This is because the path delay distribution and hence the timing violations will be different for each instance. Thus, the PSNR is a random variable and hence will have a distribution. The means μ_c and the standard deviations σ_c of the PSNR for the conventional architecture for three different clips are tabulated in Table II. We observe that the mean PSNR drops by approximately 2 dB for flower garden and mobile calendar clips and 1 dB for the football clip. This drop is quite significant and will result in a noticeable loss in image quality.

Next, we obtain the $PSNR$ distribution for the ISR-ANT architecture for the subsampling ratios m = 3,4,5 and the $\mathbf{ISR-SAD}$ input precision b = 8, 6, 5. The representative distributions of the PSNR for the conventional architecture and the ISR-ANT architecture ($m = 4, b = 8$) are shown in Fig. 15. The mean μ_i and the standard deviation σ_i of the output PSNR are tabulated in Table III. From Table III, we can see that the improvement in the mean μ_i is significant as we increase estimator complexity from $m = 5$ to $m = 4$, but provides diminishing returns as the estimator complexity increases from $m = 4$ to $m = 3$. We also note that the performance of ISR-ANT decreases as the precision of ISR-SAD block is reduced from 8 to 5. Comparing Tables II and III, we observe that the mean PSNR increases but its standard deviation decreases when we use ISR-ANT architecture instead of the conventional architecture. The relative variation (σ_i/μ_i) in PSNR is reduced by 7× for the flower garden clip, 5× for the mobile calendar clip, and 4× for the football clip. Since we want to limit the PSNR loss to 0.5 dB, we choose $m = 4, b = 8$ in the following discussion.

C. Power Versus Performance Tradeoff

In this subsection, we present the power overhead of using ISR-ANT and its impact on the PSNR for a representative clip (mobile calendar). We compare power consumption of the ISR-ANT architecture with the conventional architecture. We simulate the transistor level netlist of the conventional architecture and the ISR-ANT architecture using HSPICE with a few random input vectors to obtain the power consumption for both the architectures at different supply voltage levels. We evaluate the mean PSNR employing the procedure described in Section V-B.

We show a plot of power consumption of the two architectures along with the mean PSNR for the mobile calendar clip in Fig. 16. The first bar shows the power consumed by the conventional architecture operating under error-free conditions on a nominal process die. The supply voltage and MOS body bias voltage are adjusted using the mean delay characterization results. We note that the prediction PSNR is 23.9 dB for power consumption of 324 μW at $V_{\rm dd}$ = 1.35 V, V_b = 0.45 V at the nominal process corner. The mean performance of the conventional architecture decreases to 21.9 dB for power consumption of 248 μW at $V_{\rm dd}$ = 1.35 V, V_b = 0.45 V at the $3\sigma_g$ slow

corner. If the supply voltage is further reduced to $V_{\rm dd} = 1.1$ V, $V_b = 0.45$ V, the errors occur from process variations, as well as VOS. As a result, the mean performance degrades to 14.6 dB with a power consumption of 160 μW. If we apply the conventional ABB and ASV to reduce the gate delays and correct the timing errors then the power consumption increases to 400 μW at $V_{\rm dd} = 1.5$ V, $V_b = 0.5$ V while achieving a mean PSNR of 23.6 dB. The ISR-ANT architecture, at $3\sigma_g$ slow process corner and $V_{\rm dd} = 1.35$ V, $V_b = 0.45$ V, has a power consumption of 270 μW with a PSNR of 23.4 dB. Thus, at the same slow process corner, the PSNR of ISR-ANT is comparable to the conventional architecture while consuming 33% lower power than the conventional architecture. When process variations and VOS occur simultaneously, the ISR-ANT architecture improves the PSNR from 14.6 to 21.9 dB while consuming an additional 10% power. Thus, the ISR-ANT technique is able to tradeoff power and performance very effectively and its PSNR is robust to process variations.

VI. Conclusion

In this paper, we have proposed a low-power ME architecture based on the principle of error-resilience. Referred to as the ISR-ANT, the proposed architecture is agnostic to the actual source of errors as it operates at the algorithmic level. The work presented in this paper falls in the category of communication-inspired low-power design techniques [25], [30]–[32] that favors the notion of error-correction rather than error-avoidance. Power hungry 4G media communication kernels such as the discrete cosine transform (DCT), the fast Fourier transform (FFT), and forward error-control (FEC) decoders are all promising candidates for the development of such error-resiliency-based low-power architectures. Studying the effectiveness of these techniques in correcting errors induced by composite sources of non-idealities such as VOS, process variations and particle hits is also of great interest.

Our work has not considered the impact of VOS and process variations on SRAMs. However, recent work [40] has shown that SRAMs exhibit read and write access delay distributions in the presence of variations and voltage scaling because of random dopant fluctuations. By selecting appropriate supply voltages, [40] further shows that the intrinsic error-resiliency of video encoding and the Viterbi decoder algorithms can compensate for or hide these errors quite effectively. Future work can be directed towards the design of error-resilient media and communications systems where error-resiliency can be selectively applied to SRAMs and the datapath to reduce power.

Appendix I
Probability of Error Comparison of MVR ANT and ISR-ANT

We assume that the **SAD** values corresponding to two candidate motion vectors are compared to determine the motion vector with minimum **SAD**. We prove that VOS error detection and correction before the **MIN** block is more effective than correction after the **MIN** block by deriving the probability of error in each case. This result can be extended to the case of nine candidate motion vectors on the similar lines. We make the following assumptions.

- The error-free **SAD** values corresponding to the two candidate blocks are denoted as y_0 and y_1, respectively, and $y_0 < y_1$. This assumption is justified because there is no loss of generality.
- The **MSAD** block outputs corresponding to the two motion vectors are denoted as $y_{a0} = y_0 + \eta_0$ and $y_{a1} = y_1 + \eta_1$, where η_0 and η_1 are the VOS errors in the **MSAD** outputs. The magnitude of VOS error in **MSAD** block output is assumed to be large so that it is easy to detect them. This assumption is justified because the VOS errors occur in the most significant bits (MSBs) due to least-significant bit (LSB) first nature of computation in **MSAD**.
- The Replica SAD block outputs corresponding to the two motion vectors are denoted as $y_{p0} = y_0 + e_0$ and $y_{p1} = y_1 + e_1$, where e_0 and e_1 are the estimation errors. We assume that these estimation errors are independent identically distributed (i.i.d.) and each has a normal pdf denoted as $f_{N(\mu,\sigma)}(x) = (1/\sigma\sqrt{2\pi})e^{(-(x-\mu)^2/2\sigma^2)}$ with mean $\mu = 0$ and standard deviation of σ, and the cumulative distribution function (cdf) denoted as $F_{N(\mu,\sigma)}(x)$. This assumption is supported via simulations.
- We assume timing violation error η_i is independent of the estimation error e_i for $i = 0, 1$. This assumption is justified because timing violation occurs whenever **MSAD** input excites a path longer than the clock period. This depends upon the architecture of the arithmetic units. The estimation error occurs due to sub-sampling of the inputs. Thus, the sources of timing error and estimation error are independent of each other.

First, we determine the probability of error for the MVR ANT architecture shown in Fig. 5. In this architecture, the final motion vector $\min_f = \min_a$ when $\min_a = \min_p$, otherwise $\min_f = \min_p$. As a result, $\min_f = \min_p$ for all the cases. Therefore, the output of MVR ANT is the same as the output of the estimator. The inputs of the **MIN** block in the estimator are y_{p0} and y_{p1}. Note that we assumed e_0 and e_1, the estimation errors to be i.i.d. with normal pdf $f_{N(0,\sigma)}(x)$. Therefore, the random variable $(e_0 - e_1)$ has a normal pdf given by $f_{N(0,\sigma\sqrt{2})}(x)$ [38]. If $Q(x) = \int_x^\infty (1/\sqrt{2\pi})e^{(-x^2/2)}dx$, then

$$\begin{aligned} P_{\rm MVR}[\text{error}] &= P[y_{p0} > y_{p1}] = P[y_0 + e_0 > y_1 + e_1] \\ &= P[e_0 - e_1 > y_1 - y_0] = Q\left(\frac{y_1 - y_0}{\sqrt{2}\sigma}\right). \end{aligned} \quad (17)$$

Second, we consider the ISR ANT architecture as shown in Fig. 6. Let the inputs to the **MIN** block be denoted as y_{f0} and y_{f1}, where

$$y_{fi} = \begin{cases} y_{ai}, & \text{if } |y_{ai} - y_{pi}| \le T_h \\ y_{pi}, & \text{if } |y_{ai} - y_{pi}| > T_h \end{cases} \quad (18)$$

for $i = 0, 1$. Since T_h is chosen such that $\max(|e_i|) < T_h$, no false alarm can occur. We assumed that the VOS error is large in magnitude and hence undetectable errors are unlikely. Let ϵ_i denote the probability of VOS error detection and correction. Then there exist the following two possibilities:

- *No error:* $y_{fi} = y_{ai} = y_i$ with probability $(1 - \epsilon_i)$;
- *Error:* $y_{fi} = y_{pi} = y_i + e_i$ with probability ϵ_i.

Let $y_{fi} = y_i + n_i = y_i + \alpha_i e_i$ where α_i is a random variable with pdf of $f_{\alpha_i}(x)$ given by

$$f_{\alpha_i}(x) = (1-\epsilon_i)\delta(x) + \epsilon_i \delta(x-1) \tag{19}$$

where $\delta(x)$ is the Dirac delta function. Note that α_i and e_i are independent of each other because VOS error η_i is assumed to be independent of the estimation error e_i. Assuming $\epsilon_0 = \epsilon_1 = \epsilon$, the cdf, $F_{n_i}(x)$, and pdf, $f_{n_i}(x)$, of $n_i = \alpha_i e_i$ are given by

$$\begin{aligned} F_{n_i}(x) &= \begin{cases} (1-\epsilon) + \epsilon F_{N(0,\sigma)}(x), & \text{if } x \geq 0 \\ \epsilon F_{N(0,\sigma)}(x), & \text{if } x < 0 \end{cases} \\ f_{n_i}(x) &= (1-\epsilon)\delta(x) + \epsilon f_{N(0,\sigma)}.(x) \end{aligned} \tag{20}$$

Note that $f_{n_i}(x)$ is independent of i and it has even symmetry. Therefore, n_0 and n_1 are i.i.d. random variables, with pdf of $f_n(x) = f_{n_i}(x)$. As a result, the random variable $(n_0 - n_1)$ has a pdf given by the convolution of $f_n(x)$ with itself [38] which simplifies as follows:

$$\begin{aligned} f_{n_0-n_1}(x) &= \int_{-\infty}^{\infty} f_n(\tau) f_n(x-\tau) d\tau \\ &= \epsilon^2 f_{N(0,\sigma\sqrt{2})}(x) + 2\epsilon(1-\epsilon) f_{N(0,\sigma)}(x) \\ &\quad + (1-\epsilon)^2 \delta(x). \end{aligned} \tag{21}$$

Hence, we obtain the expression for probability of error for ISR ANT using (21) as follows:

$$\begin{aligned} P_{\text{ISR}}[\text{error}] &= P[y_{f0} > y_{f1}] \\ &= P[y_0 + n_0 > y_1 + n_1] = P[n_0 - n_1 > y_1 - y_0] \\ &= \epsilon^2 Q\left(\frac{y_1 - y_0}{\sigma\sqrt{2}}\right) + 2\epsilon(1-\epsilon) Q\left(\frac{y_1 - y_0}{\sigma}\right). \end{aligned} \tag{22}$$

Subtracting (22) from (17), we get

$$\begin{aligned} &P_{\text{MVR}}[\text{error}] - P_{\text{ISR}}[\text{error}] \\ &= (1-\epsilon^2) Q\left(\frac{y_1 - y_0}{\sigma\sqrt{2}}\right) - 2\epsilon(1-\epsilon) Q\left(\frac{y_1 - y_0}{\sigma}\right) \\ &= (1-\epsilon)\left[(1+\epsilon) Q\left(\frac{y_1 - y_0}{\sigma\sqrt{2}}\right) - 2\epsilon Q\left(\frac{y_1 - y_0}{\sigma}\right)\right] \\ &> (1-\epsilon)\left[(1+\epsilon) Q\left(\frac{y_1 - y_0}{\sigma}\right) - 2\epsilon Q\left(\frac{y_1 - y_0}{\sigma}\right)\right] \\ &= (1-\epsilon)^2 Q\left(\frac{y_1 - y_0}{\sigma}\right) > 0. \end{aligned} \tag{23}$$

Thus, we proved that the probability of error for the ISR-ANT architecture is always lower than the probability of error for the MVR ANT architecture.

REFERENCES

[1] P. Kuhn, *Algorithms, Complexity Analysis and VLSI Architectures for MPEG-4 Motion Estimation*. Boston, MA: Kluwer, 1999.

[2] R. S. Richmond, II and D. S. Ha, "A low-power motion estimation block for low bit-rate wireless video," in *Proc. ISLPED*, Aug. 2001, pp. 60–63.

[3] M. A. Elgamel, A. M. Shams, and M. A. Bayoumi, "A comparative analysis for low power motion estimation VLSI architectures," in *Proc. IEEE Workshop Signal Process.*, 2000, pp. 149–158.

[4] F. Dufaux and F. Moscheni, "Motion estimation techniques for digital TV: A review and a new contribution," *Proc. IEEE*, vol. 83, no. 6, pp. 858–876, Jun. 1995.

[5] P.-C. Tseng, Y. Chang, Y. Huang, H. Fang, C. Huang, and L. Chen, "Advances in hardware architectures for image and video coding- a survey," *Proc. IEEE*, vol. 93, no. 1, pp. 184–197, Jan. 2005.

[6] M. Ghanbari, "The cross-search algorithm for motion estimation," *IEEE Trans. Commun.*, vol. 38, no. 7, pp. 950–953, Jul. 1990.

[7] J. Minocha and N. R. Shanbhag, "A low power data-adaptive motion estimation algorithm," in *Proc. IEEE Workshop Multimedia Signal Process.*, Sep. 1999, pp. 685–690.

[8] Y. L. Chan and W. C. Siu, "New adaptive pixel decimation for block motion vector estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 1, pp. 113–118, Feb. 1996.

[9] B. Zeng, R. Li, and M. L. Liou, "Optimization of fast block motion estimation algorithms," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 6, pp. 833–844, Dec. 1997.

[10] K. Sauer and B. Schwartz, "Efficient motion estimation using integral projections," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 5, pp. 513–518, Dec. 1996.

[11] S. Kim, Y. Kim, K. Yim, H. Chung, K. Choi, Y. Kim, and G. Jung, "A fast motion estimator for real-time systems," *IEEE Trans. Consumer Electron.*, vol. 43, no. 1, pp. 24–33, Feb. 1997.

[12] S. Dutta and W. Wolf, "A flexible parallel architecture adapted to block matching motion estimation algorithms," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 1, pp. 74–86, Feb. 1996.

[13] T. Koga, "Motion compensated interframe coding for video conferencing," in *Proc. NTC*, 1981, pp. 1799–1808, Ch. 9.6.1–9.6.5.

[14] P. Lakamsani, "An architecture for enhanced three step search generalized for hierarchical motion estimation algorithms," *IEEE Trans. Consumer Electron.*, vol. 43, no. 2, pp. 221–227, May 1997.

[15] 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.

[16] 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.

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

[18] S. Borkar, T. Karnik, and V. De, "Design and reliability challenges in nanometer technologies," in *Proc. DAC*, Jun. 2004, p. 75.

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

[20] J. W. Tschanz, J. T. Kao, S. G. Narendra, D. A. Antoniadis, R. Nair, A. P. Chandrakasan, and V. De, "Adaptive body bias for reducing impact of die-to-die and within-die parameter variations on microprocessor frequency and leakage," *IEEE J. Solid-State Circuits*, vol. 37, no. 11, pp. 1396–1402, Nov. 2002.

[21] T. Chen and S. Naffziger, "Comparison of adaptive body bias (ABB) and adaptive supply voltage (ASV) for improving delay and leakage under the presence of process variation," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 11, no. 10, pp. 888–899, Oct. 2003.

[22] S. Borkar, "Designing reliable systems from unreliable components: The challenges of transistor variability and degradation," *IEEE MICRO*, vol. 25, no. 6, pp. 10–16, Dec. 2003.

[23] "International Technology Roadmap for Semiconductors (ITRS)," 2005 [Online]. Available: http://www.public.itrs.net

[24] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De, "Parameter variations and impact on circuits and microarchitecture," in *Proc. DAC*, Jun. 2003, pp. 1563–1568.

[25] R. Hegde and N. R. Shanbhag, "Soft digital signal processing," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 9, no. 6, pp. 813–823, Dec. 2001.

[26] M. Zhang and N. R. Shanbhag, "A CMOS design style for logic circuit hardening," in *Proc. IEEE Int. Reliab. Phys. Symp.*, 2005, pp. 223–229.

[27] D. Ernst and , "Razor: A low-power pipeline based on circuit-level timing speculation," in *Proc. IEEE Int. Symp. Microarch.*, Dec. 2003, pp. 7–18.

[28] N. Nakka, N. S. Kim, S. Das, S. Pant, R. Rao, T. Pham, C. Ziesler, D. Blaauw, T. Austin, K. Flautner, and T. Mudge, "An architectural framework for providing reliability and security support," in *Proc. Int. Conf. Dependable Syst. Netw.*, 2004, pp. 585–594.

[29] N. R. Shanbhag, K. Soumyanath, and S. Martin, "Reliable low-power design in presence of deep submicron noise," in *Proc. ISLPED*, Aug. 2000, pp. 295–302.

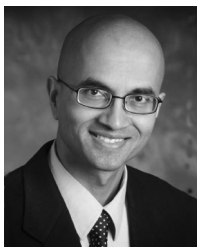
[30] B. Shim and N. R. Shanbhag, "Low-power digital signal processing via reduced-precision redundancy," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 12, no. 5, pp. 497–510, May 2004.

[31] L. Wang and N. R. Shanbhag, "Low-power filtering via adaptive error-cancellation," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 575–583, Feb. 2003.
[32] G. Varatkar and N. R. Shanbhag, "Energy-efficient motion estimation using error-tolerance," in *Proc. ISLPED*, Oct. 2006, pp. 113–118.
[33] S. C. Hsia, "VLSI implementation for low-complexity full-search Motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 7, pp. 613–619, Jul. 2002.
[34] N. S. Kim *et al.*, "Total power-optimal pipelining and parallel processing under process variations in nanometer technology," in *Proc. ICCAD*, Nov. 2005, pp. 535–540.
[35] Y. Cao, T. Sato, D. Sylvester, M. Orshansky, and C. Hu, "New paradigm of predictive MOSFET and interconnect modeling for early circuit design," in *Proc. CICC*, 2000, pp. 201–204.
[36] K. A. Bowman, S. G. Duvall, and J. D. Meindl, "Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration," *IEEE J. Solid-State Circuits*, vol. 37, no. 2, pp. 183–190, Feb. 2002.
[37] J. M. Rabaey, A. Chandrakasan, and B. Nikolic, *Digital Integrated Circuits: A Design Perspective*. Englewood Cliffs, NJ: Prentice-Hall, 2003.
[38] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.
[39] *"IBM Process Design Manual,"* IBM, May 2004.
[40] F. J. Kurdahi, A. M. Eltawil, Y. H. Park, R. N. Kanj, and S. R. Nassif, "System-level SRAM yield enhancement," in *Proc. ISQED*, 2006, pp. 179–184.

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 from the University of Illinois at Urbana-Champaign (UIUC), Urbana, in 2008.

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. 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 an Engineering Research Intern.

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



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

From July 1993 to August 1995, he worked with AT&T Bell Laboratories, 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 very high-speed digital subscriber line (VDSL) 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 a Professor. His research interests include the design of integrated circuits and systems for broadband communications including low-power/high-performance VLSI architectures for error-control coding, equalization, as well as digital integrated circuit design. He has published numerous journal articles/book chapters/conference publications in this area and holds five U.S. patents. He is also a coauthor of the research monograph *Pipelined Adaptive Digital Filters* (Kluwer, 1994).

Dr. Shanbhag was a recipient of the 2006 IEEE JOURNAL OF SOLID–STATE CIRCUITS Best Paper Award, the 2001 IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (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. In 1997, he was a Distinguished Lecturer for the IEEE Circuits and Systems Society. From 1997 to 1999 and from 1999–2002, he served as an Associate Editor for the IEEE TRANSACTION ON CIRCUITS AND SYSTEMS—PART II: BRIEF PAPERS and the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, respectively. He is currently serving on the technical program committees of major international conferences such as the IEEE International Solid-State Circuits Conference (ISSCC), the IEEE International Conference on Computer-Aided Design (ICCAD), the IEEE International Symposium on Low-Power Design (ISLPED), and the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). In 2000, he cofounded and served as the Chief Technology Officer of Intersymbol Communications, Inc., a fabless semiconductor start-up developing DSP-enhanced mixed-signal ICs for OC-192 optical links. In 2007, Intersymbol Communications, Inc., was acquired by Finisar Corporation, Inc., where he is presently serving as a Senior Scientist on a part-time basis.