ENERGY-EFFICIENT SOFT ERROR-TOLERANT DIGITAL SIGNAL PROCESSING

Byonghyo Shim, Naresh R. Shanbhag, and Seok-jun Lee

Coordinated Science Laboratory/
Department of Electrical and Computer Engineering,
University of Illinois at Urbana-Champaign,
Email: {bshim, shanbhag, slee6}@uiuc.edu

ABSTRACT

In this paper, we present energy-efficient soft error (SE)-tolerant techniques for digital signal processing (DSP) systems. The proposed technique, referred to as algorithmic soft error-tolerance (ASET), employs an low-complexity estimator of a main DSP block to guarantee reliability in presence of soft errors either in the MDSP or the estimator. For FIR filtering, it is shown that the proposed technique provides robustness to soft error rates of up to $P_{er} = 10^{-2}$ in single-event upset (SEU). It is also shown that the proposed techniques provide $40\% \sim 61\%$ savings in power dissipation over that achieved via triple modula redundancy (TMR) when the desired signal-to-noise ratio $SNR_{des} = 25 \sim 35$ dB.

1. INTRODUCTION

Rapid scaling of CMOS process technology has resulted in significant improvements in power consumption and throughput. However, reduced feature sizes and voltages make current and future systems vulnerable to *deep submicron* (DSM) noise [1] and *soft errors* due to alpha particles, high-energy neutrons, and thermal neutrons of cosmic radiation [2]-[4].

Soft errors in semiconductor memories are a known concern for many years. Error-control codes (ECC) are currently being employed to combat soft errors in memories. In the past, soft errors in logic and data-path circuits were not a problem because these get masked [6] through the following three mechanisms: 1) attenuation of noise pulses as it propagates through a logic chain (*electrical masking*), 2) logic values overiding the noisy input (*logical masking*), and 3) noise pulse missing the latch set-up and hold timing window (*latching-window masking*).

However, these masking mechanisms become less effective with reduction in feature size, increase in clock frequency, and reduction in pipelining depth of modern architectures. It has been shown that the critical charge which flips the internal state of logic gate approaches $4fC$ in $0.13\mu m$ technology [5]. Therefore, the soft error rate in logic circuits is expected to increase nine orders of magnitude from 1992 to 2011 [6] to the point where soft error rates in logic will equal that of unprotected memory structures. Indeed, the 2001 International Technology Roadmap for Semiconductors (ITRS) refers to error-tolerance as a design challenge for the next decade [7].

This research was supported in part by the Microelectronics Advanced Research Corporation (MARCO) sponsored Gigascale Silicon Research Center, NSF grants CCR 99-79381 and CCR 00-85929.

In this paper, we extend our past work on combating deep submicron (DSM) noise [11]-[14] referred to as algorithmic noise-tolerance (ANT). An ANT-based system has a noisy main DSP (MDSP) block and an error-control (EC) block that detects and corrects errors in the MDSP block. A key underlying assumption in our past work is that the EC block is error-free. The rationale for this assumption is based on the observation that EC blocks are much smaller than the MDSP and hence can be designed to be immune from DSM noise source as voltage/frequency scaling, process variation, ground bounce, etc.. Since this assumption is not valid when soft errors arise due to alpha particles and cosmic rays, we relax this assumption in this paper.

The rest of this paper is organized as follows. We introduce soft error model and TMR technique in Section II. Next, we provide the soft error-tolerant ANT (ASET) technique and its performance analysis in Section III. Simulation results and discussion are provided in Section IV.

2. SOFT ERROR

In this section, we first discuss the soft error model then review the triple modular redundancy (TMR) and algorithmic noise tolerance (ANT).

2.1. Soft Error Model

Soft errors are caused by alpha particles emitted by packaging materials and cosmic ray neutrons [3]. These high-energy particles ($10^5 \sim 10^{12}$ eV) traverse through the silicon substrate and generate electron-hole pairs which may get absorbed by drain nodes in the circuit resulting in a voltage noise pulse.

Since the particle emission is a true nuclear event, it has random characteristic and the events are independent to each other. As the technology scales, node upsetting charge Q_{crit} reduces proportionally while the number of node per chip increases and clock period/pipeline depth reduces, resulting in high soft error rate in system level.

The output $y_a[n]$ of the DSP system in the presence of soft errors can be written as

$$y_a[n] = y_o[n] + \gamma[n] = (d[n] + \eta[n]) + \gamma[n] \qquad (1)$$

where $y_o[n]$ denotes the error-free output composed of a desired signal $d[n]$ and channel noise $\eta[n]$, and $\gamma[n]$ is the soft error signal with non-zero probability.

Assuming 2's complement number representation, an $N+1$-bit representation of the original ouput $y_o[n]$ can be represented

0-7803-8104-1/03/$17.00 ©2003 IEEE

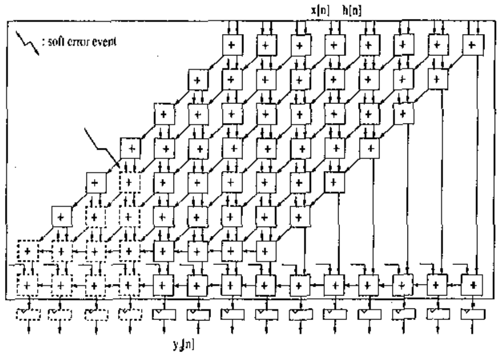

Fig. 1. Illustration of soft error event in a multiply-and-accumulate (MAC) block. The dotted blocks indicate the presence of a logic error.

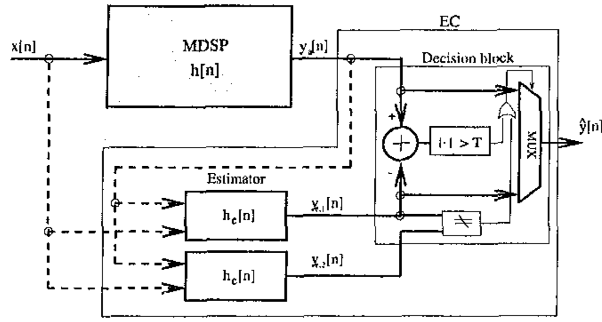

Fig. 2. The block diagram of ASET.

as

$$y_o[n] = -b_0 + \sum_{i=1}^{N} b_i 2^{-i}. \quad (2)$$

In addition, we define the flip variable f_i which indicates whether the ith output bit is flipped. If the ith bit is flipped by soft error, then $f_i = 1$ and $f_i = 0$ otherwise. The rationale for employing flip vector is that even though soft error can occur any node inside the MDSP, it is the latch that generates and propagates error. If the flip vector $\boldsymbol{f}$ at time unit n is given, i.e., $\boldsymbol{f} = (f_0, f_1, \cdots, f_N)$, then the actual output $y_a[n]$ is

$$y_a[n] = -(b_0 \oplus f_0) + \sum_{i=1}^{N} (b_i \oplus f_i) 2^{-i}, \quad (3)$$

where $\oplus$ is an XOR operation. If we assume that each f_i is i.i.d and equiprobable, and $y_o[n]$ is uniform, then the pdf of γ is shown to be

$$f_\Gamma(\gamma) \sim \frac{p_e}{4}(2 - |\gamma|) + (1 - p_e)\delta(\gamma), \ \gamma \in (-2, 2) \quad (4)$$

2.2. Error/Noise Tolerance

Redundancy based techniques are popular for combating soft errors in a computational units. Redundancy can be either spatial or temporal or both. Triple modular redundancy (TMR) [9], [15] is a redundancy based technique that includes three identical MDSPs and employs the majority voting for the error control. Since it is very unlikely for two identical units to be in error at the same time, TMR is very effective in removing soft error and conceptually easy to implement. However, it has high penalty on power and area.

In ANT technique, a low-complexity estimator $h_e[n]$ for error-free MDSP output $y_o[n]$ is employed for error control. In the decision block, the estimator output $y_e[n]$ is compared to $y_a[n]$. If the difference metric $d(y_a[n], y_e[n]) = |\, y_a[n] - y_e[n]\,|$ is greater than a prespecified threshold T_h, then $y_e[n]$ is used as a corrected output. Otherwise $y_a[n]$ is chosen. Due to the estimation noise $\epsilon[n] = y_o[n] - y_e[n]$ arised by the imperfect error correction, there exists a slight residual noise in ANT output $\hat{y}[n]$. However, if the impact of soft error is within the noise margin of system, the system does not suffer any algorithmic performance loss. In fact, ANT has been effectively used for combating deep submicron (DSM) noise [11]-[13]. Note, however, it cannot be directly applied for soft error tolerance, since the error-free assumption of estimator is no longer valid.

3. ALGORITHMIC SOFT ERROR-TOLERANCE

In this section, we assume the commonly used single event upset (SEU) model for soft error where any single internal node anywhere in the system can generate transient error during any clock period. Thus, TMR technique can perfectly detect and correct SEU and hence is a benchmark for comparison of energy-efficiency.

3.1. ASET

Figure 2 shows the block diagram of the proposed spatial ASET technique. The MDSP block executes a majority of signal processing tasks while the error control (EC) block detects and corrects errors. The decision block compares the MDSP output $y_a[n]$ and first estimator ($E1$) output $y_{e,1}[n]$. If the metric $d(y_a[n], y_{e,1}[n])$ is smaller than T_h then $y_a[n]$ is chosen as the output. Otherwise, an additional comparison checks whether the outputs of $E1$ and $E1$ are equal, i.e., $y_{e,1}[n] = y_{e,2}[n]$. If they are equal, MDSP error is declared otherwise an estimator error is declared. In case of MDSP error, we choose $\hat{y}[n] = y_{e,2}[n]$. In case of an estimator error, we choose $\hat{y}[n] = y_a[n]$. Error control in spatial ASET can be summarized as

$$\hat{y}[n] = \begin{cases} y_a[n] & \text{if } d(\,y_a[n], y_{e,1}[n]\,)\,| \le T_h \\ y_a[n] & \text{if } \ d(\,y_a[n], y_{e,1}[n]\,) > T_h, \ \ y_{e,1}[n] \ne y_{e,2}[n] \\ y_{e,1}[n] & \text{if } \ d(\,y_a[n], y_{e,1}[n]\,) > T_h, \ \ y_{e,1}[n] = y_{e,2}[n]. \end{cases} \quad (5)$$

3.2. Estimator

The role of the estimator is to generate $y_e[n]$ which is close to noiseless MDSP output $y_o[n]$ with low complexity. We first describe the RPR and then discuss prediction estimator.

1) Reduced Precision Redundancy (RPR) based Estimator: A reduced precision redundancy (RPR) estimator [13] employs a replica of the MDSP with reduced precision operands. The RPR estimator output $y_{e,rpr}[n]$ is given by

$$y_{e,rpr}[n] = \sum_{k=0}^{N-1} h_r[k] x_r[n-k] \quad (6)$$

where $h_r[n]$ and $x_r[n]$ are the truncated versions of the MDSP filter coefficient and input, respectively. When an error is detected using (5), $y_{e,rpr}[n]$ is employed as the final output. Since the LSB quantization noise $\sigma^2_{\epsilon,rpr}$ of RPR is sufficiently small for a reasonable replica precision and error happens infrequently, the overall impact on performance is insignificant. Indeed, RPR satisfies the desired performance with only small margin. In addition, RPR can be employed for a wide range of bandwidths because the quantization noise is not quite related to the bandwidth of systems.

2) Forward-predictor based Estimator: In a prediction-based estimator [11], a low-complexity forward predictor is employed for estimating the current filter output $y_o[n]$ from the past MDSP outputs $y_o[n-k], k > 0$. Since the adjacent outputs of MDSP are highly correlated for a sufficiently narrowband filter, an N_p-tap predictor generates an estimate $y_{e,prd}[n]$ to be used as an output when the soft error occurs. The output of N_p-tap forward predictor is

$$y_{e,prd}[n] = \sum_{k=0}^{N_p-1} h_p[k]y[n-k-1] \quad (7)$$

where $h_p[n]$ is obtained by solving the Wiener-Hopf equation. In general, the predictor performs well in narrowband systems and environments with low probability p_e of particle hits. However, a prediction performance degrades severely as the filter bandwidth increases and therefore fails in a wide bandwidth systems.

4. ASET PEFORMANCE ANALYSIS

4.1. Design criterion

The corrected output $\hat{y}[n]$ of ASET can be written as

$$\hat{y}[n] = y_o[n] + \gamma_r[n] \quad (8)$$

where $\gamma_r[n]$ is the residual noise of soft error after error correction. In order to meet a specific desired SNR (SNR_{des}), the SNR (signal-to-noise ratio) at the output in the presence of soft error needs to satisfy:

$$SNR_{ANT} = 10\log_{10}\frac{\sigma_d^2}{\sigma_\eta^2 + \sigma_{\gamma_r}^2} \geq SNR_{des}. \quad (9)$$

where σ_d^2 and σ_η^2 are the powers of desired signal $d[n]$ and channel noise $\eta[n]$, respectively, and $\sigma^2_{\gamma_r}$ is the power of $\gamma_r[n]$. Rewriting (9), we obtain the following condition for the residual noise power as

$$\sigma_d^2 \cdot 10^{-\frac{SNR_{des}}{10}} \geq \sigma_\eta^2 + \sigma^2_{\gamma_r} \quad (10)$$

Fundamentally, ASET exploits the trade-off between the channel noise power σ_η^2 and the residual noise power due to soft error $\sigma^2_{\gamma_r}$. As illustrated in Fig. 3, we can achieve the desired performance SNR_{des} with only slight margin.

By employing the decision rule provided in Section II.B, four possible decision events can happen. The soft error $\gamma[n]$ can go undetected if it is small enough. This undetected error (UER) event will result in $\hat{y}[n] = y_a[n]$. If $\gamma[n]$ is large enough then the error is detected. The detected error (DER) event results in the proper decision $\hat{y}[n] = y_e[n]$. In case of no error (NER) event, i.e., $\gamma[n] = 0$, original output will be chosen, i.e., $\hat{y}[n] = y_a[n] = y_o[n]$. However, when the threshold T_h is less than the magnitude of estimation noise $\epsilon[n] = y_o[n] - y_e[n]$, a false alarm error (FER) event can occur and thus $y_e[n]$ is falsely chosen as an output. We denote the probability of UER, DER, FER and NER events to be P_{uer}, P_{der}, P_{fer} and P_{ner}, respectively. In addition, we denote Y_o, Y_e, Y_e and $\hat{Y}$ as a random variable of $y_o[n]$, $y_e[n]$, $y_e[n]$ and $\hat{y}[n]$, respectively. Note, due to the stationary assumption, we drop the time index n. The following theorem describes the noise performance of an ASET-based system.

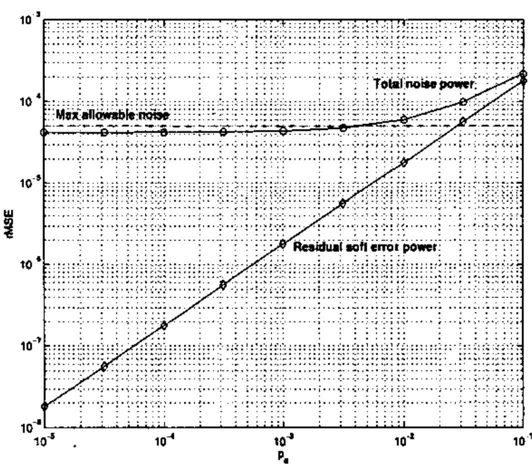

Fig. 3. Principle of ASET.

Theorem 1 *The residual mean square error (rMSE) $\sigma^2_{\gamma_r}$ of soft error $\gamma_r[n]$ is given by*

$$\sigma^2_{\gamma_r} = P_{der}\cdot\sigma^2_{der} + P_{uer}\cdot\sigma^2_{uer} + P_{fer}\cdot\sigma^2_{fer}, \quad (11)$$

where

$$\begin{aligned} P_{uer} &= P_r(|Y_a - Y_e| \leq T_h, Y_a \neq y_o) & (12)\\ P_{der} &= P_r(|Y_a - Y_e| > T_h, Y_a \neq y_o) & (13)\\ P_{fer} &= (1-p_e)P_r(|Y_o - Y_e| > T_h) & (14)\\ \sigma^2_{der} &= E[|Y_o - Y_e|^2 \,|\, |Y_a - Y_e| > T_h] & (15)\\ \sigma^2_{uer} &= E[|Y_o - Y_a|^2 \,|\, |Y_a - Y_e| \leq T_h] & (16)\\ \sigma^2_{fer} &= E[|Y_o - Y_e|^2 \,|\, |Y_o - Y_e| > T_h] & (17)\end{aligned}$$

The resulting rMSE $\sigma^2_{\gamma_r}$ in (11) depends on the 1) channel condition (p_e) and 2) effectiveness of ASET scheme, i.e., performance of estimator ($y_e[n] \sim y_o[n]$) and the decision threshold T_h.

4.2. Residual Mean Square Error

In this subsection, we present the detailed residual MSE analysis of ASET in SEU condition. In obtaining the results, we use the soft error model in (4). In addition, we assume that the estimation noise $\epsilon[n] = y_o[n] - y_e[n]$ is uniform over a sufficiently small region around 0.

We first investigate the noise power for each disjoint event. One can show that σ^2_{uer} is given by

$$\begin{aligned}\sigma^2_{uer} &= E[|Y_o - Y_a|^2 \,|\, |Y_a - Y_e| \leq T_h] \\ &\sim \sigma_r^2 + \frac{1}{3}T_h^2, & (18)\end{aligned}$$

where σ_ϵ^2 is the pure estimation noise power of estimator. In case of DER event, since the estimator output is chosen, $\sigma^2_{der} = \sigma^2_\epsilon$.

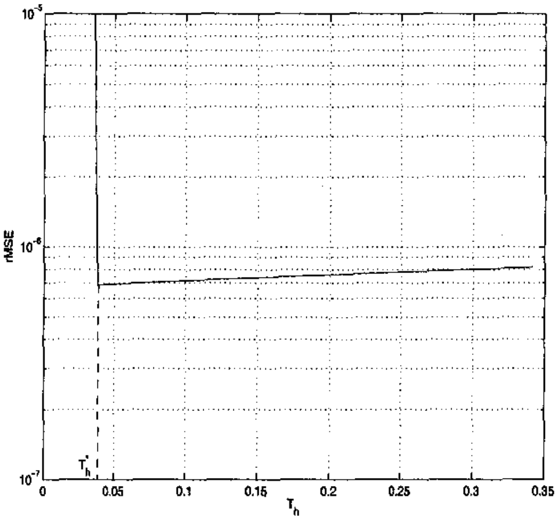

Fig. 4. rMSE $\sigma^2_{\gamma_r}$ vs. T_h.

Note that σ^2_{uer} depends quadratically on T_h and always greater than σ^2_{der}. In addition, σ^2_{fer} is

$$\begin{aligned}\sigma^2_{fer} &= E[\,|Y_o - Y_e|^2\,|\,|Y_o - Y_e| > T_h] \\ &\sim \frac{1}{3}(\epsilon^2_{\max} - \frac{T_h^3}{\epsilon_{\max}}). \end{aligned} \quad (19)$$

With these, we are now able to find the optimal threshold T_h^*.

Theorem 2 *The optimum threshold that minimizes the rMSE $\sigma^2_{\gamma_r}$ of ASET is the maximum estimation noise between $y_o[n]$ and $y_e[n]$, i.e.,*

$$T_h^* = \epsilon_{\max} = \max_{\forall n} |\, y_o[n] - y_e[n]\,|. \quad (20)$$

As an example, for the case of a 30-tap FIR filter ($\omega_c = 0.2\pi$) with 16 bit MDSP and 8 bit replica estimator, the rMSE performance for the threshold T_h variation is shown in Fig. 4. As the T_h decreases from T_h^*, the performance drops rapidly due to the abrupt increase in FER probability. Whereas, when $T_h > T_h^*$, the performance degrades gradually due to the increase in UER probability. In this case, performance is not quite sensitive to the small change in threshold. Hence we can choose the T_h which can be implemented easily.

Using the optimum threshold in (20), only DER and UER event contribute to the noise power. Thus, (11) is simplified to

$$\sigma^2_{\gamma_r} = P_{der}\,\sigma^2_\epsilon + P_{uer}\,\sigma^2_{uer} \quad (21)$$

$$\sim p_e\sigma^2_\epsilon + \frac{1}{3}P_{uer}\epsilon^2_{\max}. \quad (22)$$

It is interesting to note that the first term of (22) is due to the imperfection of error control scheme and the second term is due to the imperfection of error detection scheme. In case of TMR, which is the extreme of spatial ASET, $\sigma^2_\epsilon = 0$ and $P_{uer} = 0$ and hence $\sigma^2_{\gamma_r} = 0$.

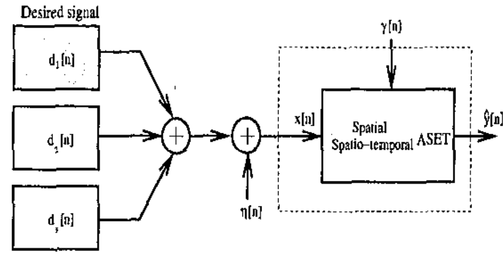

Fig. 5. Simulation setup for ASET

5. DISCUSSION AND SIMULATION RESULTS

In this section, we study the performance and power savings of the proposed ASET techniques over the MDSP, ANT, and TMR in the context of frequency selective filtering. In all simulations, we employ RPR scheme as an estimator for ASET. We first describe the simulation setup and then present the simulation results.

5.1. Simulation Setup

Figure 5 illustrates the system being considered. The MDSP block is a filter that extracts a signal $d_1[n]$ in the presence of other signals $d_2[n]$, $d_3[n]$ and white Gaussian noise source $\eta[n]$.

1) Soft error generation and injection: In order to emulate the realistic single event upset environment, we implemented a gate level multiplier-and-accumulator (MAC). When a particle hit occurs, randomly selected internal node is flipped and all the subsequent logic networks are affected accordingly. Particle hit block is chosen based on the area ratio so that the victim node location can be either the MDSP or the estimator. Assuming the probability of error in the entire system to be P_{er}, the probability of error in the MDSP p_e and each estimator p'_e of ASET are given by

$$p_e = P_{er}\left(\frac{A_{mdsp}}{A_{mdsp} + 2\,A_{est}}\right) \quad (23)$$

$$p'_e = P_{er}\left(\frac{A_{est}}{A_{mdsp} + 2\,A_{est}}\right). \quad (24)$$

where A_{mdsp} and A_{est} are the area of MDSP and estimator, respectively.

2) System specification: Each signal $d_i[n]$ used in our simulation has a bandwidth of 0.25π with a guard band of 0.1π. In addition, a Gaussian noise having -30 dB below the signal level is added at the receiver. In order to extract the desired signal, a low-pass filter having cutoff frequency of 0.3π is employed.

The error-free desired output SNR and SNR of the ASET techniques are given by

$$SNR_{aset} = 10\log_{10}\frac{\sigma^2_d}{\sigma^2_{d-\hat{y}}} \quad (25)$$

where $\sigma^2_{d-\hat{y}} = E[|d[n] - \hat{y}[n]|^2]$.

We assume $SNR_{des} = 25$ dB where the MDSP is designed to have a 1 dB margin for all techniques except TMR. Note that TMR is designed without any margin because it can perfectly correct errors under the SEU assumption. The original MDSP can provide this SNR with 33-tap (30-tap in case of TMR) MDSP filter along with 16×16 ($b = 16$) multipliers. The replica estimator has 8×8 ($b_r = 8$) multipliers with the same number of taps as the MDSP block.

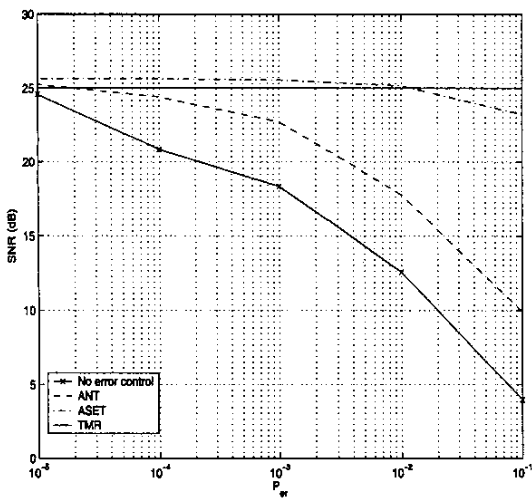

Fig. 6. SNR performance of ASET schemes vs. P_{er} in SEU environment.

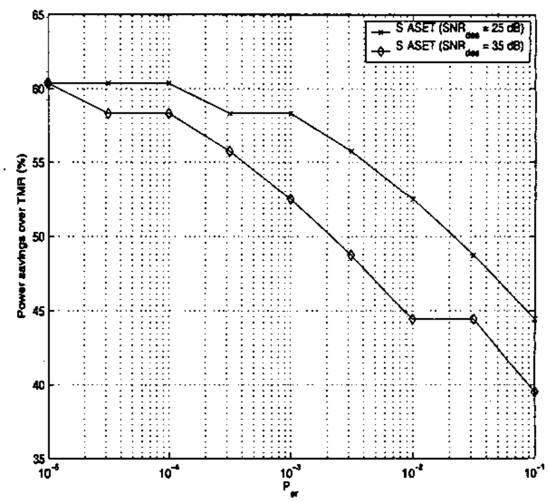

Fig. 7. Power savings vs. P_{er}.

5.2. Simulation Results

Figure 6 shows output $SNRs$ for all techniques. We see that the final output SNR for the system without any error control drops severely (18dB) as P_{er} increases from 10^{-5} to 10^{-1}. Since the error control of ANT scheme fails when the soft errors occur in the estimator, the SNR loss becomes significant, i.e., > 13 dB if $P_{er} = 10^{-1}$. Whereas, the plots of ASET lie exactly on top of each other and maintain the SNR until P_{er} approaches 10^{-2} providing more than three orders of magnitude improvement in soft error tolerance over the original MDSP.

In order to evaluate the power savings of the proposed techniques over the TMR technique, we searched the precision of replica estimator for each value of P_{er}. The resulting power savings for P_{er} variation is shown in Fig. 7, where 40% to 61% power savings can be achieved from $P_{er} = 10^{-1}$ to $P_{er} = 10^{-5}$.

6. REFERENCES

[1] K. L. Shepard and V. Narayanan, "Noise in deep submicron digital design," *Proc. of ICCAD*, vol. 1, pp. 524-531, 1996.

[2] J. R. Srour and J. M. McGarrity, "Radiation effects on microelectronics in space," *Proc. of IEEE*, vol. 76, pp. 1443-1469, 1988.

[3] 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.

[4] K. Zhang, S. Hareland, B. Senyk, and J. Maiz, "Methods for reducing soft errors in deep submicron integrated circuits," *Proc. of solid-state and integrated circuit technology*, pp. 516-519, 1998.

[5] L. Wissel, S. Pheasant, R. Loughran, and C. LeBlanc, "Managing soft errors in ASICs," *Custom Integrated Circuits Conf.*, Vol. 1, pp. 85-88, May 2002.

[6] P. Shivakumar et. al., "Modeling the effect of technology trend on the soft error rate of combinational logic," *Proc. of Intl. Conf. on Dependable Systems and Networks*, pp. 389 -398, 2002.

[7] The 2001 International Technology Roadmap for Semiconductors [On-line]. Available: http://public.itrs.net/Files/2001ITRS/Home.htm.

[8] R. Baumann, "The impact of technology scaling on soft error rate performance and limits to the efficacy of error correction," *Electron Devices Meeting*, pp. 329-332, Dec. 2002.

[9] L. Anghel, D. Alexandrescu, and M. Nicolaidis, "Evaluation of a soft error tolerance technique based on time and/or space redundancy," *Integrated Circuits and Systems Design*, pp. 237-242, 2000.

[10] P. Mazumder, "An on-chip ECC for correcting soft errors in DRAMs with trench capacitors," *IEEE Journal of Solid-State Circuits.*, vol. 27, pp. 1623 -1633, Nov. 1992.

[11] R. Hedge and N. R. Shanbhag, "Soft digital signal processing," *IEEE Trans. on VLSI*, vol. 9, pp. 813-823, Dec. 2001.

[12] B. Shim and N. R. Shanbhag, "Low-power digital filtering via reduced-precision redundancy," *Proc. of Asilomar conf.*, vol. 1, pp. 148-152, Nov. 2001.

[13] B. Shim, S. Sridhara and N. R. Shanbhag, "Low-Power digital signal processing via reduced precision redundancy," *Submitted to IEEE Trans. on VLSI.*, May 2003.

[14] B. Shim and N. R. Shanbhag, "Performance analysis of algorithmic noise-tolerance techniques," *Proc. of ISCAS.*, vol. 4, pp. 113-116, May. 2003.

[15] K. G. Shin and H. Kim, "A time redundancy approach to TMR failures using fault-state likelihoods," *IEEE Trans. on Computers*, vol. 43, pp. 1151-1162, Oct. 1994.