SOFT NMR: ANALYSIS & APPLICATION TO DSP SYSTEMS

Eric P. Kim and Naresh R. Shanbhag

University of Illinois at Urbana-Champaign
Coordinated Science Laboratory / Department of Electrical and Computer Engineering
1308 W Main St., Urbana, Illinois, USA, 61801

ABSTRACT

We have recently proposed the concept of *soft* N-modular redundancy (*soft NMR*) [1] in order to design robust and energy-efficient computing systems in nanoscale processes, where soft NMR was shown to achieve orders-of-magnitude improvement in robustness with significant power savings over NMR. In this paper, we analyze the performance of soft NMR and compare it with that of NMR and algorithmic noise-tolerance (ANT). An 8-b multiplier and a DCT-based still image compression system in a commercial $45nm$ CMOS process is considered. Two metrics of system performance: system reliability $P_{e,sys}$, and signal-to-noise ratio (SNR) are analyzed. We show that soft NMR always outperforms NMR and that our analysis predicts $P_{e,sys}$ and SNR to within 4.2% and $2.7dB$ on average, respectively, of the results of Monte Carlo simulations.

***Index Terms*—** robustness, nanotechnology, CMOS digital integrated circuits, image processing, MAP estimation

1. INTRODUCTION

Designing energy-efficient and robust DSP systems in nanoscale process technologies is a challenge. Techniques such as dynamic voltage scaling [2], algorithmic noise-tolerance (ANT) [3] (see Fig. 1(a)) and cross-layer error exploitation [4] have been proposed in the past for DSP applications. These techniques exploit the algorithmic aspects of computation, and thus are application-specific. N-modular redundancy (NMR) (see Fig. 1(b)) is a general fault-tolerance technique but comes with a heavy power and area overhead. Recently, we have proposed *soft* N-modular redundancy (soft NMR) [1] (see Fig. 1(c))as an alternative to NMR. Unlike NMR, soft NMR exploits the signal and error statistics in order to make a decision. Soft NMR enhances the robustness of NMR by two to three orders-of-magnitude, along with approximately 45% energy-savings, while preserving its generality. Soft NMR can also correct errors when all the processing element (PE) outputs are incorrect. In addition, unlike dual MR (DMR), which can only detect errors, soft dual-MR (soft DMR) can detect and correct errors.

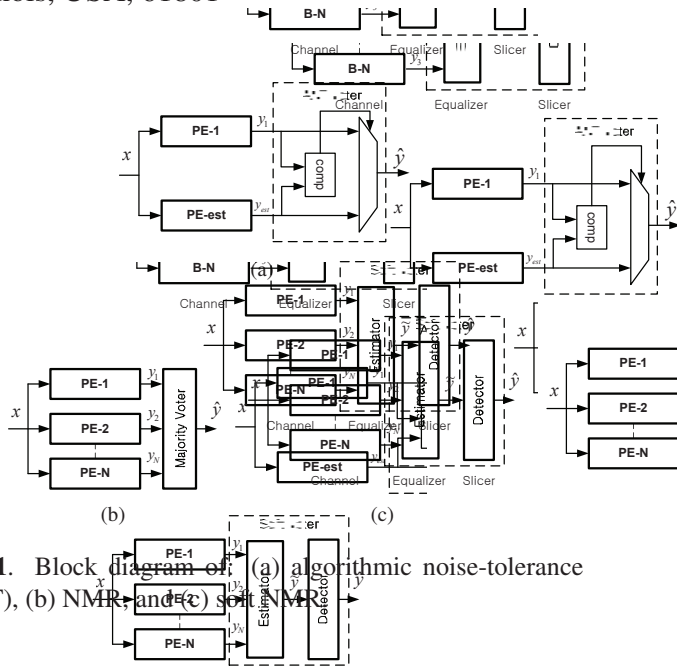


Fig. 1. Block diagram of: (a) algorithmic noise-tolerance (ANT), (b) NMR, and (c) soft NMR.

The significant benefits of soft NMR motivate us to analyze its performance systematically and compare it with ANT and NMR. Indeed, all three techniques achieve their robustness via a voter which makes a decision based on a set of observations and other relevant information. In this paper, we analyze soft NMR, NMR and ANT under a common framework. The paper is organized as follows: we present the analysis in Section 2, followed by simulations results in case of a DCT based image coder in Section 3.

2. ANALYSIS

In this section, we analyze soft NMR, NMR and ANT under a common framework.

2.1. Analysis Framework

The analysis framework is depicted in Fig. 2 in the context of soft NMR, where N PEs compute in parallel to produce an output $y_i = y_o + e_i$, where y_o is the correct value, and e_i is the error. Table 1 lists the notations employed in the analysis.

Soft NMR assumes that the error statistics at the PE output is known. Fig. 3(a) shows the error-statistics (error prob-

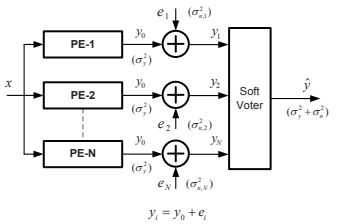


Fig. 2. Analysis framework for soft NMR.

Table 1. Definition of analysis framework parameters

Notation	Description
N	total number of replicated PEs
$\mathcal{V}$	*output space*: set of all possible outputs. Its size is m and its elements are denoted as $v_1, v_2, ..., v_m$. Note that $y_1, y_2, ..., y_N, \hat{y} \in \mathcal{V}$
$\mathcal{R}$	*observation space*: the set of all PE outputs $\{y_1, y_2, ..., y_N\}$
$\mathcal{H}$	*hypothesis space*: set of hypotheses to use in detection
r_j	the *prior* defined as the probability that $v_j \in \mathcal{V}$ is the correct output value y_o. Note that $\sum_{v_j \in \mathcal{V}} r_j = 1$.
q_j	the probability that PE output y_i is v_j given a fault has occurred
p_{e_i}	distribution of the error e_i of PE-i
P_{e_i}	error probability of PE-i
$c_j(\mathcal{R})$	occurrences of v_j in $\mathcal{R}$
$\sigma_y^2, \sigma_{n,i}^2$	signal power and noise power

ability mass function (PMF)) due to timing violations at the output of a 16-b ripple carry adder. The error PMF indicates large magnitude errors to have a higher probability than small magnitude errors. In order to simplify the analysis, we assume that the error PMF is as shown in Fig. 3(b). We define the probability of $e \neq 0$ as the error probability p_e, and the corresponding (large) error magnitude as d. With e_i ranging from $-d$ to d, $p_{e_i}(e_i = -d) = p_{e_i}(e_i = d) = \frac{p_e}{2}$, $p_{e_i}(e_i = 0) = 1 - p_e$ and $p_{e_i}(e_i) = 0$ for all other values of e_i.

Analysis is performed employing two metrics: 1) the system error probability $P_{e,sys}$, which is the probability that the final output $\hat{y} \neq y_o$ (see Fig. 2), i.e., $P_{e,sys} = 1 - P(\hat{y} = y_0)$, and 2) the signal-to-noise ratio (SNR), which is calculated as the ratio of the signal power to the noise power $\frac{\sigma_y^2}{E[(y_o - \hat{y})^2]}$, where the E operator denotes the expected value.

The computation of SNR requires the system error probability $P_{e,sys}$ as the distribution of the final output $\hat{y}$ is a mixture distribution of two PMFs, as shown in Fig. 4. From this the distribution of $\hat{y}$ is given by (see Table 1):

$$p_{\hat{y}}(v_j) = P_{e,sys}(r_j * p_{e_i}) + (1 - P_{e,sys})r_j \qquad (1)$$

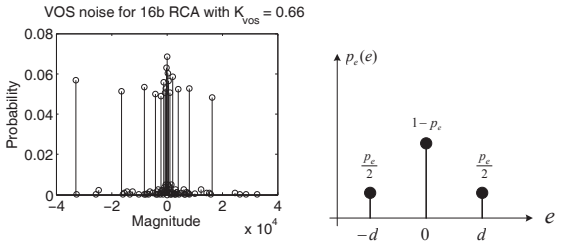


Fig. 3. Probability mass function (PMF) of errors due to timing violations: (a) error statistics from a 16-b RCA, and (b) simplified statistics used in analysis.

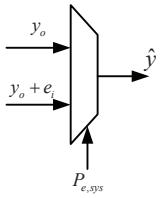


Fig. 4. The output of ANT, NMR and soft NMR can be modeled as a mixture distribution.

where $*$ denotes the convolution operation.

It can be easily shown that the first and second moments of $\hat{y}$ are given by

$$E[\hat{y}] = E[y_o] + P_{e,sys}E[e_i] \qquad (2)$$

$$E[(\hat{y} - E[\hat{y}])^2] = (1 - P_{e,sys})\sigma_{\hat{y}}^2 + P_{e,sys}\sigma_n^2 + \sigma_{y_o}^2 - \sigma_{e_i}^2 \qquad (3)$$

from which $SNR = \frac{\sigma_y^2}{E[(y_o - \hat{y})^2]}$ can be easily calculated.

2.2. Analysis of Soft NMR

The soft voter employs the *maximum a posteriori* (MAP) principle, which is optimal in the sense of minimizing the system error probability $P_{e,sys}$ by choosing the most probable value from a hypotheses set $\mathcal{H}$, given observations $\mathcal{Y}$, error statistics $p_{e_i}()$ along with the prior information r_j. The soft voter algorithm employed in this paper, is obtained by choosing $\mathcal{H} = \mathcal{Y}$ for complexity reasons, and is given as follows:

$$\hat{y} = \underset{v_j \in \{y_1, ..., y_N\}}{\arg\max} \; r_j \prod_{i=1}^{N} p_{e_i}(e_i = y_i - v_j) \qquad (4)$$

The error PMF in Fig. 3(b) indicates that there are only three possible values of PE output y_i. Thus, a simplified voting algorithm can be obtained as follows: When the voter observes three different values or only one value, the output $\hat{y} = y_o$ can be obtained. When only two values, v_i and $v_j = v_i - d$, are

observed n times and $N-n$ times, respectively, the detection rule becomes:

$$n \underset{v_j}{\overset{v_i}{\gtrless}} \frac{N}{2} + \frac{1}{2}\frac{log\frac{r_i}{r_j}}{log\frac{2-2p_e}{p_e}} \tag{5}$$

where it was assumed that $p_e < \frac{2}{3}$ or else the direction of inequality should be reversed. Thus the $P_{e,sys}$ for soft NMR can be shown to be:

$$P_{e,sys} = \sum_{i=0}^{m-1} r_i \left\{ p_e^N (1 - \frac{1}{2^{N-1}}) + 2 \sum_{k=0}^{\lfloor \frac{N}{2} + \frac{log\frac{r_i}{r_j}}{2log\frac{2-2p_e}{p_e}} \rfloor} \binom{N}{k} (1-p_e)^k \left(\frac{p_e}{2}\right)^{N-k} \right\} \tag{6}$$

which, in case of uniform priors, simplifies to:

$$P_{e,sys} = 2\sum_{i=0}^{\frac{N}{2}} \binom{N}{i} (1-p_e)^i \left(\frac{p_e}{2}\right)^{N-i} + p_e^N - 2\left(\frac{p_e}{2}\right)^N \tag{7}$$

For an arbitrary error PMF, $P_{e,sys}$ is given by

$$P_{e,sys} = \sum_{v_j \in \mathcal{V}} r_j \left(\sum_{e_i \in \mathcal{A}} p_{e_i}(e_i) \right) \tag{8}$$

$$\mathcal{A} = \left\{ e_i | r_j \left(\prod_{l=0,\ldots,m-1, l\neq i} q_l^{|c_l(R)|} \right) \cdot \left(\frac{1-P_{e_i}}{P_{e_i}} \right)^{|c_i(R)|} > r_j \left(\prod_{l=0,\ldots,m-1, l\neq j} q_l^{|c_l(R)|} \right) \cdot \left(\frac{1-P_{e_i}}{P_{e_i}} \right)^{|c_j(R)|} \right\} \tag{9}$$

which is easier to compute as compared to Monte Carlo simulations. Substituting (6) in (2) and (3) provides the SNR estimate.

2.3. Analysis of NMR

NMR has been extensively analyzed in [5] but without employing error PMFs as is done in this paper. When presented with a set of N PE outputs $Y = \{y_1, \ldots, y_N\}$, a majority voter produces an output $\hat{y}$ given by

$$\hat{y} = maj(y_1, y_2, ..., y_N) \tag{10}$$

where $maj(Y)$ selects that element of Y which occurs more than $\lfloor N/2 \rfloor$ times. In the absence of a majority, the element with the most occurrences could be chosen (plurality) or an error can be flagged. Each PE produces the correct output y_o with a probability $1-P_{e_i}$ independently (see Table 1). As the probability to choose the correct output can be given as

$$\sum_{v_j \in V} P\left(c_j(R) > \frac{N}{2} \mid v_j = y_o \right) P(v_j = y_o) \tag{11}$$

the error probability of NMR is given by,

$$P_{e,sys} = \sum_{v_j \in V} r_j \sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} \binom{N}{k} (P_{e_i})^{N-k} (1-P_{e_i})^k \tag{12}$$

As before, we substitute (12) in (2) and (3) to obtain the SNR.

Comparing (7) to the performance of NMR using a majority voter (12), we can easily show that

$$P_{e,sys-nmr} \geq P_{e,sys-softNMR} \tag{13}$$

and thus soft NMR will always be equal to or better than NMR in terms of robustness.

2.4. Analysis of ANT

ANT can be viewed as a special case of NMR with two PEs, a main block (PE-1) and an estimator (PE-est), each with different error statistics, and the detector chooses among the two PE outputs by comparing $|y_1 - y_{est}|$ to a threshold τ. In this analysis, we will assume the estimation error is $-d/2 \leq e_{est} \leq d/2$; i.e., the estimation error is smaller than hardware errors.

Considering the error PMF in Fig. 3(b), if $\tau \geq \frac{3d}{2}$, then $\hat{y} = y_1$, and $P_{e,sys} = 1 - p_{e_1}(0)$. If $\tau < \frac{d}{2}$, then $\hat{y} = y_1$ when PE-1 is error-free (which is the correct output), else $\hat{y} = y_{est}$. When $\frac{d}{2} \leq \tau < \frac{3d}{2}$, $P_{e,sys}$ will depend on the values of e_1 and e_{est}. Thus, $P_{e,sys}$ for ANT can be calculated as:

$$P_{e,sys} = \begin{cases} p_e(1 - p_{e_{est}}(0)), & \text{when } \tau < \frac{d}{2} \\ \frac{p_e}{2} \sum_{d-\tau \leq |e_{est}| \leq \frac{d}{2}} p_{e_{est}}(e_{est}), & \text{when } \frac{d}{2} \leq \tau < d \\ \frac{p_e}{2} \left\{ 1 - p_{e_{est}}(0) + \sum_{0 < |e_{est}| \leq \tau - d} p_{e_{est}}(e_{est}) \right\}, & \text{when } d \leq \tau < \frac{3d}{2} \\ p_e, & \text{when } \tau \geq \frac{3d}{2} \end{cases} \tag{14}$$

The probability of error of ANT for an arbitrary error statistic is:

$$1 - P(E_1) - P(E_2) \tag{15}$$

where E_1 is the event when $\hat{y} = y_1$ and $e_1 = 0$, and E_2 is the event when $\hat{y} = y_{est}$ and $e_{est} = 0$. Thus (15) becomes:

$$P_{e,sys} = 1 - p_{e_1}(0) \sum_{|e_{est}| < \tau} p_{e_{est}}(e_{est}) - p_{e_{est}}(0) \sum_{|e_1| > \tau} p_{e_1}(e_1) \tag{16}$$

3. SIMULATION RESULTS

In this section, we compare Monte Carlo simulation results with analytical results of Section 2. First, we show the results for an 8-b multiplier, and then we consider a discrete cosine transform (DCT) based image coder.

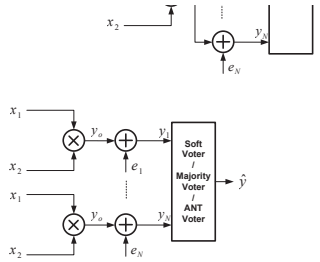
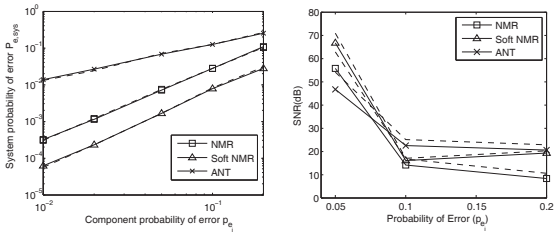


Fig. 5. Simulation setup for an 8-b multiplier.



(a) (b)

Fig. 6. Comparison of analytical (dashed) and simulation (bold) results for an 8-b multiplier: (a) $P_{e,sys}$ metric, and (b) SNR metric.

3.1. Multiplier

We employed a 6-b reduced precision version of the multiplier as an estimator for ANT. The error PMF in Fig. 3 was employed to generate independent errors for each of the N multiplier copies, as shown in Fig. 5. Figure 6 shows that the analysis predicts the simulation results to on average within 0.1% for $P_{e,sys}$ and 2dB for SNR.

3.2. Discrete Cosine Transform Based Image Coder

We now consider a discrete cosine transform (DCT) based image coder. The simulation setup is shown in Fig. 7. The DCT transmit chain is replicated, and the errors are corrected after the quantizer. A 3-b reduced precision version of the DCT block is used for the ANT estimator. The voters and the ANT estimator are all assumed to operate error free. The errors were generated via an RTL *Verilog* simulation of the DCT, where we first characterize the basic building blocks such as

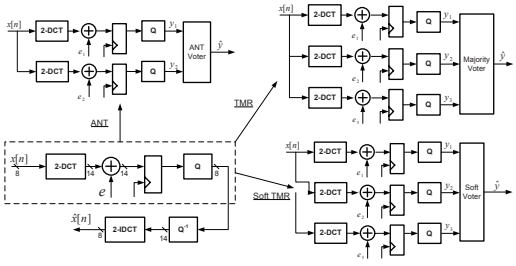
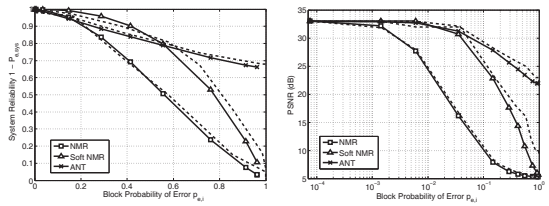


Fig. 7. Simulation setup for DCT



(a) (b)

Fig. 8. Comparison of analytical (dashed) and simulation (bold) results for a DCT image coder: (a) $P_{e,sys}$ metric, and (b) $PSNR$ metric.

a 1-b full-adder (FA) in a commercial $1.2V$, $45nm$ process technology using *HSPICE* to obtain delay vs. supply voltage curves. Thus, in this example, the errors were generated as true timing violations.

The peak signal-to-noise ratio ($PSNR$) is employed as the SNR metric, where $PSNR = 10log_{10}\left(\frac{(255)^2}{E[(y_o-\hat{y})]}\right)$ and the denominator is obtained by using the same method used for SNR calculations. The results are shown in Fig. 8, which shows that the $P_{e,sys}$ obtained via analysis is within 0.2% for NMR, 4.2% for soft NMR, and 1.5% for ANT as compared to Monte Carlo simulations. In addition, the $PSNR$ obtained by analysis is on average within 0.5dB for NMR, within 2.7dB for soft NMR, and within 1.8dB for ANT, as compared to simulations.We can also confirm that soft NMR can tolerate a magnitude higher error rate for a given PSNR compared to NMR.

These results indicates that the analysis presented in this paper captures the error correction aspects of these three techniques quite accurately and can be employed for designing such error-resilient systems in the future.

4. REFERENCES

[1] E.P. Kim, R.A. Abdallah, and N.R. Shanbhag, "Soft nmr: Exploiting statistics for energy-efficiency," in *International Symposium on System-on-Chip (SOC)*, 2009, pp. 052–055.

[2] I.S. Chong and A. Ortega, "Dynamic voltage scaling algorithms for power constrained motion estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2007, vol. 2, pp. II–101–II–104.

[3] R. Hegde and N.R. Shanbhag, "Soft digital signal processing," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 9, no. 6, pp. 813–823, 2001.

[4] A.K. Djahromi, A.M. Eltawil, F.J. Kurdahi, R. Kanj, and I. UC, "Cross layer error exploitation for aggressive voltage scaling," in *8th International Symposium on Quality Electronic Design (ISQED)*, 2007, pp. 192–197.

[5] D.M. Blough and G.F. Sullivan, "A comparison of voting strategies for fault-tolerant distributed systems," in *Proceedings of Ninth Symposium on Reliable Distributed Systems*, 1990, pp. 136–145.