STATISTICAL ANALYSIS OF ALGORITHMIC NOISE TOLERANCE

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, IL 61801
[epkim2, shanbhag]@illinois.edu

ABSTRACT

Algorithmic noise tolerance (ANT) is an effective statistical error compensation technique for digital signal processing systems. This paper proves a long held hypothesis that ANT has a strong Bayesian foundation, and develops an analytical framework for predicting the performance of, and designing performance-optimal ANT-based systems. ANT is shown to approximate an optimal Bayesian detector and an optimal minimum mean squared error (MMSE) estimator. We show that the theoretically optimum threshold and the optimal threshold obtained via Monte Carlo simulations agree to within 8%, with performance degradation of at most 2.1% for a variety of error probability mass functions. For a 2D-DCT implemented in a 45nm CMOS process, we find similar results where the thresholds have a 7.8% difference. Furthermore, both analysis and simulations indicate that ANT's probability of error detection is robust to the choice of the threshold.

Index Terms— Low-power, error-resiliency, Bayesian, detection, estimation, voltage overscaling

1. INTRODUCTION

As CMOS technology scales into the sub-22nm regime, non-idealities due to process, temperature and voltage variations, and soft errors, are increasingly becoming commonplace. These variations often result in uncertain gate delays and leakage currents leading to intermittent errors in computation. This trend is expected to worsen in the next decade [1]. Present day approaches seek to avoid these errors by designing for the worst-case scenario. These methods are often wasteful and often unaffordable in many power-limited applications. Design methods based on error resiliency are much more energy efficient.

Statistical techniques have been developed to improve the performance of signal processing systems when subject to errors and noise [2, 3]. Recently, such statistical techniques have been applied to tolerate errors in hardware [4]. Such statistical error compensation (SEC) techniques have been applied in an ad-hoc manner. In this paper, we provide a theoretical analysis for algorithmic noise tolerance (ANT) [5], which is a specific SEC technique.

ANT tolerates errors by employing a lower-complexity estimator designed to be less prone to error (see Fig. 1(a)) as compared to the main block. When the output of the estimator disagrees with the output of the main block by more than a certain threshold τ, the estimator output is used as the final corrected output. Though simple, ANT has proven to be very effective in tolerating errors, mainly because it compensates for errors statistically at the system level.

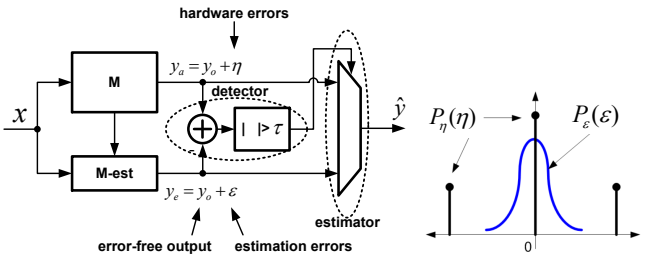
In the design of ANT, the threshold τ is an important design parameter that was previously chosen empirically. In this paper, we establish a link between ANT and the Bayesian detection and estimation framework. We show ANT can be viewed as a two step process, where detection of an error event is performed, followed by estimation of the correct value. The detection stage is a thresholding detector, while the estimation stage is an approximation to the minimum mean square error (MMSE) estimator. We further derive expressions for determining the optimal threshold τ, and verify it with Monte Carlo simulations. Thus, this paper proves the long held hypothesis that ANT has a strong Bayesian foundation. Furthermore, this work opens up the possibility of establishing a similar basis for other SEC techniques.

The remainder of the paper is organized as follows. Section 2 provides background on ANT. In Section 3 we formulate the detection and estimation problem, then derive the optimal decision rule under the Bayesian estimation framework. Section 4 compares the performance of ANT and the optimal decision rule via a simple example, and a 2D-DCT application. Section 5 concludes the paper.

2. ALGORITHMIC NOISE TOLERANCE

Algorithmic noise-tolerance (ANT) [5] in Fig. 1(a) incorporates a *main* block and an *estimator.* The main block is permit-

This work was supported in part by the Systems on Nanoscale Information fabriCs (SONIC), one of six centers supported by the STARnet phase of the Focus Center Research Program (FCRP), a Semiconductor Research Corporation program sponsored by MARCO and DARPA.



(a) (b)

Fig. 1: Algorithmic noise tolerance: (a) block diagram, and (b) error distributions.

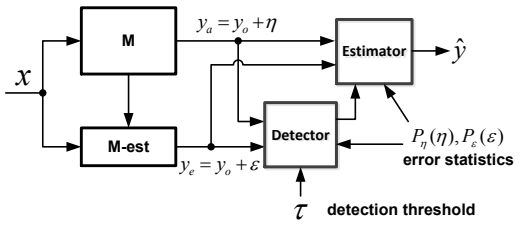


Fig. 2: The Bayesian Framework for ANT.

ted to make hardware/timing errors, but not the estimator. The estimator is a low-complexity block (typically 5%-to-20% of the main block complexity) generating a statistical estimate of the correct main block output, i.e.,

$$y_a = y_o + \eta \quad (1)$$
$$y_e = y_o + \epsilon \quad (2)$$

where y_a is the actual main block output, y_o is the error-free main block output, η is the hardware error, y_e is the estimator output, and ϵ is the estimation error. ANT exploits the difference in the statistics of η and ϵ (see Fig. 1(b)). The final/corrected output of an ANT system $\hat{y}$ is obtained via the following decision rule:

$$\hat{y} = \theta_{ANT}(\mathbf{y}) = \begin{cases} y_a, & \text{if } |y_a - y_e| < \tau \\ y_e, & \text{otherwise} \end{cases} \quad (3)$$

where $\mathbf{y} = (y_a, y_e)$ is the observation vector, and τ is an application-dependent parameter chosen empirically to maximize the performance of ANT. Under the conditions outlined above, it is possible to show that

$$SNR_{uc} \ll SNR_e \ll SNR_{ANT} \approx SNR_o \quad (4)$$

where SNR_{uc}, SNR_e, SNR_{ANT} and SNR_o are the signal-to-noise-ratios of the uncorrected main block (η dominates), the estimator (ϵ dominates), the ANT system, and the error-free main block (ideal), respectively. Thus, ANT detects and corrects errors approximately, but does so in a manner that satisfies an application-level performance specification (SNR). Several low-overhead SEC techniques have been proposed by exploiting data correlation, system architecture, and statistical signal processing techniques [6].

3. ANALYSIS OF ANT

In this section, we formulate a general detection and estimation problem with two observations. The optimal detection and estimation rule is derived and shown that ANT (Fig. 1) is a low-complexity approximation of this optimal solution.

3.1. Problem formulation

The detection and estimation problem assumes the observation set $\mathbf{y}$ (see Fig. 2) given by (1) and (2). The detector assumes binary hypotheses: 1) the error-free event H_0 (i.e. $\eta = 0$), and the error event H_1 (i.e. $\eta \neq 0$). The goal is to find the optimal decision rule $\delta(\mathbf{y}|P_\eta, P_\epsilon)$, which chooses one hypothesis based on the observation $\mathbf{y}$, and the error statistics P_η and P_ϵ. Then estimation is performed via an estimation rule $\hat{y} = \theta(\mathbf{y})$, which finds the best estimate of y_o based on a certain optimization criteria. This is different from a classical detection and estimation problem, as the hypotheses are not based on y_o, but instead on η, and that the estimation step utilizes the detection stage information to simplify the estimation process. This distinction turns out to be crucial in enabling a simplification of the error compensation scheme leading to ANT.

The following assumptions will be made. The error-free main block output y_o, and the errors η and ϵ are a fixed point number with bit width N, and defined as a random variable over the set $S = \{-2^N, ..., 2^N - 1\}$. Setting the bit width of the output as N, the cardinality of S is 2^N. Furthermore, y_o will be assumed to be uniformly distributed and thus $p(y_o) = \frac{1}{2^N}$. The errors η and ϵ are assumed to be independent. We will further assume that η is zero mean and P_η has a peak at zero (that represents the error-free probability), is symmetric about the mean, and $P_\eta(\alpha) \geq P_\eta(\beta)$ for $|\alpha| \geq |\beta|, \beta \neq 0$. On the other hand, ϵ is assumed to be zero mean, and P_ϵ is symmetric about the mean, and $P_\epsilon(\alpha) \leq P_\epsilon(\beta)$ for $|\alpha| \geq |\beta|$, as depicted in Fig. 3(a). Large magnitude values of ϵ are assumed to be extremely unlikely. These assumptions are motivated by noting that timing induced hardware errors are large in magnitude due to LSB first computation, and estimation errors are Gaussian (subsampled estimators) or uniform distributed (reduced precision estimators). An example conditional PMF of timing errors, conditioned on the error event induced by voltage overscaling (VOS) is given in Fig. 3(b). This error statistic was obtained through Verilog simulations from a 16-bit ripple carry adder with voltage overscaling applied. It can be seen that large magnitude errors occur with large probability which justifies our assumption for P_η.

3.2. Detection

We perform Bayesian hypothesis testing to detect the error event. For equally likely priors, the decision rule is given by comparing the likelihoods [7]:

$$p(H_1|\mathbf{y}) \underset{H_0}{\overset{H_1}{\gtrless}} p(H_0|\mathbf{y}) \tag{5}$$

where the likelihoods are given by:

$$\begin{aligned} p(H_0|\mathbf{y}) &= \frac{\frac{1}{2^N} P_\eta(0) P_\epsilon(y_e - y_a)}{p(\mathbf{y})} \\ p(H_1|\mathbf{y}) &= \frac{p(\mathbf{y}) - \frac{1}{2^N} P_\eta(0) P_\epsilon(y_e - y_a)}{p(\mathbf{y})} \end{aligned} \tag{6}$$

where the independence assumption between P_η and P_ϵ has been used. The posterior probability $p(\mathbf{y})$ can be shown to be:

$$\begin{aligned} p(\mathbf{y}) &= \sum_{y_o} p(y_o) P_\eta(y_a - y_o) P_\epsilon(y_e - y_o) \\ &= \frac{1}{2^N} \sum_{\eta} P_\eta(\eta) P_\epsilon(\eta + (y_e - y_a)) \end{aligned} \tag{7}$$

Defining the likelihood ratio $\mathcal{L} = \frac{p(H_1|\mathbf{y})}{p(H_0|\mathbf{y})}$, the optimal detection rule $\delta(\mathbf{y})$ will be denoted as:

$$\delta(\mathbf{y}) = \begin{cases} H_1 & \text{if } \mathcal{L} \geq 1 \\ H_0 & \text{if } \mathcal{L} < 1 \end{cases} \tag{8}$$

From (6) and (7), it can be seen that $\mathcal{L}$ only depends on the difference of the observed outputs $d = y_a - y_e$. Furthermore, given that P_ϵ is decreasing in distance from the mean and symmetric, $P_\eta(0)P_\epsilon(y_e - y_a)$ is a decreasing function in $|d|$. Also, $p(\mathbf{y})$ is symmetric, and convex in the half plane and given that P_ϵ is heavily concentrated around the mean, the minimum of $p(\mathbf{y})$ is close to 0. Thus we can approximate $p(\mathbf{y})$ as an increasing function in $|d|$. Now we can rewrite the decision rule as:

$$\delta(\mathbf{y}) = \begin{cases} H_1 & \text{if } |d| \geq \tau \\ H_0 & \text{if } |d| < \tau \end{cases} \tag{9}$$

where the theoretically optimal threshold $\tau^\star_{a,p}$ is the value of $|d|$ when $\mathcal{L}$=1, and satisfies:

$$\sum_{\eta} P_\eta(\eta) P_\epsilon(\eta + \tau^\star_{a,p}) = 2P_\eta(0) P_\epsilon(\tau^\star_{a,p}) \tag{10}$$

We can see that the resulting optimal decision rule (9) is equivalent to the detection rule (3) used in ANT. Furthermore, we have shown that the optimal threshold to be used in ANT is given by (10).

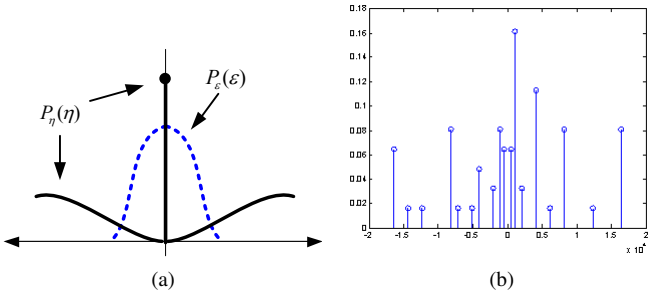


Fig. 3: Example error PMFs: (a) depiction of P_η and P_ϵ that increase or decrease in distance from the mean, and (b) voltage overscaling (VOS) induced timing errors.

3.3. Estimation

When H_0 is detected ($\eta = 0$), the main block output is used as the corrected output i.e., $\hat{y} = y_a$. When H_1 is detected, a more complex estimation needs to take place. There are many different optimality conditions available for estimation such as minimum mean squared error (MMSE), minimum mean absolute error (MMAE), and maximum *a posteriori* probability (MAP). In this paper, we will focus on MMSE.

The optimization criteria will be to minimize $E(\hat{y} - y_o)^2$ given $\mathbf{y}$ and that H_1 has been detected where the expectation is over y_o. It is well known that the MMSE estimator is the conditional mean, i.e. the estimate $\hat{y}(\mathbf{y}) = E(y_o|\mathbf{y}, H_1)$ [7]. As the posterior distribution of y_o is:

$$p(y_o|\mathbf{y}, H_1) = \frac{1}{1 - p_e} P_\eta(y_a - y_o) P_\epsilon(y_e - y_o) \tag{11}$$

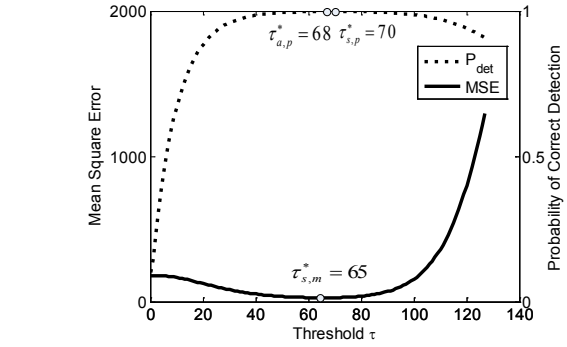
the MMSE optimal estimator will be:

$$\hat{y} = \theta(\mathbf{y}) = \begin{cases} y_a & \text{if } \delta(\mathbf{y}) = H_0 \\ \sum_{y_o \neq y_a} p(y_o|\mathbf{y}, H_1) y_o & \text{if } \delta(\mathbf{y}) = H_1 \end{cases} \tag{12}$$

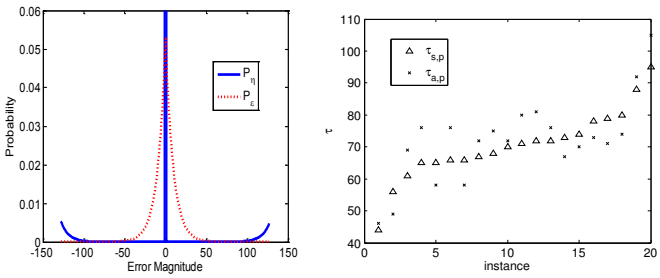
where y_a is excluded in the summation as it corresponds to H_o. Determining $\hat{y}$ is a complex and power hungry task. In ANT, an approximation to the optimal estimator has been made, such that $\hat{y} = y_e$. This estimator can be implemented by a simple mux with the output of the detection stage used as the control (Fig. 1(a)). In Section 4, we will show that such approximation results in minimal performance degradation.

4. COMPARISON OF SIMULATION AND ANALYSIS

In this section, we will first apply ANT on a simple example and compare the results with analysis. For this example, all signals are considered to be 8 bits ($N = 8$). The error PMFs P_η and P_ϵ are constructed as follows. P_ϵ is a truncated discrete Laplace distribution with parameter $a = 0.9$ normalized



(a)



(b) (c)

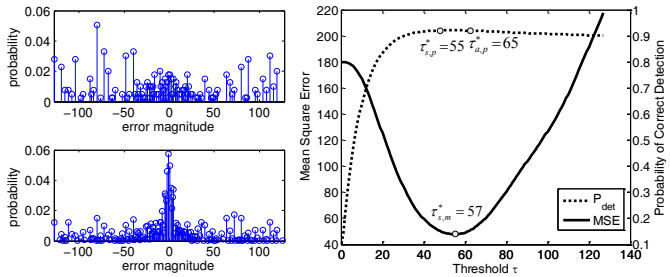
Fig. 4: Simulation results for a simple example: (a) probability of correct detection and MSE vs. τ, (b) error PMF, and (c) $\tau^{\star}_{s,p}$, and $\tau^{\star}_{a,p}$ for different PMFs.

by a constant A. P_η is a circularly shifted version of P_ϵ scaled with the probability of error $p_e = 0.1$, and a peak of $1 - p_e$ at 0. The PMF is depicted in Fig. 4(b), where the peak at $P_\eta(0)$ has been clipped for better visibility.

$$P_\epsilon(k) = A\frac{1-a}{1+a}a^{|k|}, -128 \leq k \leq 127 \quad (13)$$

$$P_\eta(k) = \begin{cases} 1-p_e & \text{if } k = 0 \\ p_e P_\epsilon(\left(k + 2^{N-1}\right)_{2^N}) & \text{if } k \neq 0 \end{cases} \quad (14)$$

The performance of ANT with varying thresholds are shown in Fig. 4(a). The design of an ANT system would go through such simulations to obtain the optimal threshold. For this example, Monte Carlo simulations show that the probability of correct detection is maximum at $\tau^{\star}_{s,p} = 70$, with $P_{det} = 0.9986$ which is very close to $\tau^{\star}_{a,p} = 68$, the theoretical optimum obtained from (10). Similarly, simulations indicate that the MSE is minimized at $\tau^{\star}_{s,m} = 65$, with MSE=27.21, which is also very close to the MSE=25.19 obtained via the optimal estimator in 12. Furthermore, the utility of our analysis in design can be seen by the fact that the MSE achieved via simulations with $\tau = \tau^{\star}_{a,p}$ is 33.55, which corresponds to a difference of 0.4% when normalized with respect to the maximum MSE obtained when $\tau = 128$. Note that the detection probability and MSE curves are flat around the optimal point. This shows that the exact value of



(a) (b)

Fig. 5: DCT Example: (a) error statistics of a voltage overscaled DCT block at $V_{dd} = 1.1V$ ($p_e = 0.0043$), and $V_{dd} = 1.0V$ ($p_e = 0.0374$), and (b) the resulting probability of correct detection and MSE vs. τ.

the optimal threshold is unnecessary to achieve near optimal performance, and that both $\tau^{\star}_{a,p}$ and $\tau^{\star}_{s,p}$ are good approximations for $\tau^{\star}_{s,m}$. Similar results (see Fig. 4(c)) were obtained with different forms for P_η and P_ϵ, as long as they met the assumptions outlined in Section 3. For the total of 20 cases shown in Fig. 4(c), the maximum difference in $\tau^{\star}_{s,p}$ and $\tau^{\star}_{a,p}$ was 11, while on average the difference was 6.6. In terms of MSE, this corresponds to a maximum degradation of 2.1%, with an average degradation of 0.6% when normalized to the maximum MSE.

We have also applied ANT to a 2D-DCT image compression application. The setup of the application is identical as in (1). The main block is an 8-bit input, 8-bit output, 8×8 2-D DCT block followed by a quantizer using Chen's algorithm [21], with mirror adders and array multipliers [22] as fundamental building blocks, implemented in a commercial 45nm, 1.2V CMOS process. P_η is characterized at various different voltages through delay based Verilog simulations, with two examples shown in Fig. 5(a). P_η has a few large amplitude errors that have a high probability of occurrence, which follows our assumptions for η but not fully. Only the main 2D-DCT block is subject to voltage overscaling (VOS), and hence is the only block that exhibits errors. The estimator is a reduced precision version of the main block. Figure 5(b) shows the simulation result at $V_{dd} = 1V$. Simulations show $\tau^{\star}_{s,p} = 55$, while $\tau^{\star}_{s,m} = 57$. Using (10), $\tau^{\star}_{a,p} = 65$. We can see that even though P_η does not satisfy the assumptions in Section 3, the values are similar. In this case as well, the detection probability exhibits a very flat behavior around $\tau^{\star}_{p}$ and shows that ANT is robust to the value of the detection threshold.

5. CONCLUSION

In this paper we have provided a statistical analysis of ANT, to aid the design of error resilient DSP systems.

6. REFERENCES

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

[2] Y. Lee and Y. Altunbasak, "A collaborative multiple description transform coding and statistical error concealment method for error resilient video streaming over noisy channels," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, 2002, pp. II–2077.

[3] A. Black, H. Zen, and K. Tokuda, "Statistical parametric speech synthesis," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, 2007, pp. IV–1229.

[4] J. Choi, B. Shim, A. Singer, and N. Cho, "Low-power filtering via minimum power soft error cancellation," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 5084–5096, 2007.

[5] R. Hegde and N. R. Shanbhag, "Soft digital signal processing," *IEEE Trans. VLSI Syst.*, vol. 9, no. 6, pp. 813–823, Dec. 2001.

[6] N. R. Shanbhag, R. A. Abdallah, R. Kumar, and D. L. Jones, "Stochastic computation," in *Proc. 47th Design Automation Conf. (DAC)*, 2010, pp. 859–864.

[7] H. Poor, *An Introduction to Signal Detection and Estimation*. New York, NY: Springer-Verlag, 1994.