Cost-Performance Tradeoffs in Fusing Unreliable Computational Units

Mehmet A. Donmez† Maxim Raginsky‡ Andrew C. Singer§ Lav R. Varshney¶

May 23, 2017

Abstract

We investigate fusing several unreliable computational units that perform the same task. We model an unreliable computational outcome as an additive perturbation to its error-free result in terms of its fidelity and cost. We analyze performance of repetition-based strategies that distribute cost across several unreliable units and fuse their outcomes. When the cost is a convex function of fidelity, the optimal repetition-based strategy in terms of incurred cost while achieving a target mean-square error (MSE) performance may fuse several computational units. For concave and linear costs, a single more reliable unit incurs lower cost compared to fusion of several lower cost and less reliable units while achieving the same MSE performance. We show how our results give insight into problems from theoretical neuroscience, circuits, and crowdsourcing.

1 Introduction

We consider the problem of fusing outcomes of several unreliable computational units that perform the same computation under cost and fidelity constraints. We formalize the relationship between the fidelity of each unit and the cost associated with it, and explore this tradeoff in a number of practical problems. Consider, for instance, the capacity of an additive white Gaussian noise (AWGN) channel, which is a logarithmic function of the signal-to-noise (SNR) ratio. In this scenario, the capacity can be increased at the expense of requiring a higher SNR, which introduces a tradeoff between cost (SNR) and performance (rate). Note also that the Fisher information in estimation is often a linear function of SNR, leading to a different cost-performance tradeoff [2].

Building reliable systems out of unreliable components has attracted substantial interest in circuits and systems [3][5], information theory [6][8], and signal processing [9]. In [3], Von Neumann investigated error in logic circuits from a statistical point of view and demonstrated that repeated
computations followed by majority logic may yield reliable results even when the underlying components are unreliable. In [4], Tryon introduced a technique called quadded logic, which corrects errors by a redundant design of logic gates. Moreover, the authors of [6–8] investigated reliable computation by formulas in the presence of noise. More recently, the authors of [9] considered energy-reliability tradeoffs in computing linear transforms implemented on unreliable components.

Fusion of the outputs collected from several sensors has been considered in distributed detection, estimation, classification, and optimization in sensor networks [10–16]. Often, spatially distributed sensors locally perform a decision-making task and send their outputs, under bandwidth constraints, to a fusion center that forms a final decision. In most practical applications, these sensors are battery-powered devices with limited accuracy and computational capabilities, so their performance is critically affected by the resources allocated to them, introducing a cost-performance tradeoff. The authors of [15] studied tradeoffs between the number of sensors, resolution of quantization at each sensor, and SNR. Similarly, [16] considered the tradeoff between reliability and efficiency in distributed source coding for field-gathering sensor networks. In general, the main goal is to make a reliable final decision in a cost-efficient manner based on these unreliable sensors subject to resource and reliability constraints.

A fundamental question that arises in fusing several unreliable computational units is how a limited budget should be allocated across several unreliable units, where adding a new unit incurs a baseline cost as well as an incremental cost, and also increases the cost of fusion. That is, what is the optimal approach in terms of cost-performance tradeoff? Although existing work in fault-tolerant computing and in-sensor networks focus on different pieces of this problem, a more general treatment that jointly considers cost and performance is necessary. This paper is an attempt to combine insights from both fields into a unified framework that captures characteristics of a range of problems. In particular, we show how our framework and results are connected to problems from neuroscience, circuits, and crowdsourcing in Section 5.

In this paper, we present an abstract framework to explore the fundamental tradeoff between cost and performance achievable through forms of redundancy. We model unreliability in any computational unit as an additive random perturbation, where the variance of the perturbation is inversely related to its fidelity. We cast the main task as inference of the error-free computation based on noisy computational outcomes. Each computational unit incurs some cost, which is a function of its fidelity, that includes a baseline cost incurred simply to operate the unit.

We define a class of repetition-based strategies, where each strategy distributes the total cost across several unreliable computational units and fuses their outputs. We note that the fusion operation also incurs some cost, which is a function of the number of individual computational units to be fused. We measure the inference performance of each strategy in terms of MSE between its final output and the error-free computation.

We consider optimal repetition-based strategies under convex, linear, and concave cost functions rather than restricting to specific cost functions. For convex costs, there are two main cases. In the first case, we prove that using only a single and more reliable computational outcome is more cost-efficient than the fusion of several lower cost but less reliable computational outcomes. In the second case, however, we demonstrate that the optimal strategy uses several computational outcomes instead of a single more reliable one. Intuitively, the convexity of the cost function disperses the cost across several less reliable computational outcomes with smaller individual costs. For linear or concave costs, the optimal strategy is to use a single and more reliable computational outcome.
2 Problem Description

Consider the problem of fusing outcomes of several unreliable computational units subject to cost and fidelity constraints. We first introduce a model of an unreliable computational outcome as an additive perturbation to its error-free result in terms of its fidelity and cost. We next consider a class of repetition-based strategies that distribute cost across several parallel unreliable units and fuse their outcomes to produce a final estimate of the error-free computation.

Suppose a vector of input signals $X = (X_1, \ldots, X_k)$ is processed to yield the error-free computation,

$$Y = f(X),$$

where $f(\cdot)$ is some arbitrary target function. Instead, we observe an unreliable computational outcome,

$$Z_\theta = Y + U_\theta,$$

where $U_\theta$ is a zero-mean perturbation with variance $\theta^{-1}$. Here, $\theta$ is the fidelity of the unreliable computational outcome $Z_\theta$. We assume that $Y$ and $U_\theta$ are uncorrelated, that is, $\mathbb{E}[YU_\theta] = \mathbb{E}[Y]\mathbb{E}[U_\theta]$ holds, whether or not $Y$ is a random variable.

By Chebyshev’s inequality, the unreliable outcome $Z_\theta$ with fidelity $\theta > 0$ satisfies, for any $\varepsilon > 0$,

$$\Pr(|Z_\theta - Y| \geq \varepsilon) \leq \frac{1}{\varepsilon^2 \theta}. \quad (1)$$

This implies the unreliable outcome $Z_\theta$ converges to the error-free computation in probability as the fidelity tends to infinity. However, as the fidelity parameter $\theta$ increases, the cost $C(\theta)$ incurred to guarantee that level of fidelity also increases, introducing a cost-fidelity tradeoff. Note that this holds both when $X_i$ for $i = 1, \ldots, k$, or $Y$, are random as well as when they are purely deterministic.

In this model, we must incur a cost $C(\theta)$ to get the unreliable outcome $Z_\theta$ with fidelity $\theta > 0$, which we assume to be a strictly increasing function of $\theta$. In particular, we assume

$$C(\theta) = C_{\min} + G(\theta),$$

where $C_{\min} \triangleq \inf_{\theta > 0} C(\theta) \geq 0$ is the minimum (baseline) cost, and $G(\theta)$ is an increasing and twice differentiable incremental cost function with $G(0) = 0$. In the sequel, we focus on three classes of cost functions: convex, linear, and concave.

We define a class of repetition-based strategies that fuse the outputs of several computational units to estimate $Y$. For any positive integer $N$, a repetition-based strategy $S_N$, with weights $w = (w_1, \ldots, w_N) \in \mathbb{R}^N$ and fidelities $\theta = (\theta_1, \ldots, \theta_N) \in (0, \infty)^N$, linearly combines the outcomes of $N$ parallel unreliable units with fidelities $\theta$ using the weights $w$. That is, if we denote each unreliable outcome with the fidelity $\theta_i$ and the cost $C(\theta_i)$ as

$$Z_{\theta_i} = Y + U_{\theta_i},$$

for $i = 1, \ldots, N$, then the final output of this strategy $S_N$ is

$$\hat{Y}_N(w; \theta) \triangleq w^T Z_\theta = Y(w^T 1) + w^T U_\theta$$

(2)
where \( Z_\theta \triangleq (Z_{\theta_1}, \ldots, Z_{\theta_N}) \), \( U_\theta \triangleq (U_{\theta_1}, \ldots, U_{\theta_N}) \), and \( \mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^N \) is a vector of ones. In particular, we assume that \( U_{\theta_i} \)'s are uncorrelated with each other.

The cost incurred by the strategy \( S_N \) with fidelities \( \theta \) is

\[
\sum_{i=1}^{N} C(\theta_i) + D(N),
\]

where \( D(N) \) is the fusion cost, i.e., the cost of linear combination. We assume that the function \( D : \mathbb{Z}_+ \to \mathbb{R}_+ \) is increasing, as fusing a larger number of computational units has higher cost than fewer. Note that the fusion cost is super-linear in \( N \) in that it requires at least \( O(N) \) multiplications and additions. In particular, we assume that \( D(N) \) is convex in \( N \).

3 Performance Analysis

Here, we consider the MSE performance of each repetition-based strategy in estimating the error-free computation \( Y \). For any positive integer \( N \), the strategy \( S_N \) with a weight vector \( w \in \mathbb{R}^N \) and a fidelity vector \( \theta \in (0, \infty)^N \) achieves the MSE

\[
\text{MSE}(w, \theta) \triangleq \mathbb{E}\left[ (\hat{Y}_N(w; \theta) - Y)^2 \right].
\]

In particular, we derive the minimum MSE (MMSE) achievable by this strategy \( S_N \) while producing an unbiased output:

\[
\text{MSE}_o(\theta) \triangleq \min_{w \in \mathbb{R}^N} \text{MSE}(w, \theta),
\]

where \( w_o \) is the corresponding minimizer.

**Lemma 1.** Suppose that for any positive integer \( N \), the strategy \( S_N \) fuses the outcomes of \( N \) parallel computational units with fidelities \( \theta \in (0, \infty)^N \). Then the MMSE achievable by this strategy \( S_N \) while producing an unbiased estimate of \( Y \), and the corresponding weights are

\[
\text{MSE}_o(\theta) = \frac{1}{\theta^T \mathbf{1}}, \quad w_o = \frac{\theta}{\theta^T \mathbf{1}},
\]

respectively.

**Proof.** We provide the proof in Appendix A.

Thus, Lemma 1 provides the strategy \( S_N \) achieving the MMSE for a given fidelity vector \( \theta \in (0, \infty)^N \). For any positive integer \( N \), whenever we refer to the strategy \( S_N \), we use the optimal weights given in (4), so that its output is

\[
\hat{Y}_N(w_o; \theta) = w_o^T Z_\theta = \frac{\theta^T Z_\theta}{\theta^T \mathbf{1}}.
\]

We next study a particular scenario, where \( U_{\theta} \) is sub-Gaussian.
3.1 Sub-Gaussian Perturbations

Here, we consider a case where the perturbation $U_\theta$ is sub-Gaussian with parameter $\theta^{-1}$, which means

$$
\mathbb{E}\left[e^{\lambda U_\theta}\right] \leq \exp\left(\frac{\lambda^2}{2\theta}\right), \quad \forall \lambda \in \mathbb{R},
$$

or equivalently, the probability of absolute deviation of $Z_\theta$ from $Y$ satisfies, for any $\varepsilon > 0$,

$$
\Pr(|Z_\theta - Y| \geq \varepsilon) \leq 2 \exp\left(-\varepsilon^2 \theta / 2\right).
$$

The tail bound in (6) decreases faster (with increasing $\theta$) than the bound in (1). Sub-Gaussian distributions can be used to model a wide range of stochastic phenomena including Gaussian and uniform distributions, or distributions with finite or bounded support. Note that a weighted sum of finitely many sub-Gaussian random variables is also sub-Gaussian [17]. By applying this result to the output of a strategy $S_N$ with $w \in \mathbb{R}^N$ and $\theta \in (0, \infty)^N$, we get, for any $\varepsilon > 0$,

$$
\Pr\left(\left|\hat{Y}_N(w; \theta) - Y\right| \geq \varepsilon\right) \leq 2 \exp\left(-\frac{\varepsilon^2}{\sum_{i=1}^N w_i^2 / \theta_i}\right).
$$

The weights minimizing the upper bound under $w^T \mathbf{1} = 1$, and the resulting bound are known to be $w_o = \theta / \theta^T \mathbf{1}$, and

$$
\Pr\left(\left|\hat{Y}_N(w_o; \theta) - Y\right| \geq \varepsilon\right) \leq 2 \exp\left(-\varepsilon^2 \theta^T \mathbf{1} / 2\right),
$$

for any $\varepsilon > 0$, respectively.

We emphasize that, in this case, even though the performance is measured in terms of probability of absolute deviation from the error-free computation, the optimal weights are exactly the same as the ones minimizing the MSE. Hence, same results apply to both cases when comparing the cost-performance tradeoff of the repetition-based strategies.

In this section, we analyzed the MSE performance of repetition-based strategies. More precisely, for any positive integer $N$ and a fidelity vector $\mathbf{\theta} \in (0, \infty)^N$, we derived the optimal weights for the strategy $S_N$ in terms of minimizing the MSE. Based on these results, we next investigate the cost-performance tradeoff for the class of repetition-based strategies under the classes of convex, linear, and concave cost functions.

4 Cost-Performance Tradeoff

We investigate the performance of repetition-based strategies under convex, linear, and concave cost functions in terms of the tradeoff between the total incurred cost and the final MSE performance in estimating the error-free computation.

We first analyze the case where the cost $C(\theta)$ is a convex function of the fidelity $\theta$. We characterize the optimal strategy, based on the desired MSE performance as well as the baseline and fusion cost functions. In particular, we show that the optimal cost-performance tradeoff may be achieved by some strategy $S_N$ with $N > 1$ under certain conditions.
We next consider the case where the cost $C(\theta)$ is a linear function of the fidelity parameter $\theta$, and show that strategy $S_1$ is optimal among repetition-based strategies. We finally study the concave cost scenario, and demonstrate results similar to the linear cost function case.

To compare cost-performance tradeoffs of repetition-based strategies, we constrain each strategy to guarantee the same MSE performance. More precisely, given some $\tau > 0$, we assume that the strategy $S_N$ with $\theta \in (0, \infty)^N$ satisfies

$$\tau = \operatorname{MSE}_o(\theta) = \frac{1}{\theta^T 1},$$

or equivalently, $\tau^{-1} = \theta^T 1$, for any positive integer $N$. We also define the total cost incurred by this strategy $S_N$, which achieves $\operatorname{MSE}_o(\theta) = \tau$, as

$$\operatorname{Cost}_\tau(N) \triangleq \sum_{i=1}^N C(\theta_i) + D(N).$$

### 4.1 Convex Cost Functions

We study the cost-performance tradeoff for the class of repetition-based strategies under a convex cost function. This case corresponds to a law of diminishing returns between cost and fidelity, which may drive the dispersion of cost across several less reliable computational units with smaller individual costs. We show that there are two main cases, where in the first case some strategy $S_N$ with $N > 1$ may incur the minimum total cost achievable by the repetition-based strategies while achieving the same MSE, whereas in the second case the strategy $S_1$ is optimal in terms of cost-performance tradeoff, i.e., no repetition or fusion is required.

Consider a uniform fidelity distribution across several unreliable computational outcomes, given by

$$\theta_i \triangleq \frac{1}{\tau N}, \quad i = 1, \ldots, N, \quad (7)$$

which implies that the constraint $\operatorname{MSE}_o(\theta) = \tau$ is satisfied. In fact, the following lemma shows that the optimal fidelity distribution satisfying the MSE constraint in terms of minimizing the total cost is in fact uniform.

**Lemma 2.** For any $\tau > 0$, the uniform fidelity distribution given by (7) is the unique solution to the optimization problem:

$$\min_{\theta \in \mathbb{R}_+^N} \sum_{i=1}^N C(\theta_i)$$

subject to $\theta^T 1 = \tau^{-1}$ when the cost function $C(\theta)$ is convex.

**Proof.** The proof is given in Appendix B. \[\square\]
Hence, we only consider the case where the strategy \( S_N \), for each positive integer \( N \), uses the fidelities in (7). The total cost incurred by this strategy \( S_N \) is:

\[
\text{Cost}_\tau(N) = \sum_{i=1}^{N} C\left(\frac{1}{\tau N}\right) + D(N) = NG\left(\frac{1}{\tau N}\right) + NC_{\text{min}} + D(N).
\]  

(8)

To investigate the behavior of the total cost, we define its continuous relaxation as

\[
\text{Cost}_\tau : [1, \infty) \to (0, \infty)
\]

\[
\text{Cost}_\tau(a) \triangleq aG\left(\frac{1}{\tau a}\right) + aC_{\text{min}} + D(a),
\]

where \( D(a) \) is a twice differentiable continuous relaxation of the fusion cost function \( D(N) \). We first demonstrate that \( \text{Cost}_\tau(a) \) is a convex function in \( a \).

**Lemma 3.** The total cost function \( \text{Cost}_\tau(a) \) is convex in \( a \).

**Proof.** The proof is provided in Appendix C.

Convexity of \( \text{Cost}_\tau(a) \) implies that it has a unique minimizer on any given compact subset of its domain \([1, \infty)\). In particular, note that \( \text{Cost}_\tau(1) = G(\tau^{-1}) + C_{\text{min}} \), and \( \text{Cost}_\tau(a) \to \infty \) as \( a \to \infty \). Therefore, the total cost function \( \text{Cost}_\tau(a) \) has a unique and finite minimizer \( a_o(\tau) \in [1, \infty) \). Also, there exists a corresponding unique optimal repetition-based strategy, which we denote as the strategy \( S_{N_o(\tau)} \) where

\[
N_o(\tau) = \arg \min_{N \in \left\{\lfloor a_o(\tau)\rfloor, \lceil a_o(\tau)\rceil\right\}} \text{Cost}_\tau(N).
\]  

(9)

is a finite positive integer (a function of \( \tau \)), that minimizes the total incurred cost while achieving the desired MSE of \( \tau > 0 \).

We next characterize conditions under which the optimal repetition-based strategy either uses a single but more reliable computational unit, that is, \( N_o(\tau) = 1 \), or distributes the cost across several unreliable computational units and fuses their outcomes, that is, \( N_o(\tau) > 1 \). In the latter case, we implicitly derive the optimal strategy as a function of the desired MSE level \( \tau \), the baseline cost \( C_{\text{min}} \), and the fusion cost function \( D(\cdot) \). The next theorem characterizes these cases in terms of the first derivative of the fusion cost and the baseline cost.

**Theorem 1.** For any given \( \tau > 0 \), the minimizer of \( \text{Cost}_\tau(a) \) satisfies \( a_o(\tau) > 1 \) if and only if

\[
C_{\text{min}} + D'(1) < V(\tau)
\]

where

\[
V(\tau) \triangleq \tau^{-1}G'(\tau^{-1}) - G'(\tau^{-1}).
\]  

(10)
Proof. We define \( \kappa_r(a) \triangleq \partial \text{Cost}_r(a)/\partial a \), and observe that from Lemma 3, \( \kappa_r(a) \) is nondecreasing and continuous in \( a \) since \( \text{Cost}_r(a) \) is a twice differentiable and convex function of \( a \). Hence, whenever \( \kappa_r(1) \geq 0 \), we have \( \kappa_r(a) \geq 0 \) for any \( a > 1 \). It implies that \( \text{Cost}_r(a) \) is a nondecreasing function of \( a \) on \([1, \infty)\), and minimized at \( a_0(\tau) = 1 \). When \( \kappa_r(1) < 0 \), \( \text{Cost}_r(a) \) is minimized at some finite \( a_0(\tau) > 1 \), since \( \text{Cost}_r(a) \to \infty \) as \( a \to \infty \). The proof follows by noting that

\[
\kappa_r(1) = G(\tau^{-1}) - \tau^{-1}G'(\tau^{-1}) + C_{\min} + D'(a) < 0
\]

if and only if \( C_{\min} + D'(1) < V(\tau) \), where \( V(\tau) \) is defined in (10).

Based on these results, we can characterize the optimal repetition-based strategy. If \( C_{\min} + D'(1) \geq V(\tau) \), then \( N_o(\tau) = 1 \) since \( a_0(\tau) = 1 \). Otherwise, we get \( a_0(\tau) > 1 \), which is in this case implicitly given by

\[
\left. \frac{\partial \text{Cost}_r(a)}{\partial a} \right|_{a=a_0(\tau)} = G\left(\frac{1}{\tau a_0(\tau)}\right) - \frac{1}{\tau a_0(\tau)}G'\left(\frac{1}{\tau a_0(\tau)}\right) + C_{\min} + D'(a_0(\tau)) = 0. \tag{11}
\]

If \( 1 < a_0(\tau) < 2 \), then we may get \( N_o(\tau) = 1 \) or \( N_o(\tau) = 2 \), based on (9). When \( a_0(\tau) \geq 2 \), we get \( N_o(\tau) > 1 \).

We finally consider the optimal repetition-based strategy as the target MSE \( \tau \) changes. In the following lemma, we investigate the function \( V(\tau) \) defined in (10) as \( \tau \) changes.

Lemma 4. \( V(\tau) \) is nonnegative and nonincreasing on \((0, \infty)\), and in particular, we have \( \lim_{\tau \to \infty} V(\tau) = 0 \), and

\[
L \triangleq \lim_{\tau \to 0} V(\tau) > 0, \tag{12}
\]

if \( V(\tau) \) is bounded as \( \tau \to 0 \), or else, the limit does not exist.

Proof. We present the proof in Appendix D.

It may appear that from (9) and (11), as the target MSE \( \tau \) decreases, the optimal repetition-based strategy may need to fuse more units, i.e., \( N_o(\tau) \) may increase. More rigorously, we next characterize the behavior of the minimizer \( a_0(\tau) \) of the total cost \( \text{Cost}_r(a) \) as the target MSE \( \tau \) changes.

Theorem 2. If the limit in (12) exists, and \( L \leq C_{\min} + D'(1) \), then \( a_0(\tau) = 1 \) for all \( \tau > 0 \). If, on the other hand, the limit does not exist, or it exists and \( L > C_{\min} + D'(1) \), we define

\[
T \triangleq \inf V^{-1}(C_{\min} + D'(1)) > 0,
\]

where \( V^{-1}(x) \) is the inverse image of a point \( x \) under the function \( V \) for any \( x > 0 \). Then we get \( a_0(\tau) = 1 \) whenever \( \tau \geq T \), and \( a_0(\tau) > 1 \) whenever \( 0 < \tau < T \).

Proof. Suppose the limit in (12) exists, and \( L \leq C_{\min} + D'(1) \). Then \( V(\tau) \leq C_{\min} + D'(1) \), and \( a_0(\tau) = 1 \), for all \( \tau > 0 \).

Suppose next the limit in (12) either does not exist, or it exists and \( L > C_{\min} + D'(1) \). Since \( V(\tau) \) is a monotone function, \( V^{-1}(C_{\min} + D'(1)) \) is either a singleton or an interval. Then for any \( \tau \geq T \), we have \( V(\tau) \leq C_{\min} + D'(1) \), which implies \( a_0(\tau) = 1 \), and when \( 0 < \tau < T \), we have \( C_{\min} + D'(1) < V(\tau) \), which implies \( a_0(\tau) > 1 \).
In this section, we investigated the cost-performance tradeoff for repetition-based strategies under convex cost functions. In particular, we characterized the optimal repetition-based strategy in terms of the baseline cost, the behaviors of the incremental and fusion cost functions with different parameters, for different values of the target MSE level $\tau$. We next study the cost-performance tradeoff under linear cost functions.

4.2 Linear Cost Functions

We consider the optimal repetition-based strategy in terms of cost-efficiency when the underlying cost function is linear, where we can express it as

$$C(\theta) = C_{\min} + \alpha \theta, \quad \theta > 0,$$

where $\alpha > 0$ is an application-dependent constant. This case corresponds to a law of proportional returns. We show that the strategy $S_1$ is the optimal repetition-based strategy for any target MSE $\tau > 0$. There is no gain in repetition-based approaches in terms of cost-efficiency for linear cost functions.

**Theorem 3.** Suppose that the cost function $C(\theta)$ is linear, that is, $C(\theta) = C_{\min} + \alpha \theta$ for some $\alpha > 0$. Then the optimal repetition-based strategy in terms of minimizing the incurred cost while achieving the same MSE is the strategy $S_1$.

**Proof.** Let $\tau > 0$ be given. The total cost of the strategy $S_N$, for any positive integer $N$, is given by

$$\text{Cost}_\tau(N) = NC_{\min} + \alpha \sum_{i=1}^{N} \theta_i + D(N) = NC_{\min} + \alpha \tau^{-1} + D(N) > C_{\min} + \alpha \tau^{-1} = \text{Cost}_\tau(1).$$

This implies the cost incurred by the strategy $S_1$ is smaller than that of the strategy $S_N$ for any $N > 1$ and $\tau > 0$. \qed

For proportional costs a single more reliable unit is always more cost-efficient than a fusion of several less reliable units in the sense that it incurs a smaller cost while achieving the same MSE. We next analyze the concave cost function case.

4.3 Concave Cost Functions

We consider the cost-performance tradeoff of each strategy in the class of strategies when the cost function is concave. This case corresponds to a law of increasing returns, as opposed to a law of diminishing returns. That is, the incremental cost for performance decreases, making single, high-cost, high performance elements more attractive. Before proving the main theorem of this section, we present a lemma that proves that the concave incremental cost function is sub-additive.

**Lemma 5.** If a function $f$ with the domain $[0, \infty)$ is concave, and $f(0) \geq 0$, then it is sub-additive, i.e., for any $x, y \geq 0$,

$$f(x) + f(y) \geq f(x + y).$$
Proof. We provide the proof in Appendix E.

The next theorem characterizes the optimal repetition-based strategy in terms of minimizing the total incurred cost while achieving the same MSE performance for a given $\tau > 0$.

**Theorem 4.** Suppose that the cost function $C(\theta)$ is concave, and each repetition-based strategy achieves the same MSE level $\tau > 0$. Then the strategy $S_1$ is always the optimal strategy in terms of incurring the smallest cost for any $\tau > 0$.

**Proof.** Let $\tau > 0$ be given. Then, for any positive integer $N$, the total cost incurred by the strategy $S_N$ is given by

$$\text{Cost}_\tau(N) = \sum_{i=1}^{N} C(\theta_i) + D(N) = \sum_{i=1}^{N} G(\theta_i) + NC_{\min} + D(N).$$

We note that by Lemma 5, the incremental cost function is sub-additive, since it is concave and $G(0) \geq 0$, implying that

$$\sum_{i=1}^{N} G(\theta_i) \geq G\left(\sum_{i=1}^{N} \theta_i\right) = G(\tau^{-1}). \quad (13)$$

Note that the cost incurred by the strategy $S_1$ is given by

$$\text{Cost}_\tau(1) = G(\tau^{-1}) + C_{\min},$$

implying $\text{Cost}_\tau(N) > \text{Cost}_\tau(1)$ for any $N > 1$. Hence, the strategy $S_1$ is the optimal strategy for any desired MSE.

Strategy $S_1$, which is formed by exhausting all available cost for a single computational unit, is more cost-efficient as compared to any strategy $S_N$ with $N > 1$, which allocates available cost across several less reliable computational units.

In this section, we considered the cost-performance tradeoff of repetition-based strategies under convex, linear, and concave cost function classes. We showed that under convex cost functions the optimal cost-performance tradeoff may be achieved either by the strategy $S_1$ or by some strategy $S_N$ with $N > 1$ under certain conditions. For linear and concave costs, optimality is always achieved by strategy $S_1$ for any target MSE performance. In the next section, we consider applications of our results into a number of contexts.

## 5 Applications

Here, we show how our cost-fidelity formulation and theoretical results are connected to problems from different fields.
5.1 Neuroscience

We review a particular application of our framework in a theoretical neuroscience context. We focus on two principal tasks of the brain where synapses play essential roles, namely, information storage and information processing. Typical central synapses exhibit noisy behavior due, for instance, to probabilistic transmitter release. The firing of the presynaptic neuron is inherently stochastic and occasionally fails to evoke an excitatory postsynaptic potential (EPSP). In this sense, we can cast each noisy synapse as an unreliable computational unit, contributing to the overall neural computation carried out by its efferent neuron. We focus on two distinct cost-fidelity formulations, where we show that experimental results \[18, 19\] agree with our theoretical predictions. We note that recall corresponds to a form of “in-memory computing” whereas processing corresponds to a form of “in-sensor computing”.

5.1.1 In-Memory Computing

Revisiting \[18\], we first consider an information-theoretic framework to study the information storage capacity of synapses under resource constraints, where memory is seen as a communication channel subject to several sources of noise. Each synapse has a certain SNR, where increasing the SNR increases the information storage capacity in a logarithmic fashion. However, this increase comes at a cost, namely, the synaptic volume. Hence, from an information storage perspective, we cast capacity as the fidelity of a noisy synapse and the volume as the cost. If we denote the information storage capacity of a synapse and its average volume by \(C_I\) and \(V\), respectively, then taking Shannon’s AWGN channel capacity formula \[20\] for concreteness:

\[
C_I = \frac{1}{2} \ln \left( 1 + \frac{V}{V_N} \right),
\]

where \(V_N\) is the volume of a synapse with a unit SNR. This relationship assumes the power law \((V/V_N) = (A/A_N)^2\), which is supported by experimental measurements \[18\], where \(A\) is the mean EPSP amplitude and \(A_N\) is the noise amplitude. We rewrite the volume as a function of capacity as

\[
V = V_N(e^{2C_I} - 1),
\]

and observe that this is an exponential cost function, a particular example of convex costs. For exponential costs, fusion of several less reliable computational units may lead to better cost-efficiency than a single more reliable computational unit. Therefore, our cost-fidelity framework applied to information recall under resource constraints recovers the principle that several small and noisy synapses should be present in brain regions performing storage and recall, rather than large and isolated synapses \[18, 21\].

Moreover, \[22, 26\] show that the noisiness of the synapses leads to efficient information transmission. That is, transmitting the same information over several less reliable but metabolically cheaper synapses requires less energy, as compared to the case where the information is transmitted over a single, more reliable but metabolically more expensive synapse. The idea that noise can facilitate information transmission is also present in neuronal networks. In particular, the authors in \[27\] show that a neuron is a noise-limited device of restricted bandwidth, and an energy-efficient nervous system will split the information and transmit it over a large number of relatively noisy neurons of lower information capacity.
5.1.2 In-Sensor Computing

We next consider an information processing perspective, and view the SNR of a synapse itself as its fidelity and the synaptic volume as the cost. We adopt a data-driven approach using two different data sets. This joining is necessary since joint electrophysiology and imaging experiments are technically difficult, where electrophysiology experiments to measure voltages require live tissue while electron micrograph imaging experiments to measure volumes require fixing and slicing the tissue [19].

The first data set [18] includes EPSP measurements across 637 distinct synapses over 43 trials for each synapse. Based on these measurements, we generate an empirical distribution of the mean EPSP measurements of a synapse. The second data set [19] includes volume measurements across 357 synapses, which is used to compute a distribution of a synapse volume.

We first generate \( T = 500 \) random variables \( \{Y_t\}_{t=1}^T \) from the calculated volume distribution. We next generate \( T \) random variables from the calculated mean EPSP distribution, and sort them assuming a monotonic relationship between the mean EPSP and the volume of synapses [18]. From the sorted mean EPSP amplitudes, we compute the corresponding SNRs \( \{X_t\}_{t=1}^T \). We plot the resulting pairs \( \{(X_t, Y_t)\}_{t=1}^T \) in Figure 1. This plot indicates that the cost function is approximately concave as a function of SNR. More rigorously, we assess convexity using a nonparametric hypothesis test based on a simplex statistic, a descriptive measure of curvature described in [28]. When applied to this data, the test yields a \( p \)-value of \( 3.25 \times 10^{-4} \), which can be interpreted as a strong evidence in favor of the hypothesis that the cost (volume) is a concave function of the SNR (fidelity). This suggests that the brain may achieve cost-efficiency by using a single large and reliable synapse, instead of several smaller and less reliable synapses, from an information processing perspective.

To compare this prediction with experimental findings, we focus on a particular synapse called the calyx of Held, the largest synapse in the mammalian auditory central nervous system that connects principal neurons within the auditory system [29,31]. The calyx of Held plays a crucial role in certain information processing tasks of the brain. For instance, the principal cells connected by
The calyx of Held enable interaural level detection, a vital role in high frequency sound localization [32,33]. The signals derived from the calyx of Held generate large excitatory postsynaptic currents with a short synaptic delay, where the transmission speed and fidelity of the calyx is very reliable in mature animals [34].

Hence, the calyx of Held may be regarded as a very reliable but costly synapse, as compared to the ones performing information storage tasks, which are noisier and less costly in terms of brain resources. We observe that these experimental findings agree with our prediction that the cost-efficiency results from employing a single reliable and costly synapse (calyx of Held), instead of several less reliable and metabolically cheaper synapses, under a concave cost function.

5.2 Circuits

Next, let us consider signal processing systems implemented on unreliable circuit fabrics. As CMOS technology scales beyond 10 nm, the operation of CMOS devices begins to suffer from static defects as well as dynamic operational non-determinism [35,37]. Moreover, spintronics, which use electron spin for computing, exhibit an unreliable behavior, where there is a tradeoff between reliability and energy consumption [38,39]. That is, probability of failure is smaller when more energy is used. Hence, deeply scaled CMOS and spintronics based systems must operate under computational errors.

In [3], von Neumann studied noise in circuits and showed that even when circuit components are unreliable, reliable computations can be performed by using repetition-based schemes. Repeated computations followed by a majority vote have also been used extensively in error-tolerant circuit design [40,41]. Also, Hadjicostis [42] investigated redundancy-based approaches to build fault-tolerant dynamical systems out of cheap but unreliable components.

Moreover, a statistical error compensation technique called Algorithmic Noise Tolerance (ANT) has been studied in [43,44]. ANT compensates for errors in computation in a statistical manner.
by fusing outcomes of several unreliable computational branches that operate at different points along energy-reliability tradeoffs. The ANT framework can also be cast as a CEO problem in multiterminal source coding [45].

Stochastic behavior in circuit fabrics may arise when computation is embedded into either memory, which leads to in-memory computing [46], or sensing, which leads to in-sensor computing [47], to achieve cost-efficiency [48]. Note that in-memory computing and in-sensor computing may lead to fundamentally different cost-performance tradeoffs. In particular, we demonstrate that the difference between in-memory computing and in-sensor computing may be modeled through our framework by using different cost-fidelity function classes.

5.2.1 Example case

Here, we present an application of the results of this section into spintronics. In particular, exponential cost has been shown to approximately model the functional dependence between energy and reliability for a typical spin device [39]. Consider the exponential cost

\[ C(\theta) = C_{\text{min}} + \alpha \left( e^{\beta \theta} - 1 \right), \quad \theta > 0, \]  

(14)

for some \( \alpha, \beta > 0 \). Moreover, for illustration purposes, we assume that the fusion cost function is \( D(N) = \gamma (N - 1) \), for \( N \geq 1 \) and \( \gamma > 0 \). Then the total cost function is given by

\[ \text{Cost}_\tau(N) = \alpha N \left( e^{\frac{\beta}{\tau N}} - 1 \right) + N(C_{\text{min}} + \gamma) - \gamma, \]  

(15)

for any positive integer \( N \). In Fig. 2, we plot this total cost function with parameters \( \alpha = 1, \beta = 1, \gamma = 1, C_{\text{min}} = 7 \) for different values of the target MSE \( \tau > 0 \). We observe that Fig. 2 illustrates how \( N_0(\tau) \) increases as \( \tau \) decreases, as discussed in this section. In particular, we note that \( N_0(\tau) = 1, 6, 13 \) for \( \tau = 2, 0.1, 0.05 \), respectively.

Finally, the total cost function (15) yields

\[ V(\tau) = \alpha \exp\left( \frac{\beta}{\tau \gamma} \right) \left( \frac{\beta}{\tau \gamma} - 1 \right) + \alpha, \]  

(16)

implying \( V(\tau) \to \infty \) as \( \tau \to 0 \). Hence there exists a threshold

\[ T = V^{-1}(C_{\text{min}} + \gamma) > 0 \]

such that \( a_o(\tau) = 1 \) when \( \tau \geq T \), and \( a_o(\tau) > 1 \) when \( \tau < T \). These cases are illustrated in Fig. 3 for \( C_{\text{min}} = 7, \gamma = 1 \).

5.3 Crowdsourcing

Crowdsourcing assigns a task to a large number of less expensive but unreliable workers, instead of a small number of more expensive and reliable experts. Monetary payment to incentivize workers has been shown to affect the quality and the quantity of work in such scenarios [49]. Recently, motivated by reliability issues of crowdsourced workers and limited budgets, several researchers have pursued the limits of achievable performance from estimation-theoretic [49], information-theoretic [50], optimization [51,52], and empirical [53] perspectives.

The authors of [53] studied the relation between monetary incentives and work quality in a knowledge task. More precisely, they performed an experiment on 451 unique workers on Amazon
Figure 3: The function $V(\tau)$ [16] to illustrate the optimal strategy regions.

Mechanical Turk, and investigated the effect of bonus payments on the work quality in the task of proofreading an article. They measured the quality by the number of typographical errors found in a given article. In this scenario, each worker is paid a base salary (minimum cost), and an additional bonus (incremental cost), which is shown to yield an improvement in the work quality. In this sense, the bonus payment, i.e., the incremental cost, can be viewed as a function of the number of errors found. In particular, experiments in [53] showed that increasing the bonus payment has diminishing returns in terms of the work quality. That is, the incremental cost is a convex function of the work quality.

More recently, Lahouti and Hassibi [50] considered the crowdsourcing problem as a human-based computation problem where the main task is inference. They formulated an information-theoretic framework, where unreliable workers are modeled as parallel noisy communication channels. They represented the queries of the workers and the final inference using a joint source channel encoding/decoding scheme. Similarly, Khetan and Oh [52] studied the tradeoff between budget and accuracy in crowdsourcing scenarios under the generalized Dawid-Skene model, where they introduced an adaptive scheme to allocate a budget across unreliable workers.

We observe that there is a tradeoff between cost (monetary payments, bonus) and fidelity (quality of work) in a wide range of crowdsourcing scenarios. In particular, assigning a task to several workers, distributing the limited budget among them, and fusing their unreliable outputs have been problems of interest in the crowdsourcing literature. In this sense, our cost-fidelity formulation and repetition-based approaches may have relevance in crowdsourcing problems.

6 Conclusion and Future Directions

We considered fusing outcomes of several unreliable computational units that perform the same task. We modeled unreliability in a computational outcome using an additive perturbation, where the fidelity is inversely related to the variance of the perturbation. We investigated cost-performance tradeoffs achievable through repetition-based approaches. Here, each computational unit incurs a
baseline cost as well as an incremental cost, which is a function of its fidelity.

We defined a class of repetition-based strategies, where any repetition-based strategy distributes the cost across several unreliable computational units and fuses their outcomes to produce a final output, where it incurs cost to perform the fusion operation. We considered the MSE of each strategy in estimating the error-free computation. In particular, we defined the optimal repetition-based strategy as the one incurring the smallest cost while achieving the desired MSE performance.

When the cost is a convex function of fidelity, the optimal repetition-based strategy may distribute cost across several less reliable computational units instead of using a single more reliable unit under certain conditions. For the classes of concave and linear cost functions we preserved that the optimal strategy uses only a single and relatively reliable computational unit, instead of a fusion of several less costly but less reliable units.

We assumed that outcomes produced by different computational units are uncorrelated. This framework can be extended to a correlated outcome setting. When studying the fundamental tradeoff between cost and performance, we assumed that the fusion operation is error-free. We can extend this to the case where the fusion operation also produces noisy results under cost and fidelity constraints. Moreover, we focused on a particular fusion operation, i.e., linear combination, which is common in certain applications. More generally, we can consider nonlinear fusion rules to compute the final estimate of the error-free computation. For instance, midrange \[49\] and median-of-means \[54\] estimators have been considered as alternatives to linear estimators under different scenarios to improve performance. Extension of this setup would be of interest for different network topologies, as opposed to the centralized fusion setting of this paper, as in \[55\].

Acknowledgements

We thank Dmitri B. Chklovskii for providing data from \[19\].

A Proof of Lemma 1

The MSE of the strategy \(S_N\) with a given \(\theta \in (0, \infty)^N\) is

\[
\text{MSE}(w, \theta) = \mathbb{E} \left[ (Y (w^T 1 - 1) + w^T U_\theta)^2 \right],
\]

where (2) is substituted in (3). Since \(Y\) and \(U_\theta\) are uncorrelated:

\[
\text{MSE}(w, \theta) = \mathbb{E} [Y^2] (w^T 1 - 1)^2 + w^T \Sigma_{U_\theta} w,
\]

(A.1)

where \(\Sigma_{U_\theta}\) is the covariance matrix of the perturbation vector \(U_\theta\). If we impose the condition that \(w^T 1 = 1\) in (A.1), then

\[
\text{MSE}(w, \theta) = w^T \Sigma_{U_\theta} w.
\]

To minimize this over weights that satisfy \(w^T 1 = 1\), we first form the Lagrangian

\[
J(w, \lambda) = \frac{1}{2} w^T \Sigma_{U_\theta} w + \lambda (1 - w^T 1),
\]

16
and then compute the gradient with respect to $w$ to get

$$\Sigma_{U\theta} w - \lambda \mathbf{1} = 0,$$

which is satisfied iff $w = \lambda \Sigma_{U\theta}^{-1} \mathbf{1}$. With $w^T \mathbf{1} = 1$, it yields

$$\lambda = \frac{1}{1^T \Sigma_{U\theta}^{-1} \mathbf{1}},$$

which yields the optimal weights

$$w_o = \frac{1}{1^T \Sigma_{U\theta}^{-1} \Sigma_{U\theta}^{-1} \mathbf{1}}.$$

Finally, when substituted in $\text{MSE}(w, \theta)$, we achieve

$$\text{MSE}_o(\theta) = w_o^T \Sigma_{U\theta} w_o = \frac{1}{1^T \Sigma_{U\theta}^{-1} \mathbf{1}}.$$

The proof follows by noting $\Sigma_{U\theta} = \text{diag}(\theta_1^{-1}, \ldots, \theta_N^{-1})$.

**B Proof of Lemma 2**

We solve this optimization problem using the method of Lagrange multipliers, where we first form the Lagrangian

$$J(\theta_1, \ldots, \theta_N, \lambda) \triangleq \sum_{i=1}^{N} C(\theta_i) + \lambda \left( \tau^{-1} - \sum_{i=1}^{N} \theta_i \right).$$

Then, we set the derivative of the Lagrangian with respect to $\theta_j$ to 0, which is given by

$$\frac{\partial J}{\partial \theta_j}(\theta_1, \ldots, \theta_N, \lambda) = C'(\theta_j) - \lambda = 0,$$

for each $j = 1, \ldots, N$. Hence the necessary conditions for optimality are given by $\lambda = C'(\theta_j)$ for $j = 1, \ldots, N$.

Here, we note that the cost function $C(\theta)$ is convex and strictly increasing in $\theta$, and its derivative $C'(\theta)$ is nondecreasing. This implies that it is invertible, so we can write

$$\theta_j = (C')^{-1}(\lambda),$$

for each $j = 1, \ldots, N$, where $(C')^{-1}$ is the inverse of the function $C'$. That is, $\theta_1 = \cdots = \theta_N$. Moreover, by imposing the MSE constraint, we get $\theta_j = (\tau N)^{-1}$ for any $j = 1, \ldots, N$, which yields the desired result.
C Proof of Lemma 3

We first differentiate the total cost function as

$$\frac{\partial \text{Cost}_\tau(a)}{\partial a} = G\left(\frac{1}{\tau a}\right) - \frac{1}{\tau a} G'\left(\frac{1}{\tau a}\right) + C_{\min} + D'(a).$$

We next find its second derivative as

$$\frac{\partial^2 \text{Cost}_\tau(a)}{\partial a^2} = -\frac{1}{\tau a^2} G''\left(\frac{1}{\tau a}\right) + \frac{1}{\tau^2 a^3} G''\left(\frac{1}{\tau a}\right) + \frac{1}{\tau a^3} G''\left(\frac{1}{\tau a}\right) + D''(a),$$

which is nonnegative since the incremental cost function $G(\cdot)$ and the fusion cost function $D(\cdot)$ are both convex and $a > 0$.

D Proof of Lemma 4

We first observe that from (10)

$$V'(\tau) = -\frac{1}{\tau^2} G''(\tau^{-1}) - \frac{1}{\tau} G''(\tau^{-1}) + \frac{1}{\tau^2} G'(\tau^{-1})$$

$$= -\frac{1}{\tau^3} G''(\tau^{-1}) \leq 0,$$

for any $\tau > 0$, as $G(\cdot)$ is convex and twice differentiable. Thus, the function $V(\tau)$ is decreasing on $(0, \infty)$. We next note that

$$\lim_{\tau \to \infty} V(\tau) = \lim_{\tau \to \infty} (\tau^{-1} G'(\tau^{-1}) - G'(\tau^{-1}))$$

$$= \lim_{\tau \to \infty} \tau^{-1} G'(\tau^{-1}) - G(0) = 0,$$

since $G(0) = 0$ and $G'(0)$ is finite. Therefore, $V(\tau)$ is nonnegative on $(0, \infty)$. This implies that the function $V(\tau)$ either converges to a finite limit (if and only if $V(\tau)$ is bounded on $(0, \infty)$), or is unbounded as $\tau \to 0$.

E Proof of Lemma 5

Suppose that $\lambda \in [0, 1]$. Since $f$ is concave, we have

$$f(\lambda x) = f(\lambda x + (1 - \lambda)0) \geq \lambda f(x) + (1 - \lambda) f(0) \geq \lambda f(x).$$
Then, for any $x, y > 0$, we can write

$$f(x) + f(y) = f\left((x + y) \frac{x}{x+y}\right) + f\left((x + y) \frac{y}{x+y}\right),$$

$$\geq \frac{x}{x+y} f(x + y) + \frac{y}{x+y} f(x + y)$$

$$= f(x + y),$$

where we use $x/(x+y), y/(x+y) \in [0, 1]$.

References

[1] Mehmet A. Donmez, Maxim Raginsky, Andrew C. Singer, and Lav R. Varshney. Cost-performance tradeoffs in unreliable computation architectures. In Conf. Rec. 50th Asilomar Conf. Signals, Syst. Comput., pages 215–219, November 2016.

[2] Steven M. Kay. Fundamentals of Statistical Signal Processing: Estimation Theory. Prentice Hall, Upper Saddle River, NJ, USA, 2010.

[3] J. von Neumann. Probabilistic logics and the synthesis of reliable organisms from unreliable components. Automata Studies, 34:43–98, 1956.

[4] J. G. Tryon. Quadded logic. In Richard H. Wilcox and William C. Mann, editors, Redundancy Techniques for Computing Systems, pages 205–228. Spartan Books, Washington, 1962.

[5] Shmuel Winograd and Jack D. Cowan. Reliable Computation in the Presence of Noise. MIT Press, Boston, MA, USA, 1963.

[6] Nicholas Pippenger. Reliable computation by formulas in the presence of noise. IEEE Trans. Inf. Theory, 34(2):194–197, March 1988.

[7] B. Hajek and T. Weller. On the maximum tolerable noise for reliable computation by formulas. IEEE Trans. Inf. Theory, 37(2):388–391, March 1991.

[8] William Evans and Nicholas Pippenger. On the maximum tolerable noise for reliable computation by formulas. IEEE Trans. Inf. Theory, 44(3):1299–1305, May 1998.

[9] Y. Yang, P. Grover, and S. Kar. Computing linear transforms with unreliable components. In Proc. 2016 IEEE Int. Symp. Inf. Theory, pages 1934–1938, July 2016.

[10] R. Viswanathan and P. K. Varshney. Distributed detection with multiple sensors part I. fundamentals. Proc. IEEE, 85(1):54–63, January 1997.

[11] Prakash Ishwar, Rohit Puri, Kannan Ramchandran, and S. Sandeep Pradhan. On rate-constrained distributed estimation in unreliable sensor networks. IEEE J. Sel. Areas Commun., 23(4):765–775, April 2005.

[12] Alejandro Ribeiro and Georgios B. Giannakis. Bandwidth-constrained distributed estimation for wireless sensor networks—part I: Gaussian case. IEEE Trans. Signal Process., 54(3):1131–1143, March 2006.
[13] Alejandro Ribeiro and Georgios B. Giannakis. Bandwidth-constrained distributed estimation for wireless sensor networks—part II: Unknown probability density function. IEEE Trans. Signal Process., 54(7):1131–1143, July 2006.

[14] Sergio Barbarossa, Stefania Sardellitti, and Paolo Di Lorenzo. Distributed detection and estimation in wireless sensor networks. arXiv prep., 2013.

[15] S. A. Aldosari and J. M. F. Moura. Fusion in sensor networks with communication constraints. In Proc. 3rd Int. Symp. Inf. Processing Sensor Netw. (IPSN’04), pages 108–115, April 2004.

[16] D. Marco and D. L. Neuhoff. Reliability vs. efficiency in distributed source coding for field-gathering sensor networks. In Proc. 3rd Int. Symp. Inf. Processing Sensor Netw. (IPSN’04), pages 161–168, April 2004.

[17] Stephane Boucheron, Gabor Lugosi, and Pascal Massart. Concentration Inequalities. Oxford University Press, Oxford, UK, 2013.

[18] Lav R. Varshney, Per Jesper Sjöström, and Dimitri B. Chklovskii. Optimal information storage in noisy synapses under resource constraints. Neuron, 52(3):409–423, November 2006.

[19] Yurii Mishchenko, Tao Hu, Josef Spacek, John Mendenhall, Kristen M. Harris, and Dmitri B. Chklovskii. Ultrastructural analysis of hippocampal neuropil from the connectomics perspective. Neuron, 67(6):1009–1020, September 2010.

[20] Claude Shannon. A mathematical theory of communication. Bell Syst. Tech. J., 27:379–423, July 1948.

[21] Nicolas Brunel, Vincent Hakim, Philippe Isope, Jean-Pierre Nadal, and Boris Barbour. Optimal information storage and distribution of synaptic weights: perceptron versus Purkinje cells. Neuron, 43(5):745–757, September 2004.

[22] Simon B. Laughlin, Rob R. de Ruyter van Steveninck, and John C. Anderson. The metabolic cost of neural information. Nat. Neurosci., 1(1):36–41, May 1998.

[23] Anthony Zador. Impact of synaptic unreliability on the information transmitted by spiking neurons. J. Neurophysiol., 79(3):1219–1229, March 1998.

[24] A. Manwani and C. Koch. Detecting and estimating signals over noisy and unreliable synapses: Information-theoretic analysis. Neural Comput., 13(1):1–33, January 2001.

[25] William B. Levy and Robert A. Baxter. Energy-efficient neuronal computation via quantal synaptic failures. J. Neurosci., 22(11):4746–4755, June 2002.

[26] M. S. Goldman. Enhancement of information transmission efficiency by synaptic failures. Neural Comput., 16(6):1137–1162, June 2004.

[27] Simon B. Laughlin and Terrence J. Sejnowski. Communication in neuronal networks. Science, 301(5641):1870–1874, September 2003.

[28] Jason Abrevaya and Wei Jiang. A nonparametric approach to measuring and testing curvature. J. Bus. Econ. Stat., 23(1):1–19, September 2005.
[29] D. K. Morest. The collateral system of the medial nucleus of the trapezoid body of the cat, its neuronal architecture and relation to the olivocochlear bundle. *Brain Res.*, 9(2):288–311, July 1968.

[30] P. H. Smith, P. X. Joris, L. H. Carney, and T. C. Yin. Projections of physiologically characterized globular bushy cell axons from the cochlear nucleus of the cat. *J. Comp. Neurol.*, 304(3):387–407, February 1991.

[31] Joachim Hermann. *Information Processing at the Calyx of Held Synapse Under Natural Conditions*. PhD thesis, Ludwig Maximilian University of Munich, January 2008.

[32] K. M. Spangler, W. B. Warr, and C. K. Henkel. The projections of principal cells of the medial nucleus of the trapezoid body in the cat. *J. Comp. Neurol.*, 283(3):249–262, August 1985.

[33] C. Tsuchitani. Input from the medial nucleus of trapezoid body to an interaural level detector. *Hearing Res.*, 105(1-2):211–224, March 1997.

[34] K. Futai, M. Okada, K. Matsuyama, and T. Takahashi. High-fidelity transmission acquired via a developmental decrease in NMDA receptor expression at an auditory synapse. *J. Neurosci.*, 21(10):3342–3349, May 2001.

[35] H. P. Wong, D. J. Frank, P. M. Solomon, C. H. J. Wann, and J. J. Welser. Nanoscale CMOS. *Proc. IEEE*, 87(4):537–570, April 1999.

[36] L. Wang and N. R. Shanbhag. Low-power filtering via adaptive error-cancellation. *IEEE Trans. Signal Process.*, 51(2):575–583, February 2003.

[37] J. W. Choi, B. Shim, A. C. Singer, and N. I. Cho. Low-power filtering via minimum power soft error cancellation. *IEEE Trans. Signal Process.*, 55(10):5084–5096, October 2007.

[38] W. H. Butler, T. Mewes, C. K. A. Mewes, P. B. Visscher, W. H. Ripppard, S. E. Russek, and R. Heindl. Switching distributions for perpendicular spin-torque devices within the macrospin approximation. *IEEE Trans. Magn.*, 48(12):4684–4700, December 2012.

[39] A. Patil, S. Manipatruni, D. E. Nikonov, I. Young, and N. Shanbhag. Enabling spin logic via Shannon-inspired statistical computing. In *13th Joint MMM-Intermag Conf.*, January 2016.

[40] R. A. Abdallah and N. R. Shanbhag. An energy-efficient ECG processor in 45-nm CMOS using statistical error compensation. *IEEE J. Solid-State Circuits*, 48(11):2882–2893, November 2013.

[41] Jie Han, Eugene Leung, Leibo Liu, and Fabrizio Lombardi. A fault-tolerant technique using quadded logic and quadded transistors. *IEEE Trans. VLSI Syst.*, 23(8):1562–1566, August 2015.

[42] Chris N. Hadjicostis. *Coding Approaches to Fault Tolerance in Dynamic Systems*. PhD thesis, EECS Department, Massachusetts Institute of Technology, Cambridge, MA, August 1999.

[43] R. Hegde and N. R. Shanbhag. Soft digital signal processing. *IEEE Trans. VLSI Syst.*, 9(6):813–823, December 2001.
[44] Eric P. Kim and Naresh R. Shanbhag. Statistical analysis of algorithmic noise tolerance. In Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2013), pages 2731–2735, May 2013.

[45] Daewon Seo and Lav R. Varshney. Information-theoretic limits of algorithmic noise tolerance. In Proc. IEEE Int. Conf. on Reboot. Comput. (ICRC), pages 1–4, November 2016.

[46] M. Kang, M. S. Keel, N. R. Shanbhag, S. Eilert, and K. Curewitz. An energy-efficient VLSI architecture for pattern recognition via deep embedding of computation in SRAM. In Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2014), pages 8323–8330, May 2014.

[47] Y. Hu, W. Rieutort-Louis, J. Sanz-Robinson, K. Song, J. C. Sturm, S. Wagner, and N. Verma. High-resolution sensing sheet for structural-health monitoring via scalable interfacing of flexible electronics with high-performance ICs. In Proc. 2012 Symp. VLSI Circuits, pages 120–121, June 2012.

[48] Naresh R. Shanbhag. Energy-efficient machine learning in silicon: A communications-inspired approach. arXiv prep., 2016.

[49] Song Jianhan, Vei Wang Isaac Phua, and Lav R. Varshney. Distributed estimation via paid crowd work. In Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2016), pages 6200–6204, March 2016.

[50] Farshad Lahouti and Babak Hassibi. Fundamental limits of budget-fidelity trade-off in label crowdsourcing. In Proc. 29th Annu. Conf. Neural Inf. Process. Syst. (NIPS), pages 5058–5066, December 2016.

[51] David R. Karger, Sewoong Oh, and Devavrat Shah. Budget-optimal task allocation for reliable crowdsourcing systems. Oper. Res., 62(1):1–24, February 2014.

[52] Ashish Khetan and Sewoong Oh. Achieving budget-optimality with adaptive schemes in crowdsourcing. In Proc. 29th Annu. Conf. Neural Inf. Process. Syst. (NIPS), pages 4844–4852, December 2016.

[53] Chien-Ju Ho, Aleksandrs Slivkins, Siddharth Suri, and Jennifer Wortman Vaughan. Incentivizing high quality crowdwork. In Proc. 24nd Int. Conf. World Wide Web (WWW’15), pages 419–429, May 2015.

[54] Luc Devroye, Matthieu Lerasle, Gabor Lugosi, and Roberto I. Oliveira. Sub-Gaussian mean estimators. arXiv prep., 2015.

[55] Aolin Xu and Maxim Raginsky. Information-theoretic lower bounds for distributed function computation. arXiv prep., 2015.