# Binary Recursive Estimation on Noisy Hardware

Elsa Dupraz† and Lav R. Varshney‡
† IMT Atlantique, Lab-STICC, UBL
‡ University of Illinois at Urbana-Champaign

***Abstract*—Recursive estimation is a basic operation in statistical inference that may be implemented and deployed on faulty hardware with error rates governed by energy consumption. We analyze the loss in estimation performance due to noise in recursive probability computation for the binary case, and develop an optimal energy allocation strategy. Simulations show the validity of analytical bounds.**

## I. Introduction

Recursive estimation is a basic technique for prediction of hidden Markov models, implemented in circuits and deployed in countless application areas including target tracking, speech, and image processing [1]. This consists of estimating a sequence of statistically dependent hidden states from their noisy observations. State estimation can be realized by filtering, which consists of recursively calculating the state posterior probabilities given the sequence of observations. The binary alphabet case [2]–[5] is specifically relevant in settings such as neural prosthetics [6], where there are very strong limitations on circuit size and power consumption.

Indeed, a key resource limit in future nanoscale computing systems is energy. Lowering this energy consumption is necessary to both increase battery life and to reduce environmental footprint, and can be achieved by reducing power supply. However, due to basic thermodynamic reasons, further reducing power of computational units will make them unreliable and introduce transient faults at a probability governed by the power usage [7]. In this paper, we consider recursive estimation algorithms implemented on such unreliable hardware.

The general problem of linear (Boolean) computation on faulty hardware has been studied in a variety of works, e.g. [8]–[11]. The aim in noisy linear computation is to (almost) produce the exact value of the noiseless function output, making use of appropriate circuit redundancy. Unlike general noisy linear computation, there are many information processing and artificial intelligence problems that are naturally robust to errors introduced by noisy hardware. Examples include hypothesis testing and parameter estimation [12], logistic regression [13], and general belief propagation [14]. Noisy LDPC decoders have specifically been widely investigated [15]–[18]. Note that despite their intrinsic robustness, some of these algorithms are not optimal even on fault-free hardware.

Here we consider recursive binary estimation under computation noise, and investigate intrinsic robustness. We assume filtering is realized on faulty hardware that introduces errors in the recursive probability computation. The level of noise in the computation depends on the hardware energy consumption through the chip power supply. In this context, we aim to optimize the filtering estimation performance while limiting the hardware energy consumption.

We first develop a theoretical analysis of binary recursive estimation performance, realized on faulty hardware. This analysis provides bounds on the expected gap between the noisy and noiseless filtering recursion, and establishes convergence of this expected gap for a large number of observations. This analysis allow us to predict the loss in estimation performance due to the noise in the recursion. As a second step, we introduce an energy scheduling strategy to allocate different energy from observation to observation in the recursive computation. Based on the theoretical analysis, we optimize the filtering estimation performance under a given energy constraint, finding that uniform allocation is best.

The paper is organized as follows. Sec. II describes the noiseless and noisy filtering problems. Sec. III presents a theoretical performance analysis of noisy filtering. Sec. IV introduces the energy scheduling strategy. Sec. V gives simulation results. Sec. VI concludes.

## II. Filtering problem

We first introduce the hidden Markov model with binary states and observations [1], and then describe the noiseless and noisy filtering recursions to estimate hidden states from the observations. For the noisy case, we also provide noise models for recursive computation.

### A. Signal model

Denote by $\{S_k\}_k$ a first-order stationary Markov process, where the random variables $S_k$ are called the hidden states. The random variables $S_k$ take values in $\{0,1\}$ and we denote state transition probabilities as

$$\alpha = P(S_k = 1 | S_{k-1} = 0), \tag{1}$$

$$\beta = P(S_k = 0 | S_{k-1} = 1). \tag{2}$$

In general, $\alpha \neq \beta$. We assume without loss of generality that $1 - \alpha - \beta \geq 0$. In addition, let $\{X_k\}_k$ be the sequence of binary observations. The random variables $X_k$ take values in $\{0,1\}$ and for all $i,j \in \{0,1\}$, we denote the observation probabilities as

$$P(i,j) = P(X_k = j | S_k = i).$$

We want to estimate the sequence of hidden states $S_1, \dots, S_k$, from the sequence of observations $X_1, \dots, X_k$

This work was supported in part by the "Make our Planet Great Again" Initiative of the Thomas Jefferson Fund and by grant ANR-17-CE40-0020 of the French National Research Agency ANR (project EF-FECtive).

by filtering. Filtering is based on the the recursive computation of the successive state posterior probabilities $P(S_k = s|x_1, \dots, x_k)$. We assume this recursive computation is implemented on faulty hardware, which introduces noise in the computation operations. Next we describe standard noiseless filtering and then introduce noisy filtering.

### B. Noiseless filtering

Filtering consists of producing an estimate $\hat{s}_k$ of the state $S_k$ from the sequence of observations $(x_1, \dots, x_k)$ as [2]

$$\hat{s}_k = \arg\max_{s\in\{0,1\}} P(S_k = s|x_1, \dots, x_k). \tag{3}$$

The probability $P(S_k = s|x_1, \dots, x_k)$ represents the posterior probability of state $S_k$ given the sequence of observations $(x_1, \dots, x_k)$ available at time $k$.

Since the hidden states $S_k$ are binary, estimation problem (3) can be restated in terms of the log-likelihood ratio (LLR) defined as $L_k = \log \frac{P(S_k=1|x_1,\dots,x_k)}{P(S_k=0|x_1,\dots,x_k)}$. The estimation rule then becomes $\hat{s}_k = 1$ if $L_k > 0$, and 0 otherwise. It is shown in [2] that the successive LLRs $L_k$ can be calculated recursively as $L_0 = 0$ and for all $k \geq 1$,

$$L_k = \log \frac{P(1, x_k)}{P(0, x_k)} + h(L_{k-1}), \tag{4}$$

where the function $h : \mathbb{R} \to \mathbb{R}$ is defined as

$$h(y) = \log \frac{\alpha + e^y(1-\beta)}{(1-\alpha) + e^y\beta}. \tag{5}$$

Standard works [2]–[6] assume the computation of LLRs is perfect and that the state decision is taken on the exact values of $L_k$ given by (4) . Here, we consider that recursion (4) is realized on faulty hardware, which introduces errors in the computation of the succesive $L_k$.

### C. Noisy filtering

If (4) is computed on unreliable hardware, we only observe noisy versions $\widetilde{L}_k$ of the LLRs $L_k$. We assume additive noise:

$$\widetilde{L}_k = \log \frac{P(1, x_k)}{P(0, x_k)} + h(\widetilde{L}_{k-1}) + B_k, \tag{6}$$

where $B_k$ is a continuous random variable. We assume that the successive $B_k$ are i.i.d. and that $E[|B_k|] = \overline{B} < \infty$. Without loss of generality, we assume the noise is unbiased in the sense that $E[B_k] = 0$. A wide range of physical noise models satisfy these simplifying assumptions (additive model and i.i.d.), cf. [10].

As a result of the unbiased assumption, the decisions on the estimated states $\hat{s}_k$ are still taken from the rule $\hat{s}_k = 1$ if $\tilde{L}_k > 0$, and 0 otherwise. At the end, due to the recursive LLR computation, there is a risk that the successive noise $B_k$ accumulate, and that the noisy recursion $\tilde{L}_k$ diverges from $L_k$. This is why the theoretical analysis we propose studies the convergence of the expected gap $E[|\tilde{L}_k - L_k|]$ between the noisy and the noiseless recursion.

## III. Performance Analysis of Noisy Filtering

We develop a theoretical analysis that gives lower and upper bounds on the expected gap $E[|\widetilde{L}_k - L_k|]$ between the noiseless LLR $L_k$ and its noisy version $\widetilde{L}_k$ for all $k$. We then study convergence and fixed points of the sequence of upper bounds.

### A. Bounds on the expected gap $E[|\widetilde{L}_k - L_k|]$

The recursive expression for noiseless $L_k$ depends on the function $h$ in (5). Hence, before giving bounds on the expected gap $E[|\widetilde{L}_k - L_k|]$, we first bound this function.

*Proposition 1:* For $b \in \mathbb{R}$, function $h$ in (5) satisfies

$$-S(|b|) \leq h(y+b) - h(y) \leq S(|b|) \tag{7}$$

where the function $S : \mathbb{R} \to \mathbb{R}$ is defined as

$$S(b) = \log\left(1 + \frac{(e^b-1)\mu}{2ve^{\frac{1}{2}b} + \alpha\beta e^b + (1-\alpha)(1-\beta)}\right) \tag{8}$$

and $\mu := 1 - \alpha - \beta$, $v := \sqrt{\alpha\beta(1-\alpha)(1-\beta)}$. The function $S$ is convex and for all $b \in \mathbb{R}$, $S(|b|) \geq 0$.

*Proof:* Assume $b \geq 0$. From the expression of $h$ in (5):

$$h(y+b) - h(y) \tag{9}$$
$$= \log\left(1 + \frac{(e^b-1)\mu}{\alpha(1-\alpha)e^{-y} + (1-\alpha)(1-\beta) + \alpha\beta e^b + e^{y+b}\beta(1-\beta)}\right),$$
$$h(y) - h(y-b) \tag{10}$$
$$= \log\left(1 + \frac{(1-e^{-b})\mu}{\alpha(1-\alpha)e^{-y} + (1-\alpha)(1-\beta)e^{-b} + \alpha\beta + e^{y-b}\beta(1-\beta)}\right).$$

By function differentiation, we then show that the function $f : y \mapsto \alpha(1-\alpha)e^{-y} + e^{y+b}\beta(1-\beta)$ achieves its minimum in $y_+^\star = \frac{1}{2}\log\frac{\alpha(1-\alpha)}{\beta(1-\beta)} - \frac{b}{2}$ and we have $f(y_+^\star) = 2ve^{\frac{b}{2}}$. Again by function differentiation, the function $g : y \mapsto \alpha(1-\alpha)e^{-y} + e^{y-b}\beta(1-\beta)$ achieves its minimum in $y_-^\star = \frac{1}{2}\log\frac{\alpha(1-\alpha)}{\beta(1-\beta)} + \frac{b}{2}$ and we have $g(y_-^\star) = 2ve^{-\frac{b}{2}}$. In addition, we have that $\lim_{y\to+\infty} f(y) = \lim_{y\to-\infty} f(y) = +\infty$ and $\lim_{y\to+\infty} g(y) = \lim_{y\to-\infty} g(y) = +\infty$. At the end, injecting $f(y_+^\star)$ into (9) and $g(y_-^\star)$ into (10) and considering the limit condition shows that for all $b \geq 0$,

$$0 \leq h(y+b) - h(y) \leq S(b), \tag{11}$$
$$0 \leq h(y) - h(y-b) \leq S(b). \tag{12}$$

Eqs. (11) and (12) hold for $b \geq 0$. Changing $b$ into $-b$ in (12) gives the result for all $b \leq 0$, $-S(-b) \leq h(y+b) - h(y) \leq 0$, which concludes the proof. That $S$ is convex is direct. ∎

Prop. 1 gives lower and upper bounds on the difference $h(y+b) - h(y)$. These bounds do not depend on the value of $y$, but only on the value of $b$. Further, both bounds are defined by the function $S$ given in (8). This function $S$ appears in the upper bound on the expected gap $E[|\widetilde{L}_k - L_k|]$, as follows.

*Proposition 2:* Consider the noiseless LLR recursion in (4), and the noisy LLR recursion in (6). For any $k > 0$,

$$\overline{B} \leq E[|\widetilde{L}_k - L_k|] \leq A_k, \tag{13}$$

where $A_1 = \overline{B}$ and for all $k > 1$,

$$A_k = S(A_{k-1}) + \overline{B}. \tag{14}$$

*Proof:* For all $k \geq 1$, let $\log \frac{P(1,x_k)}{P(0,x_k)} = \mathcal{L}_0(x_k)$. First, consider the recursion $A'_1 = |B_1|$ and for all $k' > 1$, $A'_k = S(A'_{k-1}) + |B_k|$. Since $S(|b|) > 0$, it follows by induction that for all $k \geq 1$, $A'_k \geq 0$. We first show by induction that $|\widetilde{L}_k - L_k| \leq A'_k$.

Since $L_1 = \mathcal{L}_0(x_1)$ and $\widetilde{L}_1 = \mathcal{L}_0(x_1) + B_1$, we have $-|B_1| \leq \widetilde{L}_1 - L_1 \leq |B_1|$, which is the base case. Then, assume that

$$-A'_k \leq \widetilde{L}_k - L_k \leq A'_k. \tag{15}$$

We want to show this inequality holds for $\widetilde{L}_{k+1} - L_{k+1}$. Since $L_k = \mathcal{L}_0(x_k) + h(L_k)$ and $\widetilde{L}_{k+1} = \mathcal{L}_0(x_k) + h(\tilde{L}_k) + B_k$, we have that $\widetilde{L}_{k+1} - L_{k+1} = h(\tilde{L}_k) - h(L_k) + B_k$. Then, according to (15) and since $h$ is increasing,

$$\begin{aligned} &h(L_k - A_k) - h(L_k) + B_k \\ &\quad \leq \widetilde{L}_{k+1} - L_{k+1} \leq h(L_k + A_k) - h(L_k) + B_k, \\ &-S(A_k) - |B_k| \leq \widetilde{L}_{k+1} - L_{k+1} \leq S(A_k) + |B_k| \end{aligned}$$

where the second set of bounds are from Prop. 1. This shows $-A'_{k+1} \leq \widetilde{L}_{k+1} - L_{k+1} \leq A'_{k+1}$ and so $|\widetilde{L}_k - L_k| \leq A'_k$.

Then (13) and (14) follow from the above recursion on $|L'_k - L_k|$ and from the fact $S$ is convex. ∎

In Prop. 2, the expected gap $E[|\tilde{L}_k - L_k|]$ is upper bounded by a value $A_k$ whose recursive expression is (14), determined by function $S$ in (8) and parameter $\overline{B}$. From this expression, we study convergence and fixed points of $A_k$, in order to study the asymptotic behavior of the expected gap $E[|\widetilde{L}_k - L_k|]$. This will verify whether this expected gap diverges due to error accumulation caused by the successive $B_k$.

### B. Asymptotic behavior of $A_k$

Consider convergence and fixed points of the sequence $A_k$.

*Proposition 3:* Sequence $(A_k)_k$ is positive, increasing, and converges to a unique fixed point $A^\star$. The fixed point $A^\star = 2\log(X^\star)$, where $X^\star$ is the unique real positive solution of

$$\begin{aligned} 0 = \alpha\beta e^{-\overline{B}} X^4 + 2ve^{-\overline{B}} X^3 - 2vX - \alpha\beta \\ + (1-\alpha)(1-\beta)(e^{-\overline{B}} - 1)X^2. \end{aligned} \tag{16}$$

*Proof:* We first show $A_k$ is positive and increasing. Since according to Prop. 1, $S(b) \geq 0\ \forall b \geq 0$, it follows that $\forall k \geq 0$, $A_k \geq 0$. The proof $(A_k)_k$ is increasing is by induction. First, since $A_1 = \overline{B} \geq 0$ and $A_2 = S(A_1) + \overline{B}$, we have $A_2 - A_1 \geq S(A_1) \geq 0$. As a result, $A_2 \geq A_1$. Second, assume $A_k \geq A_{k-1}$. We have $A_{k+1} - A_k = S(A_k) - S(A_{k-1})$. Since $S$ is increasing and since $A_k \geq A_{k-1}$, we have $A_{k+1} - A_k \geq 0$, which implies $(A_k)_k$ is increasing.

We now show $A_k$ has a unique fixed point. First, compute:

$$S'(x) = \frac{\mu\left(e^x((1-\alpha)(1-\beta)+\alpha\beta)+ve^{\frac{3}{2}x}+ve^{\frac{1}{2}x}\right)}{\left(2ve^{\frac{1}{2}x}+\alpha\beta e^x+(1-\alpha)(1-\beta)\right)\left(2ve^{\frac{1}{2}x}+\alpha\beta+(1-\alpha)(1-\beta)e^x\right)}.$$

We see that the terms in $e^x$, $e^{\frac{1}{2}x}$, $e^{\frac{3}{2}x}$ in the numerator of $S'$ are strictly less than the terms in $e^x$, $e^{\frac{1}{2}x}$, $e^{\frac{3}{2}x}$, respectively, in the denominator of $S'$. This point and the fact $\mu \leq 1$ imply that $S'(x) < 1$. As a result, the function $R(x) = S(x) + \overline{B}$ is a contractor. Thus the sequence $A_k$ has a unique fixed point $A^\star$ that is the unique solution of the equation $A^\star = S(A^\star) + \overline{B}$. Solving this equation is equivalent to finding the unique real positive solution of (16). ∎

Prop. 3 shows the sequence $A_k$ converges to a fixed point $A^\star$ that can be calculated from (16). As a result, whatever the value of $k$, the expected gap $E[|\tilde{L}_k - L_k|]$ is upper bounded by $A^\star$. Thus, despite the successive noise $B_k$, the noisy recursion $\tilde{L}_k$ does not diverge from the noiseless recursion $L_k$. Moreover, simulation results in Sec. V show the fixed point $A^\star$ is actually close to the parameter $\overline{B}$.

The value of fixed point $A^\star$ can be used as a criterion for optimizing the performance of the noisy filtering recursion, since it bounds the expected gap between noiseless and noisy recursions. Although $A^\star$ can be calculated numerically, for easier optimization an explicit upper bound is as follows.

*Proposition 4:* Denote $\lambda = \max_{x\in\mathbb{R}} S'(x)$. Then, the fixed point $A^\star$ is bounded by $\overline{B} \leq A^\star \leq \frac{\overline{B}}{1-\lambda}$.

*Proof:* Since $S'(x) \leq \lambda < 1$ (see proof of Prop. 3), we have $A_{k+1} = S(A_k) + \overline{B} \leq \lambda A_k + \overline{B}$. By induction, we have that $A_{k+1} \leq \frac{1-\lambda^{k+1}}{1-\lambda}\overline{B}$. To finish, the sequence $\left(\frac{1-\lambda^{k+1}}{1-\lambda}\overline{B}\right)$ converges to $\overline{B}/(1-\lambda)$ when $k \to \infty$. ∎

Based on this analysis, we now want to optimize the filtering performance under given energy constraints. For this, we introduce an energy scheduling framework.

## IV. Energy Scheduling

Theoretical analysis in Sec. III shows the upper bound on the expected gap $E[|\widetilde{L}_k - L_k|]$ is increasing with $k$ and converges to a fixed point. In Sec. III the noise parameter $E[|B_k|]$ was assumed to be fixed with $k$. We now allow this parameter to vary, depending on the energy level allocated at time $k$. To do so, we first develop a physical model to relate the energy level at time $k$ to the noise parameter $E[|B_k|]$.

### A. Energy and noise model

The noise model in Sec. II-C is generic. Here, we consider a specific relationship between energy and noise, to demonstrate hardware energy optimization for the noisy LLR recursion $\tilde{L}_k$. There are two sources of noise: LLR quantization and noise from the memory for successive quantized values of $\tilde{L}_k$.

To implement LLR recursion on-chip, we assume LLR values $\tilde{L}_k$ are uniformly quantized to a large number of $q$ bits. As a result, the quantization noise can be assumed independent of the LLR value $\tilde{L}_k$ [19]. To incorporate the quantization noise into the noise model, we only need to determine $E[|L_k - Q(L_k)|]$, where $Q(L_k)$ is the quantized value of $L_k$. Letting $B_Q$ be the quantization noise, $E[|B_Q|] = E[|L_k - Q(L_k)|] = \frac{\mu}{4}$, where $\mu$ is the quantizer bin width.

We further consider noise coming from storing successive values $\tilde{L}_k$ into memory; memory noise is applied to the binary representation of the quantized LLR values. We consider a thermodynamic CMOS energy-error model [10] to relate error probability $\epsilon$ of a bit to the memory energy consumed $e$ as:

$$\epsilon = \delta_0 \exp(-ce^{\beta}), \tag{17}$$

where $e$ is the cell energy consumption and $\delta_0, c$, and $\beta$ are technology-dependent constants.

Consider a sign-magnitude representation for quantized LLR values, and denote the error probability on the $j$th bit ($j = 0$ is the bit sign) at time $k$ as $\epsilon_{j,k}$. According to (17), $\epsilon_{j,k} = \delta_0 exp(-ce_{j,k}^{\beta})$, where $e_{j,k}$ is the energy level for the $j$th bit at time $k$. We also consider a sign-preserving error model, in which the error probability on the bit sign satisfies $\epsilon_{0,k} = 0$. Such a model can be obtained by proper circuit design [17]. Under the sign-preserving assumption, the noise from the memory is independent of LLR values $L_k$, and so: $E[|B_M|] = \sum_{j=1}^{q} \epsilon_{j,k} 2^{j-1}$, where $B_M$ is the memory noise.

Finally we have $E[|B_Q + B_M|] \leq E[|B_Q|] + E[|B_M|]$. In our analysis, we therefore consider a random variable $B_k$ such that

$$E[|B_k|] = \frac{\mu}{4} + \sum_{j=1}^{q-1} \epsilon_{j,k} 2^{j-1}. \tag{18}$$

In the following, we always assume the number of bits $q$ is fixed and we consider two different settings. First, that at a given time $k$, all bits have the same energy allocation. In this case, we want to optimize energy allocation from time to time. Second, that the energy allocation is fixed in time but can vary from bit to bit. In this case, we want to optimize the energy level for each bit.

### B. Energy allocation over time

Since we assume all bits have the same energy, we denote $e_{j,k} = e_k$ and omit the bit index $j$. To optimize the energy scheduling, we assume the LLR values $\tilde{L}_k$ are divided into blocks of length $K$. The $t$th block includes LLR values $\tilde{L}_{K(t-1)+1}, \ldots, \tilde{L}_{Kt}$ and we denote $\tilde{L}_{Kt+k} = \tilde{L}_k^{(t)}$. Consider $K$ energy levels $e_1, \ldots, e_K$ such that $\sum_{k=1}^{K} e_k \leq E$, where $E$ is the block energy constraint, and assume that energy level $e_k$ is used to store the LLR values $\tilde{L}_k^{(t)}$ for all $t$. This give $K$ error probabilities $\epsilon_k$ and $K$ different parameters $\overline{B}_k$ that can be obtained from the model in Sec. IV-A.

Now, consider the upper bound $A_k^{(t)}$ on the expected gap $E[|\tilde{L}_k^{(t)} - L_k^{(t)}|]$. Since $A_k^{(t)} \leq \lambda A_{k-1}^{(t)} + \overline{B}_k$ (see proof of Prop. 4), we show $A_k^{(t)}$ can be bounded by

$$A_k^{(t)} \leq \lambda^K A_k^{(t-1)} + \sum_{j=k+1}^{K} \lambda^{K+k-j} \overline{B}_j^{(t-1)} + \sum_{j=1}^{k} \lambda^{k-j} \overline{B}_j^{(t)} \tag{19}$$

As a result, the $K$ sequences $(A_k^{\star})_t$ all converge to a different fixed point that is bounded by

$$A_k^{\star} = \frac{1}{1-\lambda^K} \left( \sum_{j=k+1}^{K} \lambda^{K+k-j} \overline{B}_j^{(t-1)} + \sum_{j=1}^{k} \lambda^{k-j} \overline{B}_j^{(t)} \right). \tag{20}$$

Since the $A_k^{\star}$ are all different, the asymptotic behaviors of the $K$ sequences of expected gaps $\left( E[|\tilde{L}_k^{(t)} - L_k^{(t)}|] \right)_t$ vary with the symbol position $k$ in a block. Thus to optimize the energy scheduling, we minimize the sum $\sum_{k=1}^{K} A_k^{\star}$ which considers the filtering performance over all $K$ symbols of a block, as follows:

$$\min_{e_1,\ldots,e_K} \sum_{k=1}^{K} A_k^{\star} \ \text{ s.t. } \ \sum_{k=1}^{K} e_k \leq E \ \text{ and } \ e_k \geq 0 \ \forall k. \tag{21}$$

We restate as minimizing the functional $J_1$:

$$J_1(e_1, \cdots, e_K, \gamma) = \sum_{k=1}^{K} A_k^{\star} + \gamma_0 \left( \sum_{k=1}^{K} e_k - E \right) - \sum_{k=1}^{K} \gamma_k e_k. \tag{22}$$

Exponential model (17) implies the optimization problem is convex. From the expressions of the $A_k^{\star}$ in (20), we get $\sum_{k=1}^{K} A_k^{\star} = \frac{1}{1-\lambda} \sum_{k=1}^{K} \overline{B}_k$. As a result, by differentiation of $J_1(e_1, \ldots, e_K, \gamma)$, the optimal energy scheduling is given by uniform energy allocation $e_k = \frac{E}{K}$ for all $k$.

### C. Energy allocation over quantization bits

Now since we assume energy allocation does not vary with time, we denote $e_{j,k} = e_j$ and omit time index $k$. We consider $q-1$ energy levels $e_1, \ldots e_{q-1}$ such that $\sum_{j=1}^{q-1} e_j \leq E$. In this case, the theoretical analysis of Sec. III directly applies, and we consider the optimization problem

$$\min_{e_1,\cdots,e_{q-1}} \sum_{k=1}^{K} A^{\star} \ \text{ s.t. } \ \sum_{j=1}^{q-1} e_j \leq E \ \text{ and } \ e_j \geq 0 \ \forall j. \tag{23}$$

In (23), $A^{\star}$ is the upper bound on the fixed point of the sequence of expected gaps $(E[|\tilde{L}_k - L_k|])_k$. Taking the expression of $A^{\star}$ from Prop. 4, the optimization (23) is convex and can be restated as minimizing the functional $J_2$:

$$J_2(e_1, \ldots, e_{q-1}, \gamma) = A^{\star} + \gamma_0 \left( \sum_{j=1}^{q-1} e_j - E \right) - \sum_{j=1}^{q-1} \gamma_j e_j. \tag{24}$$

By differentiation of the functional $J_2(e_1, \cdots, e_{q-1}, \gamma)$, we show that optimal bit energy allocation is given by

$$e_j = \begin{cases} 0 & \text{if } \frac{c\delta_0 2^{j-1}}{1-\lambda} \leq \nu, \\ \frac{1}{c} \log \frac{c\delta_0 2^{j-1}}{1-\lambda} - \frac{1}{c} \log \nu, & \text{otherwise,} \end{cases} \tag{25}$$

where $\nu \geq 0$ is such that $\sum_{j=1}^{q-1} e_j = E$. Note that energy level $e_j = 0$ leads to $\epsilon_j = \delta_0$.

Here the optimal energy allocation varies from bit to bit, and is given by a water-filling solution [20]. In (25), the condition $\frac{c\delta_0 2^{j-1}}{1-\lambda} \leq \nu$ that leads to $e_j = 0$ depends on the bit position $j$ only through the term $2^{j-1}$. The most significant bits correspond to the highest values $2^{j-1}$ and they are more likely to be allocated an energy level $e_j > 0$.

## V. Simulations

Here we use simulations to verify the accuracy of the lower and upper bounds on the expected gap $E[|\tilde{L}_k - L_k|]$ in Sec. III. We consider two setups for the Markov model in Sec. II-A. In the first, we consider $\alpha = 0.05$, $\beta = 0.9$, and $P(0,1) = P(1,0) = 0.1$. In the second, we consider $\alpha = 0.1$, $\beta = 0.7$, and $P(0,1) = P(1,0) = 0.1$. For both settings, we

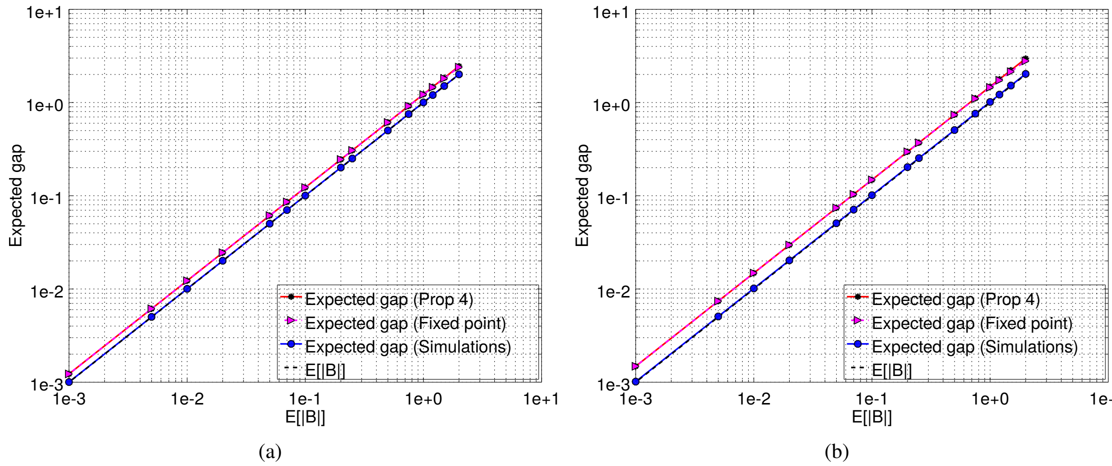


Fig. 1. Expected gap between $L_k$ and $\widetilde{L}_k$, evaluated both from numerical simulations and from the theoretical analysis on two setups: (a) Setup 1, Markov source with parameters $\alpha = 0.05$, $\beta = 0.9$, and $P(0,1) = P(1,0) = 0.1$, (b) Setup 2, $\alpha = 0.1$, $\beta = 0.7$, and $P(0,1) = P(1,0) = 0.1$.

have a Markov chain of length $N = 100000$, a Laplacian distribution for $B$, and various values of $\overline{B} = E[|B_k|]$. For every considered value of $\overline{B}$, we perform 1000 trials, and compute the resulting expected value $E[|\widetilde{L}_k - L_k|]$. We compare the resulting expected value to 1) the lower bound $E[|B_k|]$, 2) The fixed point $A^\star$ from Prop. 3 , and 3) the upper bound $\overline{B}/(1-\lambda)$ on $A^\star$ from Prop. 4.

Results are in Fig. 1. In the two cases, we see that lower and upper bounds are close to each other. We also observe that the upper bound $\overline{B}/(1-\lambda)$ on $A^\star$ coincides with the value of $A^\star$, indicating the accuracy of this bound. To finish, we observe in both cases that the expected value $E[|\widetilde{L}_k - L_k|]$ is between its lower and upper bounds, and is also very close to the lower bound $E[|B_k|]$. Although the upper bound given by the fixed point is pessimistic, we can conclude that Props. 3 and 4 satisfactorily represent the behavior of the expected gap between the noiseless LLR $L_k$ and its noisy version $\widetilde{L}_k$.

## VI. Conclusion

This paper has introduced the problem of recursive estimation on faulty hardware and shown that there is a kind of intrinsic robustness. Further, a fixed point analysis yielding $A^*$ provides a criterion for optimal energy scheduling, which itself turns out to involve waterfilling over bits and identical allocation over time. Various generalizations of this work are possible, but we are particularly interested in characterizing $A^*$ using a thermodynamic entropy production-dissipation argument for recursive estimation [18], [21].

## References


[1] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[2] E. Ordentlich and T. Weissman, "On the optimality of symbol-by-symbol filtering and denoising," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 19–40, Jan. 2006.

[3] A. Critch, "Binary hidden Markov models and varieties," *Journal of Algebraic Statistics*, vol. 4, no. 1, pp. 1–30, 2013.

[4] P. Jacquet, G. Seroussi, and W. Szpankowski, "On the entropy of a hidden Markov process," *Theoretical Computer Science*, vol. 395, no. 2–3, pp. 203–219, May 2008.

[5] J. B. Ruebeck, R. G. James, J. R. Mahoney, and J. P. Crutchfield, "Prediction and generation of binary Markov processes: Can a finite-state fox catch a Markov mouse?" *Chaos*, vol. 28, p. 013109, 2018.

[6] A. C. Smith, L. M. Frank, S. Wirth, M. Yanike, D. Hu, Y. Kubota, A. M. Graybiel, W. A. Suzuki, and E. N. Brown, "Dynamic analysis of learning in behavioral experiments," *J. Neurosci.*, vol. 24, no. 2, pp. 447–461, Jan. 2004.

[7] N. R. Shanbhag, N. Verma, Y. Kim, A. D. Patil, and L. R. Varshney, "Shannon-inspired statistical computing for the nanoscale era," *Proc. IEEE*, vol. 107, no. 1, pp. 90–107, Jan. 2019.

[8] N. Pippenger, "On networks of noisy gates," in *Proc. 26th Annu. Symp. Found. Comput. Sci.*, Oct. 1985, pp. 30–38.

[9] P. Gács and A. Gál, "Lower bounds for the complexity of reliable Boolean circuits with noisy gates," *IEEE Trans. Inf. Theory*, vol. 40, no. 2, pp. 579–583, Mar. 1994.

[10] A. Chatterjee and L. R. Varshney, "Energy-reliability limits in nanoscale circuits," in *Proc. 2016 Inf. Theory Appl. Workshop*, Feb. 2016.

[11] Y. Yang, P. Grover, and S. Kar, "Computing linear transformations with unreliable components," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3729–3756, Jun. 2017.

[12] H. Chen, L. R. Varshney, and P. K. Varshney, "Noise-enhanced information systems," *Proc. IEEE*, vol. 102, no. 10, pp. 1607–1621, Oct. 2014.

[13] Y. Yang, P. Grover, and S. Kar, "Fault-tolerant distributed logistic regression using unreliable components," in *Proc. 54th Annu. Allerton Conf. Commun. Control Comput.*, Sep. 2016, pp. 940–947.

[14] C.-H. Huang, Y. Li, and L. Dolecek, "Belief propagation algorithms on noisy hardware," *IEEE Trans. Commun.*, vol. 63, no. 1, pp. 11–24, Jan. 2015.

[15] L. R. Varshney, "Performance of LDPC codes under faulty iterative decoding," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4427–4444, Jul. 2011.

[16] C.-H. Huang, Y. Li, and L. Dolecek, "Gallager B LDPC decoder with transient and permanent errors," *IEEE Trans. Commun.*, vol. 62, no. 1, pp. 15–28, Jan. 2014.

[17] E. Dupraz, D. Declercq, B. Vasić, and V. Savin, "Analysis and design of finite alphabet iterative decoders robust to faulty hardware," *IEEE Trans. Commun.*, vol. 63, no. 8, pp. 2797–2809, Aug. 2015.

[18] L. R. Varshney, "Toward limits of constructing reliable memories from unreliable components," in *Proc. IEEE Inf. Theory Workshop (ITW'15)*, Oct. 2015, pp. 114–118.

[19] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[20] Y. Kim, M. Kang, L. R. Varshney, and N. R. Shanbhag, "Generalized water-filling for source-aware energy-efficient SRAMs," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4826–4841, Oct. 2018.

[21] S. K. Mitter and N. J. Newton, "Information and entropy flow in the Kalman-Bucy filter," *J. Stat. Phys.*, vol. 118, no. 1-2, pp. 145–176, Jan. 2005.