ISIT'2013 1569715819

"Information-Friction" and its Impact on Minimum Energy Per Communicated Bit

Pulkit Grover
ECE, Carnegie Mellon University
Email: pulkit @ cmu.edu

***Abstract*—Just as there are frictional losses associated with moving masses on a surface, what if there are frictional losses associated with moving information on a substrate? We propose to model these losses as proportional to "bit-meters' *i.e.*, the product of mass of information (*i.e., the number of bits*) and the distance of information transport. For communication across a binary input AWGN channel decoded by decoders implemented using a simple circuit model, we derive unavoidable lower bounds on** bit-meters **for decoding computation. These bounds are translated into limits on energy consumption in decoding under the information-friction model. Using these lower bounds we show that the *total* (transmit + decoding) energy-per-bit must diverge to infinity as the target error probability is lowered.**

I. Introduction

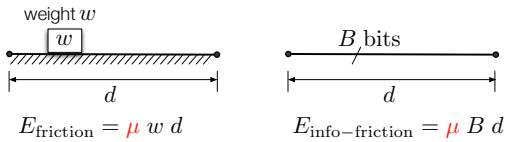

Fig. 1. A Newtonian inspiration for the information-friction model.

[1] Just as there are frictional losses associated with moving masses on a surface, there can be frictional losses associated with moving information between the computational gates (see Fig. 1). Within the context of communication, these frictional losses can be a significant part of the energy consumed in computations at the transmitter and receiver (e.g. encoding and decoding an error-correcting code), which in turn can be a significant fraction of total energy for short-distance communication [3].

What models allow us to account for these frictional losses? Communication complexity, introduced by Andrew Yao in [5], accounts for moving of information on a computational substrate by counting the number of bits that need to be moved. However, for many implementations [2], energy of computation depends not only on the number of bits, but also on the Euclidean distance to which those bits are moved. VLSI complexity, introduced by Thompson in [6], [7], accounts for these distances by measuring the wiring infrastructure required to compute a function, and multiplying it with the number of clock-cycles needed to obtain an understanding of the energy consumption. There are two limitations of the model that we overcome in this paper. First, modern technology is exploring and using interconnects that can be optical, or even wireless [8], and it is important to study models that subsume such implementations. Second, even for metal-interconnects, if one assumes that most of the information is passed along the shorter wires in the "bisection-cut" of the circuit [1], then Thompson's energy-model overestimates the required energy: the bounding technique allows for long wires not being charged and discharged as frequently as the short ones.

In Section II-B, we introduce the "information-friction" model of computation and energy consumption towards addressing these limitations of Thompson's model (see Fig. 1). The model accounts for the number of bits moved and the amount of distance those bits are moved by counting "bit-meters": the product of the number of bits, and the distance this information is moved in order to compute a function on a circuit. A similar "bit-meters" metric was used as a measure of "transport capacity" supported *by* a communication network in the work of Gupta and Kumar [9]. Here, we are interested in the opposite question: what are the bit-meters *needed to support* a computation?

Why is "bit-meters" an appropriate metric for circuit communication energy? There is an intuitive appeal to the metric. If the bits are independent, the metric only increases if subsequent transmissions on a link carry "new" information: exactly when a switching happens on a circuit wire. However, the metric has its shortcomings. In particular, at extremely low speeds of computation, it may be possible to reduce the coefficient of information-friction, consistent with results in thermodynamics of computation [10]. Due to space-constraints, the issue is further discussed in [1]. Thus, while the model is a good approximation to energy requirements in many techniques of circuit implementations [1], studying its limitations could also design energy-efficient codes and decoding architectures that outperform the bounds here, and thereby also the current implementations.

In Section III, we use the implementation model and an AWGN-based hard-decision channel model to derive the bit-meters cost for decoding an error correcting code. Intellectually, our work builds on [2] that uses Thompson's model (and thus suffers from its drawbacks in estimation of energy consumption). Our model of information-friction is simpler than Thompson's and more general in that it does not assume that the circuit has metal-wire interconnects. We show that the required bit-meters for decoding can be no smaller than $\Omega\left(\sqrt{\log \frac{1}{P_e^{blk}}/P_T}\right)$, where P_e^{blk} is the block-error probability, and P_T is the transmit power. Under the information-friction model, optimizing over P_T, we show that the total (transmit + decoding) energy per bit is at least $\Omega\left(\sqrt[3]{\log \frac{1}{P_e^{blk}}}\right)$. Thus, for any implementation consistent with our implementation model that experiences information-frictional losses, the total energy per bit must diverge to infinity as the error probability is driven to zero.

[1]The introduction is shortened to allow space for the definitions and the proofs. The full introduction and proofs will appear in [1]. Interested readers can see of [2]–[4] for relevant literature in circuits and information theory.

II. System model

A. Channel model

We consider a point-to-point communication link. An information sequence of k fair coin flips $\mathbf{b}_1^k$ is encoded into 2^{nR} binary-alphabet codewords $\mathbf{X}_1^n$, hence the rate of the code is $R = \frac{k}{n}$ bits/channel use, which is assumed to be fixed. The codeword $\mathbf{X}_1^n$ is modulated using BPSK modulation and sent through an Additive White Gaussian Noise (AWGN) channel of bandwidth W, with W channel uses per second. The decoder estimates the input sequence $\widehat{\mathbf{b}}_1^k$ by first performing a hard-decision on the received channel symbols before using these hard-decisions $\mathbf{Y}_1^n$ to decode the input sequence. The overall channel $\mathbf{X}_1^n \to \mathbf{Y}_1^n$ is therefore a Binary Symmetric Channel (BSC) with raw bit-error probability $p_{ch} := \mathbb{Q}\left(\sqrt{\frac{\zeta P_T}{\sigma_z^2}}\right)$, where $\mathbb{Q}(x) = \int_x^\infty \frac{1}{2\pi} e^{-\frac{x^2}{2}} dx$, ζ is the path-loss associated with the channel, P_T is the transmit power of the BPSK-modulated signal, and σ_z^2 is the variance of the Gaussian noise in the hard-decision estimation. The encoder-channel-decoder system operates at an average block-error probability P_e^{blk} given by $P_e^{blk} = \Pr\left(\widehat{\mathbf{b}}_1^k \neq \mathbf{b}_1^k\right)$.

Definition 1 (Channel Model (ζ, σ_z^2)): Channel Model (ζ, σ_z^2) denotes (as described above) a BSC(p_{ch}) channel that is a result of hard-decision at the receiver across an AWGN channel of average transmit power P_T, path loss ζ and noise variance σ_z^2.

B. Implementation, computation, and energy models

The computation is performed using a "circuit" on a "substrate." This section formally defines these terms allowing for decoding analysis in Section III.

Definition 2 (Substrate): A Substrate is a square $\mathrm{Sq}(l)$ of side l in $\mathbb{R}^2$ with vertices at $(0,0)$, $(0,l)$, $(l,0)$, and (l,l).

Definition 3 ($\mathrm{Grid}(\lambda)$): $\mathrm{Grid}(\lambda)$ is the intersection of a square lattice with minimum separation between points λ with the substrate $\mathrm{Sq}(l)$.

The parameter λ determines how close computational nodes in the circuit can be brought to each other, and depends on the technology of implementation. For large circuits, $\lambda \ll l$.

Definition 4 (Circuit, computational nodes): The substrate $\mathrm{Sq}(l)$ together with a collection $\mathcal{S} \subset \mathrm{Grid}(\lambda)$ of points (called *computational nodes*, or simply *nodes*) inside $\mathrm{Sq}(l)$, is called a *Circuit*, and is denoted by $\mathrm{Ckt} = (\mathrm{Sq}(l), \mathcal{S})$.

For instance, $\mathrm{Sq}(10\lambda)$ along with the set $\mathcal{S} = \{(\lambda,\lambda), (5\lambda, 4\lambda)\}$ constitutes a Circuit.

Nodes can be *input nodes*, *output nodes*, or *helper nodes*. Input nodes store the input of computation (one bit each; at the beginning of computation), output nodes store the output (one bit each; at the end of computation), and helper nodes help perform the computation.

Definition 5 (Subcircuit): A subcircuit $\mathrm{SubCkt}_1 = (F_1, \mathcal{S}_1)$ of a circuit $\mathrm{Ckt} = (Sq(l), \mathcal{S})$ is constituted by an open and convex subset F_1 of $\mathrm{Sq}(l)$ and by the subset of computational nodes $\mathcal{S}_1 = F_1 \cap \mathcal{S}$.

That is, all the computational nodes within the sub-substrate F_1 must lie in the subcircuit SubCkt_1.

Definition 6 (Link): A (unidirectional) link connects two nodes in that it allows for noiseless communication between nodes in one direction.

In a circuit with n nodes, there are $n(n-1)$ unidirectional links, which can be used more than once during a computation.

Definition 7 (Communication on a circuit): Computational nodes use messages received thus far in computation, and stored memory values, to generate messages that can be (asynchronously) communicated to other nodes over links.

Definition 8 (Computation on a circuit): The computation starts with the arrival of the input of computation at the input nodes. Each input node stores one bit of the input. The computation then proceeds with communication of messages over communication links in a predetermined sequence. Each message is a (potentially randomized) function of the messages (including a part of the input of computation) that the transmitting computational node has received thus far in the computation. Each message is assumed to communicate a predetermined constant number of bits. At the end of the computation, the output is stored in the memories of the output nodes.

A computation may use some or all of the communication links in the circuit. Each link can be used as many times as needed, and at each use, the message can be of any chosen size with the associated costs as described in the following definitions.

Definition 9 (bit-meters cost of a link and of a circuit): The bit-meters cost of a *link* in a computation Comp is the product of the number of bits carried on that link and the Euclidean distance between the nodes at the ends of the link. The bit-meters for the entire circuit Ckt is the sum of bit-meters for all the links in Comp.

Definition 10 (bit-meters for a link within a subcircuit): For a link that connects two nodes within a subcircuit in a computation Comp, the bit-meters for that link *within the subcircuit* is the same as the bit-meters for the link in the original circuit. However, if only one of the nodes lies within the subcircuit, then bit-meters for this link within the subcircuit is the product of the number of bits of the message passed along this link and the length of link from the node inside the subcircuit to the boundary of the subcircuit.

Definition 11 (bit-meters for a subcircuit): The bit-meters for a subcircuit $\mathrm{SubCkt}_1 = (F_1, \mathcal{S}_1)$ in computation Comp is the sum of bit-meters for all the links within the subcircuit (wholly or partially, as defined in Definition 10), and is denoted by bit-meters(SubCkt_1).

The definition also holds for bit-meters for the entire circuit.

Definition 12 (Coefficient of information-friction): The coefficient of information-friction, denoted by μ, characterizes the energy required for computation in our model. This energy is given by $E = \mu \times \mathrm{bm}$, where bm is the bit-meters for executing the given computation on the given circuit.

Definition 13 (Implementation Model (λ, μ)): Implementation Model (λ, μ) denotes the implementation model (as described in Section II-B) with λ minimum distance between computational nodes, and coefficient of information-friction μ.

III. The lower bound on bit-meters and information-friction energy for decoding

To obtain lower bounds on bit-meters for decoding, similar to analysis in [2], [4], [11], we will need to break the decoding circuit into many disjoint subcircuits.

Definition 14 (Disjoint subcircuits): Two subcircuits $\mathrm{SubCkt}_1 = (F_1, \mathcal{S}_1)$ and $\mathrm{SubCkt}_2 = (F_2, \mathcal{S}_2)$ of a circuit $\mathrm{Ckt} = (\mathrm{Sq}(l), \mathcal{S})$ are said to be *disjoint subcircuits* if $F_1 \cap F_2 = \phi$, the null set. Similarly, $\{\mathrm{SubCkt}_i\}_{i=1}^{N_{\mathrm{subckt}}}$ are said to be mutually disjoint subcircuits if $F_i \cap F_j = \phi$ for every $i, j \in \{1, 2, \ldots, N_{\mathrm{subckt}}\}, i \neq j$.

It follows that any two disjoint subcircuits cannot share computational nodes or communication links that connect two nodes *within* one of the subcircuits. In fact, two disjoint subcircuits do not share $\mathrm{bit\text{-}meters}$ of computation:

Lemma 1: Let $\{\mathrm{SubCkt}_i\}_{i=1}^{N_{\mathrm{subckt}}}$, where $\mathrm{SubCkt}_i = (F_i, \mathcal{S}_i)$, be a set of mutually disjoint subcircuits of the circuit $\mathrm{Ckt} = (\mathrm{Sq}(l), \mathcal{S})$. Then for any computation Comp,

$$\text{bit-meters}(\mathrm{Ckt}) \geq \sum_{i=1}^{N_{\mathrm{subckt}}} \text{bit-meters}(\mathrm{SubCkt}_i). \quad (1)$$

Proof: The lemma follows from the observation that in Definition 9, no $\mathrm{bit\text{-}meters}$ are double-counted in disjoint subcircuits. We note that there are potential situations when $\bigcup_{i=1}^{N_{\mathrm{subckt}}} F_i = \mathrm{Sq}(l)$ for which (1) is not satisfied with equality. This happens when there is a long link in a circuit which has a part that does not lie within either of the subcircuits of the two nodes. ■

The decoder circuit is partitioned into multiple subcircuits via a "stencil" that can be moved over the circuit by changing its origin.

Definition 15 (Stencil): A *Stencil*(a, η, O) in $\mathbb{R}^2$ is a pattern of equally spaced "inner" squares that are concentric with "outer" squares which form a grid (as shown in Fig. 2). The length of a side of each outer square is a, and the origin O lies in the center of an "inner" square. The side of each inner square is of length $s = (1 - 2\eta)a$, and the distance between any two adjacent inner squares is $d = 2\eta a$.

A node in a circuit is said to be *covered* by a Stencil that is overlaid on the circuit substrate if it lies inside an inner-square of the Stencil. For the decoder, the n input nodes store the channel observations, and the k output nodes (also called "bit-nodes") store the decoded message bits. Let k_i^{inside} denote the number of bit-nodes that lie inside the *inner* square, and n_i denote the number of input nodes that lie inside the *outer* square, of the i-th subcircuit.

Definition 16 (Stencil-partition): The outer squares of *Stencil*(a, η, O) induce a partition (see Fig. 2) of a circuit into subcircuits, each occupying substrate area a^2. If any computational node lies on the boundary of an outer square, then it is arbitrarily included in one of the subcircuits.

Lemma 2: For any circuit implemented in Implementation Model (λ, μ), for any $\eta > 0$, there exists an origin O of *Stencil*(a, η, O) such that the number of bit-nodes covered by the stencil is lower bounded by

$$\sum_i k_i^{\mathrm{inside}} \geq k(1 - 2\eta)^2. \quad (2)$$

Proof: The proof uses the probabilistic method [12]. Let $O \sim \mathbb{U}\{[0, a), [0, a)\}$, that is, uniformly distributed in the square formed by $(0, 0), (0, a), (a, a), (a, 0)$. Now, the average number of bit-nodes covered by the stencil (averaged over O) is:

$$\begin{aligned} \mathbb{E}\left[\sum_{i=1}^{k} \mathbb{1}_{\{i \text{ covered}\}}\right] &= \sum_{i=1}^{k} \mathbb{E}\left[\mathbb{1}_{\{i \text{ covered}\}}\right] \\ &= \sum_{i=1}^{k} \Pr(i \text{ covered}) \\ &\overset{(a)}{=} \sum_{i=1}^{k} (1 - 2\eta)^2 = k(1 - 2\eta)^2 \quad (3) \end{aligned}$$

where the key step (a) follows from the observation that for any point, as we move the origin O around uniformly, the fraction of possible origins for which the point is covered by the stencil is the fraction of area occupied by the stencil, which is $(1 - 2\eta)^2$. Thus there exists at least one value of the origin O such that the number of nodes covered is larger than the average. ■

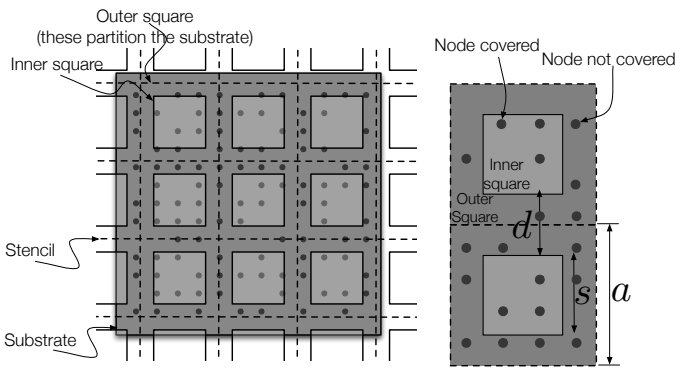

Fig. 2. A Stencil overlaid on the Substrate. Also shown are the computational nodes of the Circuit on the Substrate. A zoomed-in version shows the dimensions of the Stencil. For each square in the zoomed-in version, $k_i^{\mathrm{inside}} = 3$.

Consider the stencil shown in Fig. 2. The distance between the inner and the outer squares is ηa. If B bits are communicated from outside an outer square to inside an inner square in a subcircuit, then, intuitively, the $\mathrm{bit\text{-}meters}$ associated with the subcircuit should be at least $\eta a B$. The following lemma shows this rigorously:

Lemma 3: Consier a circuit implemented in Implementation Model (λ, μ), and any subcircuit SubCkt obtained using the stencil-partition defined in Definition 16. For communicating B bits from outside an outer-square to inside the corresponding inner-square, $\mathrm{bit\text{-}meters} \geq \eta a B$.

Proof: B bits of communication need to pass concentric N_{cut} square-shaped cuts on the circuit-network, starting with the outer square as a cut, with distance λ separating these cuts, as shown in Fig. 3. The cuts end when distance from the inner square is smaller than λ. This distance is denoted by $\alpha\lambda$ for some $\alpha \leq 1$. The inner square is now included as the final N_{cut}-th cut. Except for the inner square, across each cut, each link has to cross at least λ distance. Thus, if the number of bits on a link across the cut is B_{link} bits, then $B_{\mathrm{link}}\lambda$ $\mathrm{bit\text{-}meters}$ cross each cut that the link crosses.

Further, if the number of bits across any cut, which is the summation of B_{link} for links across the cut, is smaller than

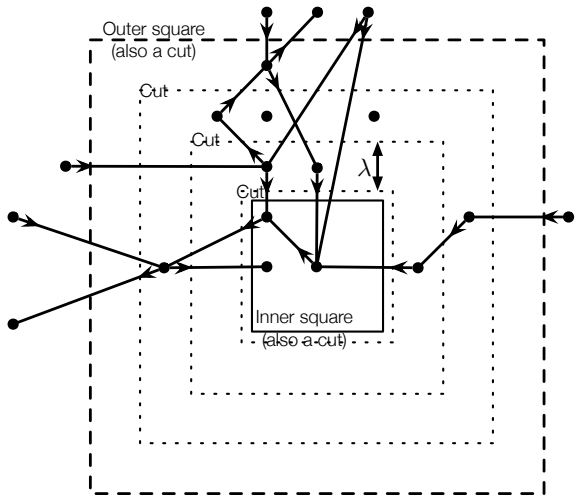

Fig. 3. Square cuts are made in order to use the cut-set bounding technique. The directed edges show the links along which information flows in the computation. However, the links do not indicate the time of information flow during the computation, or the amount of information they carry.

B, then B bits cannot be delivered to the inner square. Thus across each cut, the total number of bit-meters should be at least B. This yields the lemma. ■

Lemma 4: For any subcircuit SubCkt_i obtained via stencil-partitioning of Implementation Model (λ, μ), with $k_i^{\text{inside}} \geq 1$, if $\text{bit-meters}(\mathrm{SubCkt}_i) < \eta a \frac{k_i^{\text{inside}}}{3}$, then $P_e^{blk} \geq \frac{(2p_{ch})^{n_i}}{9}$.

Proof: From Lemma 3, since bit-meters for the subcircuit is smaller than $\eta a \frac{k_i^{\text{inside}}}{3}$, and the distance between the outer square and the inner square is ηa meters, at most $\frac{k_i^{\text{inside}}}{3}$ bits can be delivered from outside the outer square to inside the inner square.

We first observe that a BSC(p_{ch}) is a stochastically degraded version of a BEC($2p_{ch}$). That is, a decoder that receives channel outputs that pass through BEC($2p_{ch}$) can simulate a BSC(p_{ch}) channel by randomly assigning the value 0 or 1 to an erased bit, *i.e.* without any increase in bit-meters. Supplying the decoder with outputs of the erasure channel, we examine the event $\mathcal{E}$ when *all* the n_i channel outputs inside the outer square are erased. This event has probability $(2p_{ch})^{n_i}$.

Conditioning on the erasure event $\mathcal{E}$, let the (block) probability of not recovering all of the bits inside the i-th inner square, denoted by $\vec{b}_i^{in}$, be $P_{e,i}^{\mathcal{E}}$. From Fano's inequality [13] applied to reconstructing the message bits $\vec{b}_i^{in} \in \mathcal{B}_i$, $|\mathcal{B}_i| = 2^{k_i^{\text{inside}}}$, given the communicated message T of entropy at most $k_i^{\text{inside}}/3$ bits,

$$\begin{aligned} & P_{e,i}^{\mathcal{E}} \log(|\mathcal{B}_i| - 1) + h_b(P_{e,i}^{\mathcal{E}}) \\ \geq\ & H(\vec{b}_i^{in}|T) = H(\vec{b}_i^{in}) - H(T) + H(T|\vec{b}_i^{in}) \\ \geq\ & H(\vec{b}_i^{in}) - H(T) \geq k_i^{\text{inside}} - \frac{k_i^{\text{inside}}}{3} = \frac{2k_i^{\text{inside}}}{3}, \end{aligned} \quad (4)$$

where $h_b(\cdot)$ is the binary entropy function. For $k_i^{\text{inside}} = 1$, $|\mathcal{B}_i| = 2$ and $\log(|\mathcal{B}_i| - 1) = 0$, and thus

$$h_b(P_{e,i}^{\mathcal{E}}) \geq \frac{2}{3}. \quad (5)$$

Since $h_b(x) \leq 2\sqrt{x(1-x)} \leq 2\sqrt{x}$ for $x \in (0, 0.5)$ (see, e.g. [14]), $x \geq \frac{(h_b(x))^2}{4}$. From (4), for $k_i^{\text{inside}} = 1$,

$$P_{e,i}^{\mathcal{E}} \geq \frac{(h_b(P_{e,i}^{\mathcal{E}}))^2}{4} \geq \frac{4}{9 \times 4} = \frac{1}{9}. \quad (6)$$

For $k_i^{\text{inside}} \geq 2$, $|\mathcal{B}_i| \geq 4$, and thus using a looser form of (4),

$$\begin{aligned} P_{e,i}^{\mathcal{E}} \log(|\mathcal{B}_i|) + h_b(P_{e,i}^{\mathcal{E}}) &\geq \frac{2k_i^{\text{inside}}}{3} \\ \Rightarrow P_{e,i}^{\mathcal{E}} \log(|\mathcal{B}_i|) + 1 &\geq \frac{2k_i^{\text{inside}}}{3} \\ \Rightarrow P_{e,i}^{\mathcal{E}} \geq \frac{\frac{2k_i^{\text{inside}}}{3} - 1}{\log(|\mathcal{B}_i|)} = \frac{2\frac{k_i^{\text{inside}}}{3} - 1}{k_i^{\text{inside}}} &= \frac{2}{3} - \frac{1}{k_i^{\text{inside}}} \\ \overset{(k_i^{\text{inside}} \geq 2)}{\geq} \frac{2}{3} - \frac{1}{2} &= \frac{1}{6} > \frac{1}{9}. \end{aligned}$$

Thus, for any $k_i \geq 1$, the unconditional error probability for recovering the k_i^{inside} bits correctly is lower bounded by $\frac{(2p_{ch})^{n_i}}{9}$. Since the block error probability P_e^{blk} for the entire code is larger than block error probability in recovering the k_i^{inside} bits in i-th subcircuit, we obtain the lemma. ■

Theorem 1: For decoding an error correcting code transmitted over a channel with Channel Model (ζ, σ_z^2) and decoded in a circuit Ckt in Implementation Model (λ, μ) with block error probability P_e^{blk}, the decoder bit-meters are lower bounded as:

$$\text{bit-meters}(\mathrm{Ckt}) \geq \frac{k}{48\sqrt{2}} \sqrt{\frac{\log \frac{1}{10P_e^{blk}}}{\log \frac{1}{2p_{ch}}}}, \quad (7)$$

$$\text{as long as } \log \frac{1}{10P_e^{blk}} > 10 \log \frac{1}{2p_{ch}}. \quad (8)$$

Proof: First, Appendix A shows that choosing stencil parameter a to be $\frac{1}{\sqrt{2}} \sqrt{\frac{\log \frac{1}{10P_e^{blk}}}{\log \frac{1}{2p_{ch}}}} \lambda$, under condition (8), ensures that $n_i \leq \frac{\log \frac{1}{10P_e^{blk}}}{\log \frac{1}{2p_{ch}}}$. The outer squares of the stencil partition the circuit into subcircuits. Let the i-th subcircuit have n_i channel output nodes available within the *outer* square and k_i^{inside} source bits to decode inside the *inner* square. Using Lemma 2, we choose the origin O of the stencil so that at least $(1 - 2\eta)$ fraction of the k bit-nodes are covered by the *inner* squares, *i.e.*,

$$\sum_i k_i^{\text{inside}} \geq (1 - 2\eta)^2 k. \quad (9)$$

From Lemma 4, if bit-meters for any subcircuit are smaller than $\frac{k_i^{\text{inside}}}{3} \eta a$, then the error probability is lower bounded as

$$P_e^{blk} \geq \frac{(2p_{ch})^{n_i}}{9} \geq \frac{1}{9}(2p_{ch})^{\frac{\log \frac{1}{10P_e^{blk}}}{\log \frac{1}{2p_{ch}}}} = \frac{10}{9} P_e^{blk},$$

a contradiction. Thus, for each subcircuit in the stencil-partition,

$$\text{bit-meters}(\mathrm{SubCkt}_i) \geq \frac{k_i^{\text{inside}} \eta a}{3}.$$

From Lemma 2, $\sum_i k_i^{\text{inside}} \geq (1-2\eta)^2 k$, therefore,

$$\begin{aligned}\sum_{i=1}^{N_{\text{subckt}}} \text{bit-meters}(\text{SubCkt}_i) &\geq \frac{(1-2\eta)^2 k\eta a}{3} \\ &= \frac{(1-2\eta)^2 k\eta}{3\sqrt{2}} \sqrt{\frac{\log \frac{1}{10P_e^{blk}}}{\log \frac{1}{2p_{ch}}}}.\end{aligned}$$

Choosing $\eta = \frac{1}{4}$ yields the theorem. ∎

The following corollary provides lower bounds on energy-per-bit which can easily be translated into bounds on power consumption [1].

Corollary 1 (Unavoidable limits on energy-per-bit): For decoding an error correcting code transmitted over a channel with Channel Model (ζ, σ_z^2) and implemented in Implementation Model (λ, μ), the total energy per bit for communication at error probability P_e^{blk} is lower bounded as:

$$\frac{E_{total}}{k} \geq \Omega\left(\sqrt[3]{\log \frac{1}{P_e^{blk}}}\right). \tag{10}$$

Proof overview: Because the channel is used W times per second, the transmit energy used is $\frac{nP_T}{W}$. The total energy-per-bit under condition (8) can therefore be lower bounded as:

$$\begin{aligned}\frac{E_{total}}{k} &\geq \frac{nP_T}{kW} + \frac{\mu k}{48\sqrt{2}k}\sqrt{\frac{\log \frac{1}{10P_e^{blk}}}{\log \frac{1}{2p_{ch}}}} \\ &= \frac{P_T}{RW} + \frac{\mu}{48\sqrt{2}}\sqrt{\frac{\log \frac{1}{10P_e^{blk}}}{\log \frac{1}{2p_{ch}}}}.\end{aligned}$$

In our hard-decision channel model, the term $\log \frac{1}{2p_{ch}}$ scales proportionally to the received power ζP_T (see, e.g. [2]). Thus

$$\frac{E_{total}}{k} \geq \frac{P_T}{RW} + \frac{\beta}{48\sqrt{2}}\sqrt{\frac{\log \frac{1}{10P_e^{blk}}}{P_T}},$$

for some $\beta > 0$. It is clear that the choice of P_T that minimizes the RHS is $P_T^* = \Theta\left(\sqrt[3]{\log \frac{1}{P_e^{blk}}}\right)$. Substituting,

$$\frac{E_{total}}{k} \geq \Omega\left(\sqrt[3]{\log \frac{1}{P_e^{blk}}}\right). \tag{11}$$

If (8) is not satisfied, then it is easy to show that the required transmit energy itself is larger than (11). ∎

IV. Conclusions

Our attempt here is to begin to fill a void in our understanding of energy required for communication. In a relatively little-known paper [15], paralleling his results on zero-energy reversible computation [10], Landauer argues that by reducing friction and noise in the communication medium (to effectively zero), one can communicate with *arbitrarily small energy*. From this perspective, information-theoretic works of Golay [16] and Verdú [17] derive capacity per-unit energy for various communication media that do have friction and noise, but implicitly assume that computation at the transmitter and receiver is frictionless and noiseless. In this paper, we explicitly account for frictional losses in both communication and computation media. Complementary upper bounds and a survey of related physics literature appear in [1]. Ongoing work shows that bounds for an even more general model can be derived for encoding. We believe the next step is to account for noise *and* friction in encoding and decoding computation.

Appendix A
Choosing the stencil parameter a

The number of nodes in a stencil cell is approximately $\frac{a^2}{\lambda^2}$. The actual number could however be larger because of boundary effects. Taking the boundary into account, the number of nodes are bounded by $\frac{a^2}{\lambda^2} + 4\frac{a}{\lambda} + 4$. This quantity is smaller than $2\frac{a^2}{\lambda^2}$ as long as $\frac{a^2}{\lambda^2} \geq 5$.

Choosing $a = \frac{1}{\sqrt{2}}\sqrt{\frac{\log \frac{1}{10P_e^{blk}}}{\log \frac{1}{2p_{ch}}}}\lambda$, $n_i < \frac{2a^2}{\lambda^2} = \frac{\log \frac{1}{10P_e^{blk}}}{\log \frac{1}{2p_{ch}}}$ as desired, as long as $\frac{a^2}{\lambda^2} = \frac{1}{2}\frac{\log \frac{1}{10P_e^{blk}}}{\log \frac{1}{2p_{ch}}} > 5$, which is the condition of the theorem.

References

[1] P. Grover, ""Informational-friction" and its impact on minimum energy per communicated bit," 2013. [Online]. Available: http://tinyurl.com/pulkitgrover/files/ISIT13Extended.pdf

[2] P. Grover, A. Goldsmith, and A. Sahai, "Fundamental limits on the power consumption of encoding and decoding," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Jul. 2012.

[3] P. Grover, K. Woyach, and A. Sahai, "Towards a communication-theoretic understanding of system-level power consumption," *IEEE Journal on Selected Areas in Communication*, Sep. 2011.

[4] P. Grover, "Fundamental limits on the power consumption for lossless signal recovery," in *IEEE Inf. Theory Workshop (ITW)*, Sep. 2012.

[5] A. C.-C. Yao, "Some complexity questions related to distributive computing (preliminary report)," in *ACM symposium on Theory of computing (STOC)*, New York, NY, USA, 1979, pp. 209–213.

[6] C. D. Thompson, "Area-time complexity for VLSI," in *Proceedings of the 11th annual ACM symposium on Theory of computing (STOC)*. New York, NY, USA: ACM, 1979, pp. 81–88.

[7] ——, "A complexity theory for VLSI," Ph.D. dissertation, Pittsburgh, PA, USA, 1980.

[8] M. F. Chang, V. P. Roychowdhury, L. Zhang, H. Shin, and Y. Qian, "RF/wireless interconnect for inter-and intra-chip communications," *Proceedings of the IEEE*, vol. 89, no. 4, pp. 456–466, 2001.

[9] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388 –404, Mar. 2000.

[10] R. Landauer, "Computation: A fundamental physical view," *Physica Scripta*, vol. 35, pp. 88–95, 1987.

[11] A. El Gamal, J. Greene, and K. Pang, "VLSI complexity of coding," in *The MIT Conf. on Adv. Research in VLSI*, Cambridge, MA, Jan. 1984.

[12] N. Alon and J. H. Spencer, *The Probabilistic Method*, 2nd ed. New York, NY: Wiley-Interscience, 2000.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. New York: Wiley, 1991.

[14] T. Courtade, "Properties of the binary entropy function," Oct. 2012, blog entry. [Online]. Available: https://blogs.princeton.edu/blogit/2012/10/26/properties-of-the-binary-entropy-function/

[15] R. Landauer, "Minimal energy requirements in communication," *Science*, vol. 272, no. 5270, pp. 1914–1918, 1996.

[16] M. J. E. Golay, "Note on the theoretical efficiency of information reception with PPM," in *Proc. IRE*, Sep. 1949, p. 1031.

[17] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 9, pp. 1019–1030, Sep. 1990.