Little Green Codes: Energy-Efficient Short-Range Communication

Pulkit Grover and Anant Sahai
Wireless Foundations, Department of EECS
University of California at Berkeley, CA-94720, USA
{pulkit, sahai}@eecs.berkeley.edu

***Abstract*— A green code attempts to minimize the total energy per-bit required to communicate across a noisy channel. The classical information-theoretic approach neglects the energy expended in processing the data at the encoder and the decoder and only minimizes the energy required for transmissions. Since there is no cost associated with using more degrees of freedom, the traditionally optimal strategy is to communicate at rate zero.**

In this work, we use our recently proposed model for the power consumed by iterative message passing. Using generalized sphere-packing bounds on the decoding power, we find lower bounds on the total energy consumed in the transmissions and the decoding, allowing for freedom in the choice of rate. We show that contrary to the classical intuition, the rate for green codes is bounded away from zero for any given error probability. In fact, as the desired bit-error probability goes to zero, the optimizing rate for our bounds on the energy consumption of green codes converges to 1 in the context of AWGN channels with BPSK signalling and hard-decision detection.

I. Introduction

With the development of billion transistor chips, the range of communication has come down dramatically from hundreds of kilometers (e.g. deep space communication) to a few meters (e.g. ad-hoc wireless networks) or a few millimeters or even lesser (e.g. on chip communication). To communicate over smaller distances, the transmit power required is much smaller. At these distances, the energy used in transmissions can be comparable to that expended by the system processes. The small size limits the ability of these chips to dissipate heat. Further, the chip might be battery operated, imposing stringent constraints on its energy usage. It is therefore of interest to design of coding techniques that minimize the *total energy* consumed, which includes the transmission energy as well as the processing energy. We refer to the coding techniques that minimize the total energy as *green codes*.

Classical information theoretic approach finds the minimum *transmission* energy required to communicate reliably across the channel. The approach is motivated by long-range communication, that corresponds to power constrained channels. Shannon [1] first characterized the minimum energy required to communicate across a channel with fixed rate. The resulting bounds are expressed using 'waterfall' curves that convey the revolutionary idea that unboundedly low probabilities of bit-error are attainable using only finite transmit power. This characterization raises a natural question: what is the minimum energy required for communication that is free of rate constraint? The classical approach [2] [3] gives the minimum transmission energy required (on average) to communicate one bit reliably across the channel. For example, for AWGN channel of noise variance 1, this minimum energy is

$$\lim_{P_T \to 0} \frac{P_T}{C(P_T)} = 2\ln(2) \text{ Joules.} \quad (1)$$

For BSC with hard decision detection (see Section II), this limit is about 2.18 Joules. Since there is no penalty associated with lower rates, it is good to use as many degrees of freedom as are available, and the optimal transmission rate is zero.

The problem of minimizing combined transmission and processing energy is well studied in networks. The common thread in [4], [5], [6], [7], [8], [9] is that the energy consumed in processing the signals can be a substantial fraction of the total power. In [7], an information-theoretic formulation is considered. The authors model the processing energy by a constant ϵ per unit time when the transmitter is transmitting (and hence, is in the 'on' state). A total of r channel uses are allowed, and the total energy available is $r \times \mathcal{E}$. Let P_i be the transmit power at i-th time instant, and let $C(P_i)$ be the capacity of the corresponding channel. Then the problem is to transmit maximum number of bits with the total power less than $r \times \mathcal{E}$. That is,

$$\max \quad \sum_{i=1}^{r} 1_i C(P_i) \quad (2)$$

$$\text{subject to} \quad \sum_{i=1}^{r} 1_i \times (P_i + \epsilon) \leq r\mathcal{E} \quad (3)$$

where 1_i is 1 if a symbol is transmitted in the i-th channel use, and is 0 otherwise. This is equivalent to dividing the channel into r sub-channels, with independent coding on each sub-channel. Since the capacity function $C(P)$ is concave in its argument, for maximizing the total number of information bits communicated, the transmission energy P_i should be equal for all i where $1_i = 1$. Without the energy consumed by the system processes, the optimal strategy would be to use all the r parallel channels, and share the energy equally between all of them. However, the energy consumed by the system processes imposes a fixed penalty on each channel use. The authors quantify this tension by measuring 'burstiness' Θ of signaling defined as $\Theta = \frac{1}{r}\sum_{i=1}^{r} 1_i$.

The transmissions should not be too bursty because of law of diminishing returns associated with the $\log(\cdot)$ function. On the other hand, the transmission strategy should not make use of all degrees of freedom either, since there is an ϵ cost associated with use of each degree of freedom[1]. The authors conclude that for minimum total energy, $0 < \Theta < 1$. Contrary to conventional information theoretic wisdom, it is no longer optimal to use all available degrees of freedom. Since the optimal strategy is not to use all available degrees of freedom, evidently, the optimal rate that minimizes the total energy consumption is bounded away from zero. That is, *if processing energy is taken into account, green codes may not communicate at zero rate!*

The objective in [7] [5] [9] is to reduce the energy consumption for wireless devices that ". . . when inserted . . . consume energy continuously" [9]. Examples of such devices are handheld computers, high-end laptops, etc. Energy consumption per unit time for such devices is indeed well modeled by a constant possibly independent of the coding strategy being used. We consider the energy expended by the decoding process itself. The decoding circuit requires some non-zero energy to perform each operation.As opposed to energy consumed by system processes in [7], [5], [9], the decoding energy depends significantly on the code construction, the rate and the desired error probability, and therefore needs more careful modeling.

In this work, we study explicit models of energy expended at the decoder. We concentrate on the message passing decoder. We derive lower bounds on the combined transmission and decoding energy, with no constraint on the rate. We show that the optimizing rate for green codes is indeed bounded away from zero. In fact, as the error probability converges to zero, the optimizing rate converges to 1!

The organization of the paper is as follows : In Section II, we introduce the channel model, the decoder model, and the energy model. In Section III, we summarize some of our results in [10] and [11]. In Section IV, we then build on the results in [10] and [11] to find bounds on the minimum total energy required to communicate across a channel, with no rate constraint, taking into account the decoding energy as well. We conclude in Section V.

II. System model

Consider a point-to-point communication link. An information sequence $\mathbf{B}_1^k$ is encoded into 2^{mR} codeword symbols $\mathbf{X}_1^m$, using a possibly randomized encoder. The observed channel output is $\mathbf{Y}_1^m$. The information sequences are assumed to consist of iid fair coin tosses and hence the rate of the code is $R = k/m$.

The channel model considered is the BSC that we consider to have resulted from BPSK modulation followed by hard-decision detection on an AWGN channel of noise variance $\sigma_P^2 = 1$. The true channel is always denoted P, and its crossover probability by p. Thus, $p = \mathcal{Q}\left(\sqrt{P_T}\right)$.

[1]The authors do not allow for the possibility of communicating using timing by strategically turning the transmitter on and off. This is well founded, since it is unrealistic that frequent turning on and off of the device will save any energy.

For maximum generality, we do not impose any *a priori* structure on the code itself. Instead, inspired by [12], [13], [14], [15], we focus on the parallelism of the decoder and the energy consumed within it. We assume that the decoder is physically made of computational nodes that pass messages to each other in parallel along physical (and hence unchanging) wires. A subset of nodes are designated 'message nodes' in that each is responsible for decoding the value of a particular message bit. Another subset of nodes (not necessarily disjoint) has members that are each initialized with at most one observation of the received channel output symbols. There may be additional computational nodes that are just there to help decode.

The implementation technology is assumed to dictate that each computational node is connected to at most $\alpha + 1 > 2$ other nodes[2] with bidirectional wires. No other restriction is assumed on the topology of the decoder. In each iteration, each node sends (possibly different) messages to all its neighboring nodes. **No restriction is placed on the size or content of these messages except for the fact that they must depend on the information that has reached the computational node in previous iterations.** If a node wants to communicate with a more distant node, it has to have its message relayed through other nodes. The neighborhood size at the end of l iterations is denoted by $n \leq \alpha^{l+1}$. Each computational node is assumed to consume a fixed E_{node} joules of energy at each iteration.

Let the average probability of bit error of a code be denoted by $\langle P_e \rangle_P$ when it is used over channel P. The goal is to derive a lower bound on the neighborhood size n as a function of $\langle P_e \rangle_P$ and R. This then translates into a lower bound on the number of iterations which can in turn be used to lower bound the required decoding power.

Throughout this paper, we allow the encoding and decoding to be randomized with all computational nodes allowed to share a common pool of common randomness. We use the term 'average probability of error' to refer to the probability of bit error averaged over the channel realizations, the messages, the encoding, and the decoding.

III. Lower bounds on the decoding complexity and total energy

In this section we summarize our results from [10] and [11]. In Section III-A, we provide lower bounds on the decoding complexity for the BSC given the rate and the error probability. These bounds are then used to derive lower bounds on total energy consumed in the transmissions and the decoding in Section III-B.

A. Lower bounds on the decoding complexity for the BSC

The main bounds are given by theorems that capture a local sphere-packing effect. These can be turned around to give a family of lower bounds on the neighborhood size n as a

[2]In practice, this limit could come from the number of metal layers on a chip. $\alpha = 1$ would just correspond to a big ring of nodes and is therefore uninteresting.

function of $\langle P_e \rangle_P$. This family is indexed by the choice of a hypothetical channel G and the bounds can be optimized numerically for any desired set of parameters.

Theorem 3.1: Consider a BSC with crossover probability $p < \frac{1}{2}$. Let n be the maximum size of the decoding neighborhood of any individual bit. The following lower bound holds on the average probability of bit error

$$\langle P_e \rangle_P \geq \sup_{C^{-1}(R) < g \leq 0.5} \frac{h_b^{-1}(\delta(G))}{2} 2^{-nD(g||p)} \left(\frac{p(1-g)}{g(1-p)} \right)^{\epsilon \sqrt{n}} \tag{4}$$

where $h_b(\cdot)$ is the usual binary entropy function, $D(g||p) = g \log_2 \left(\frac{g}{p}\right) + (1-g) \log_2 \left(\frac{1-g}{1-p}\right)$ is the usual KL-divergence, and

$$\delta(G) = 1 - \frac{C(G)}{R} \tag{5}$$

where $C(G) = 1 - h_b(g)$, and

$$\epsilon = \sqrt{\frac{1}{K(g)} \log_2 \left(\frac{2}{h_b^{-1}(\delta(G))} \right)}, \tag{6}$$

where

$$K(g) = \inf_{0<\eta<1-g} \frac{D(g+\eta||g)}{\eta^2}. \tag{7}$$

Proof: See [11]. ■

Taking log on both sides of (4), it is clear that the term $nD(g||p)$ dominates the other terms in the RHS at large values of n. For low $\langle P_e \rangle_P$, g can be taken close to $C^{-1}(R)$. Neglecting the other two terms, we get $n \geq \frac{\log(1/\langle P_e \rangle)}{D(C^{-1}(R)||p)}$, an approximate lower bound on n that helps understand how the lower bound behaves with $\langle P_e \rangle$ and R for a fixed P_T.

B. Joint optimization of the weighted total power

Consider the total energy spent in transmission. For transmitting k bits at rate R, the number of channel uses is $m = k/R$. If each transmission has power $\xi_T P_T$, the total energy used in transmission is $\xi_T P_T m$.

At the decoder, let the number of iterations be l. Assume that each node consumes E_{node} joules of energy in each iteration. The number of computational nodes can be lower bounded by the number m of received channel outputs.

$$E_{dec} \geq E_{node} \times m \times l. \tag{8}$$

This gives a lower bound of $P_D \geq E_{node} l$ for decoder power. There is no lower bound on the encoder complexity and so the encoder is considered free. This results in the following bound for the weighted total power

$$P_{total} \geq \xi_T P_T + \xi_D E_{node} \times l. \tag{9}$$

Using $l \geq \frac{\log(n)}{\log(\alpha)}$ as the natural lower bound on the number of iterations given a desired neighborhood size,

$$\begin{aligned} P_{total} &\geq \xi_T P_T + \frac{\xi_D E_{node} \log(n)}{\log(\alpha)} \\ &\propto \frac{P_T}{\sigma_P^2} + \gamma \log(n) \end{aligned} \tag{10}$$

where $\gamma = \frac{\xi_D E_{node}}{\sigma_P^2 \xi_T \log(\alpha)}$ is a constant that summarizes all the technology and environmental terms. Figure 1 provides example[3] behavior of γ with distance. The neighborhood size n itself can be lower bounded by plugging the desired average probability of error into Theorem 3.1.

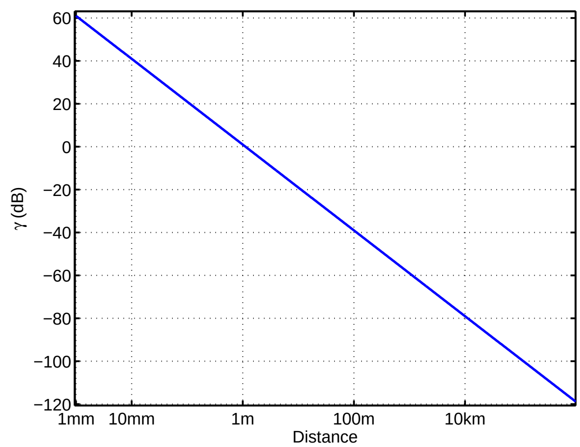

Fig. 1. The plot shows the behavior of γ with distance d for path loss $\frac{1}{\xi_T} = \frac{1}{d^2}$ for $d > 0.1$mm (and path loss 1 for smaller d). E_{node} is 1pJ, $\alpha = 4$, $\xi_D = 1$, and $\sigma_P^2 = 4 \times 10^{-21}$J.

IV. Minimization of total energy by optimizing over rate

We can now translate the expression for total power (10) into an expression on the total energy,

$$\begin{aligned} E_{\text{total}} &= m \times (P_T + \gamma \log(n)) \\ &= \frac{k}{R} P_T + \frac{k}{R} \gamma \log(n) \\ \Rightarrow E_{\text{per bit}} &= \frac{1}{R} P_T + \frac{1}{R} \gamma \log(n) \end{aligned} \tag{11}$$

For minimum energy per information bit, we can minimize (11) over rate R and transmit power P_T.

Observe that in (11), the decoding energy increases as the error probability decreases for the same transmit power. This behavior is not reflected by using the model inspired from (12) for decoding energy. We plot the two bounds against each other in Figure 2 for $k = 10,000$ bits. Since the bounds in [7] are for error probability converging to zero, in Appendix I we derive bounds for non-zero error probability based on the model in [7].

We choose $\epsilon = 2$, for which the total energy per bit for the bound inspired from [7] equals the energy per bit for $\gamma = 0.4$ for our bound for $\langle P_e \rangle = 10^{-4}$. The figure shows that for $\langle P_e \rangle$ larger than this threshold, the model inspired from [7] underestimates the total energy. It is because this model treats

[3]The energy cost of one iteration at one node $E_{node} \approx 1$ pJ is arrived at by an optimistic extrapolation from the reported values in [16], [17], thermal noise energy per sample $\sigma_P^2 \approx 4 \times 10^{-21}$J from kT with T around room temperature.

the decoder as a black-box where ϵ does not change with error probability.

Figure 3 shows the behavior of our lower bound on sum energy with $\langle P_e \rangle$ for various values of γ.

It is interesting to observe what values of R optimize (11). Fix a small desired error probability $\langle P_e \rangle$. At rate converging to zero, the number of output nodes is much larger than the number of information bits, and hence the decoding energy tends to get large. Looking at (11), to attain finite per bit energy, it is necessary to have $P_T \to 0$ as $R \to 0$. However, as $P_T \to 0$, $p \to 0.5$. Theorem 3.1 and the ensuing small $\langle P_e \rangle_P$ approximation gives $n \geq \frac{\log(1/\langle P_e \rangle)}{D(C^{-1}(R)||p)}$.

Observe that $D(g||p)$ as a function of g is continuous and bounded above in a small neighborhood around $p = 0.5$. Since R and P_T converge to zero, choose R and P_T small enough such that $D(C^{-1}(R)||p)$ is bounded above by 0.5. Then n is bounded *below* by $2\log(1/\langle P_e \rangle)$. Therefore, as $R \to 0$, the decoding energy (which is $\frac{1}{R}\gamma \log(n)$) increases to infinity. For finite $\langle P_e \rangle$, R is bounded away from zero.

Evidently, there exists an optimal rate $R_{\text{opt}} > 0$ that minimizes the combined energy consumed. We plot the behavior of the optimal rate with $\langle P_e \rangle$ in Figure 4.

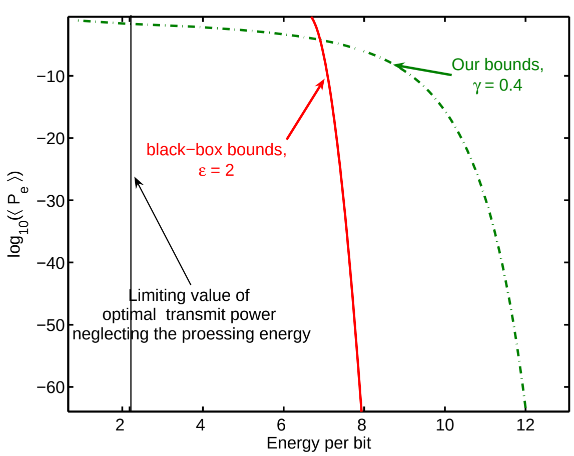

Fig. 2. The plot shows the comparison of lower bounds on the minimum energy for $k = 10,000$ bits. The 'black-box bounds' plot is based on model in [7], where the details of the processor are ignored. Our bounds take into account the decoder structure as well. Since the neighborhood size, and hence the decoding energy increases as $\langle P_e \rangle \to 0$, the gap between the two bounds increases as $\langle P_e \rangle \to 0$.

V. Discussions and Conclusions

In this work, we derived lower bounds on the combined transmission and decoding energy for iterative message passing decoding for unconstrained rates for a BSC. An interesting feature of the resulting bounds is that the optimizing rate for green codes is bounded away from zero. In fact, it converges to 1 as the error probability converges to zero. However, the plots show that total energy is much below that predicted by uncoded transmission. Hence uncoded transmission is not asymptotically optimal.

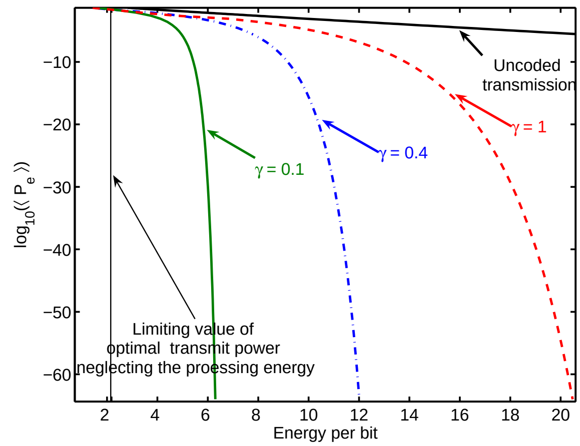

Fig. 3. The plot shows the behavior of lower bound on sum energy with $\langle P_e \rangle$ for various values of γ. The plots perform better than uncoded transmission. In Figure 4, the optimal rate converges to 1 as $\langle P_e \rangle \to 0$. Even so, this plot shows that the optimal strategy is not uncoded transmission at low $\langle P_e \rangle$.

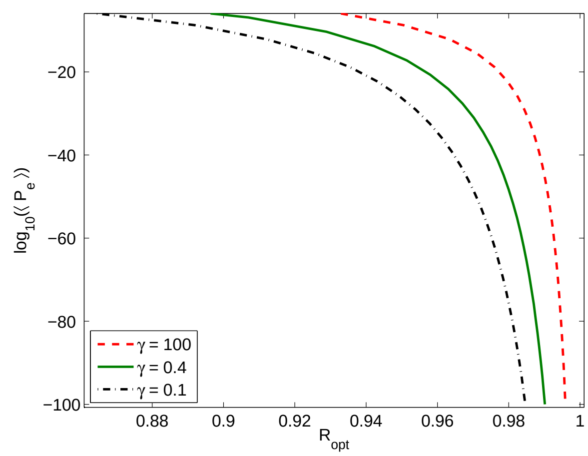

Fig. 4. Optimal value of rate vs error probability: As $\langle P_e \rangle$ converges to 0, the optimizing rate converges extremely slowly to 1.

In this work our emphasis is on the BSC. An extension to the AWGN channel using bounds in [10], [11] would also be interesting. However, for the AWGN channel the rate can be greater than 1, in which case the number of output nodes is smaller than the input nodes, and lower bounding the total number of nodes by the number of output nodes would be loose. Instead, the bounding would be by the maximum of the number of information nodes and output nodes, which saturates to the number of information nodes above rate 1. In a sense, this saturation is captured by BSC, so the derived bounds may be qualitatively similar.

Appendix I
Bounds in [7] for Non-Zero Error Probability

Observe that the bounds in [7] are for the reliable communication, that is, $\langle P_e \rangle \to 0$. Parallel to our analysis for message passing decoding, in this section, we build on the analysis in [7] to derive bounds on the minimum energy required for communicating with a non-zero error probability $\langle P_e \rangle$.

Assume k bits are to be transmitted across the channel, with desired error probability $\langle P_e \rangle$. In [7], the authors maximize the information bits communicated under a total energy constraint. Turning around the problem in [7], we can instead minimize the total energy consumed given the number of bits transmitted. Now we can add an error probability constraint to the bits transmitted. Assume that a block code is used to communicate across the channel. The corresponding error exponent is bounded by the sphere-packing bound. Assuming optimistically that the code actually achieves the sphere-packing bound in exponent,

$$\langle P_e \rangle \leq \langle P_e \rangle_{block} \approx 2^{-mE_{sp}(P_T,R)}$$

where $E_{sp}(P_T,R)$ is the sphere-packing bound at rate R and transmit power P_T across a channel. The objective, therefore, is

$$\begin{aligned} \min_{P_T,m} \quad & m \times (P_T + \epsilon) \\ \text{subject to} \quad & m \times E_{sp}\left(P_T, \frac{k}{m}\right) = \log_2(\langle P_e \rangle). \end{aligned} \quad (12)$$

References

[1] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, Jul./Oct. 1948.

[2] J. Pierce, “Optical channels: Practical limits with photon counting,” *IEEE Transactions on Communications*, pp. 1819–1821, Dec. 1978.

[3] S. Verdu, “On channel capacity per unit cost,” *IEEE Trans. Inform. Theory*, vol. 36, pp. 1019–1030, 1990.

[4] P. Agrawal, “Energy efficient protocols for wireless systems,” in *IEEE International Symposium on Personal, Indoor, Mobile Radio Communication*, 1998, pp. 564–569.

[5] S Cui, AJ Goldsmith and A Bahai, “Energy Constrained Modulation Optimization,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 1–11, 2005.

[6] A. J. Goldsmith and S. B. Wicker, “Design challenges for energy constrained ad hoc wireless networks,” *IEEE Trans. Wireless Commun.*, pp. 8–27, 2002.

[7] P. Massaad, M. Medard, and L. Zheng, “Impact of Processing Energy on the Capacity of Wireless Channels,” in *International Symposium on Information Theory and its Applications (ISITA)*, 2004.

[8] S. Vasudevan, C. Zhang, D. Goeckel, and D. Towsley, “Optimal power allocation in wireless networks with transmitter-receiver power tradeoffs,” *Proceedings of the 25th IEEE International Conference on Computer Communications INFOCOM*, pp. 1–11, Apr. 2006.

[9] R. Kravertz and P. Krishnan, “Power management techniques for mobile communications,” *The Fourth Annual ACM/IEEE International Conference on Mobile Computing and Networking*, 1998.

[10] P. Grover and A. Sahai, “A general lower bound on the decoding complexity of sparse-graph codes,” in *IEEE Workshop on Information Theory submitted*, 2008.

[11] A. Sahai and P. Grover, “The price of certainty : “waterslide curves” and the gap to capacity,” *Submitted to IEEE Transactions on Information Theory, available online at http://arXiv.org/abs/0801.0352v1*, Dec. 2007.

[12] P. P. Sotiriadis, V. Tarokh, and A. P. Chandrakasan, “Energy reduction in VLSI computation modules: an information-theoretic approach,” *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 790–808, Apr. 2003.

[13] N. Shanbhag, “A mathematical basis for power-reduction in digital VLSI systems,” *IEEE Trans. Circuits Syst. II*, vol. 44, no. 11, pp. 935–951, Nov. 1997.

[14] T. Koch, A. Lapidoth, and P. P. Sotiriadis, “A channel that heats up,” in *Proceedings of the 2007 IEEE Symposium on Information Theory*, Nice, France, 2007.

[15] ——, “A hot channel,” in *2007 IEEE Information Theory Workshop (ITW)*, Lake Tahoe, CA, 2007.

[16] M. Bhardwaj and A. Chandrakasan, “Coding under observation constraints,” in *Proceedings of the Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sep. 2007.

[17] S. L. Howard, C. Schlegel, and K. Iniewski, “Error control coding in low-power wireless sensor networks: when is ECC energy-efficient?” *EURASIP Journal on Wireless Communications and Networking*, pp. 1–14, 2006.