Coding Under Observation Constraints

Manish Bhardwaj Anantha Chandrakasan
Microsystems Technology Laboratory
Massachusetts Institute of Technology
Cambridge MA, 02139, USA.
{manishb,anantha}@mtl.mit.edu

***Abstract*—We consider coding schemes when performance is measured by the average signal observation time to reliably decode an information bit, as opposed to conventional metrics of transmit energy per bit or spectral efficiency. This formulation is motivated by energy constrained communications devices where sampling the signal, rather than transmitting or processing it, dominates energy consumption. We show that sequentially observing samples with the maximum *a posteriori* entropy can significantly reduce observation costs. Equivalently, observation costs identical to traditional coding are achieved at blocklengths that are an order of magnitude smaller. To put this in perspective, our sampling strategy can be applied to realizing feedback systems that surpass the cutoff rate limit using the (24,12) Golay code, the highest such performance reported over the AWGN channel at these blocklengths.**

I. Introduction

The traditional performance measures in a communications system are the transmitted energy per information bit, or the spectral efficiency, achieved under a specified reliability constraint. Systems are classified as power or bandwidth-limited accordingly and different design approaches used.

Our work deals with communications systems where the measure of performance is the average number of observations a receiver makes to decode an information bit, or more generally, to infer a signal or channel parameter. This formulation is motivated by low-data rate, short-range ultra-wideband (UWB) wireless systems where electronics energy dissipation is often dominated by signal observation time. We propose techniques that trade higher transmit costs (radiated power, bandwidth) for significantly reduced observation costs.

We begin with a discussion of the devices that motivate our idealized observation cost formulation. This is followed by illustrating the connections between observation cost and feedback coding. We then use the sequential probability ratio test (SPRT) as a simple example of the significant gains possible via intelligent sampling techniques. The more general minimum sampling problem when unlimited copies of a codeword are available to the receiver is then discussed, and performance of sampling criteria compared with existing alternatives. We end with a note on computational complexity.

II. Motivation

There has been a lot of activity recently in the domain of short range (10 m), low data-rate (100s of kbps) ultra-wideband (UWB) devices for applications like positioning, tagging and telemetry. The IEEE recently approved a draft standard, 802.15.4a, aimed at this space. The target applications dictate that devices must be very low in cost and have a small form factor. Hence, voluminous batteries cannot be used, and battery replacement is not an option. This, along with an expected operational lifetime of several years, leads to extreme demands on energy efficiency.

The problem of minimizing energy consumption reduces to that of minimizing observation if two conditions are satisfied. First, sampling the received signal should be the dominant source of energy consumption in the communications link. Second, sampling energy must be proportional to the number of samples taken. We now determine when these assumptions are valid.

Figure 1 shows the makeup of an example wireless transmitter and receiver.

We divide transmit electronics into signal generation and amplification. The receiver is divided into signal sampling and the subsequent computation. We use $\mathcal{E}_{\text{A}}$ to denote the *electronics* energy consumed by component *A*, per bit. Hence, $\mathcal{E}_{\text{tx}} = \mathcal{E}_{\text{gen}} + \mathcal{E}_{\text{amp}}$, and $\mathcal{E}_{\text{rx}} = \mathcal{E}_{\text{samp}} + \mathcal{E}_{\text{comp}}$.

Transmit versus receive energy: Of the many factors that determine how $\mathcal{E}_{\text{tx}}$ and $\mathcal{E}_{\text{rx}}$ compare, the regulatory limit on output (i.e. radiated) power is usually the governing one. When a high output power is permitted, as in wireless LANs, the transmitter is likely to dominate due to $\mathcal{E}_{\text{amp}}$. When regulatory limits are tight, as in UWB systems, the receiver, with its significantly more complex signal conditioning and processing, dominates consumption. For instance, Lee, Wentzloff and Chandrakasan have recently demonstrated a UWB system with $\mathcal{E}_{\text{tx}}$=47 pJ/bit and $\mathcal{E}_{\text{rx}}$=2.5 nJ/bit, i.e. the receiver dominates energy consumption by a factor of 50 [1], [2].

Sampling versus processing energy: If $\mathcal{E}_{\text{rx}} \gg \mathcal{E}_{\text{tx}}$, the next question is whether $\mathcal{E}_{\text{samp}} \gg \mathcal{E}_{\text{comp}}$. This comparison is difficult because notions of complexity, and sources of energy consumption are often markedly different in analog and digital circuits. One prevailing view is that the energy efficiency of digital circuits scales more aggressively than that of analog circuits as technology progresses. Thus the $\mathcal{E}_{\text{samp}}/\mathcal{E}_{\text{comp}}$ ratio can be expected to grow with time.

The receiver of Lee et al. quoted above is among the most energy efficient at the data rates of interest to us [1]. Also, virtually all its energy is consumed in sampling.

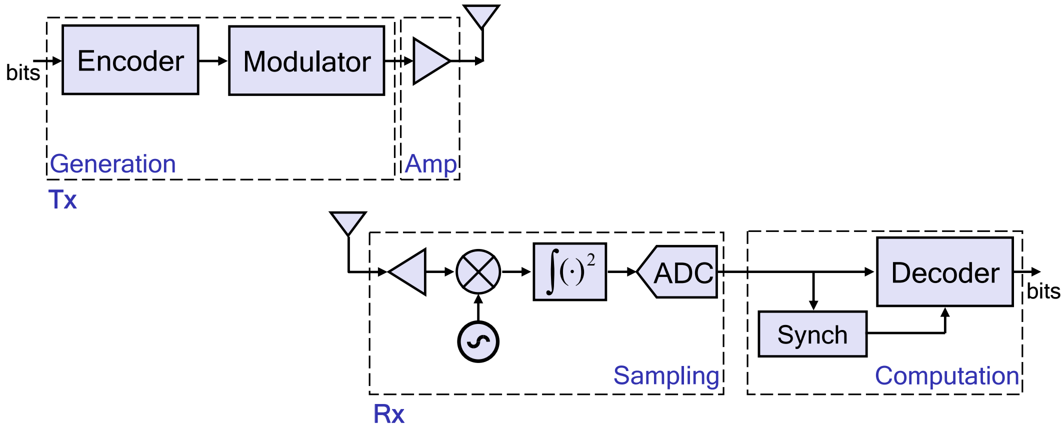


Fig. 1: Components of a typical wireless transmitter and a non-coherent receiver.

Hence, 2.5nJ/bit is indicative of the limits of energy efficient sampling in current technology. Compare this with the 0.2 nJ/bit consumed by a 1024 b LDPC decoder operating 3 dB away from capacity at a BER of 10^{-6} [3].This illustrates that $\mathcal{E}_{\text{samp}} \gg \mathcal{E}_{\text{comp}}$ is a tenable assumption, provided the receiver bears a 'reasonable' computational burden.

Proportionality of sampling energy and duration: There are physical and architectural constraints on how fast a sampler can turn on and off. These fixed delays are unrelated to the duration of observation, and can weaken the proportionality assumption. A related point is that some components of the sampler might need to stay on even while not sampling. Again, Lee et al. have demonstrated a negligible turn-off time and a turn-on time of 3 ns in a system with a minimum observation duration of 30 ns. Also, the most energy hungry components are turned off when not sampling. Hence, a proportional model would be appropriate for their receiver.

Finally, overall system constraints must permit sacrificing degrees of freedom (i.e. rate or range) to reduce sampling costs. This energy efficiency vs. capacity tradeoff is driven by the target application. E.g., the 15.4a standard includes non-coherent signaling techniques that sacrifice on the order of 5-10 dB in link margin to permit lower complexity, energy efficient devices. This suggests that a subset of short-range, low-data-rate UWB applications permit trading off capacity for energy efficiency.

III. Preliminaries

We begin with the setup and some definitions. Consider the communications system in figure 2.

Information is conveyed over a discrete-time, binary-input AWGN channel. The system uses a code $\mathcal{C}$ and BPSK modulation (not shown explicitly). A procedure $\mathcal{P}$ observes noisy channel outputs and infers the information bits. The observation cost is equal to the *expected* number of samples observed by $\mathcal{P}$ to declare a result[1].

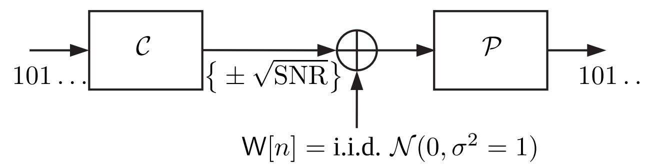


Fig. 2: Setup of the problem of coding under observation cost constraints.

Problem 1 (Coding under observation cost constraints)**.** What choice of $\mathcal{C}$ and $\mathcal{P}$ minimizes the expected cost per information bit under a specified SNR and BER constraint?

Note that the transmitter can send as many coded bits as it desires. What matters is that the receiver judiciously pick the bits (samples) it observes.

The terms 'conventional' and 'traditional' when applied to a code, decoding technique or system refer to a fixed-length code with ML decoding i.e. a codeword is transmitted exactly once, the receiver samples the entire codeword, and uses a ML decoder.

Given a (n, k) block code $\mathcal{C}$, we define the *code rate*, $r_{\mathcal{C}} = k/n$ and 'SNR per bit', $E_b/N_0 = \text{SNR}/(2r_{\mathcal{C}})$. The *coding gain* at a specified probability of bit-error is denoted by $\gamma_{\mathcal{C}}(p_{\text{b}})$ and measures the decrease in required E_b/N_0 compared with an uncoded system.

The code rate must be distinguished from the overall *transmit rate* $r_{\text{tx}} = k/n'$ where n' is the total number of bits transmitted for k information bits. In conventional systems, $r_{\text{tx}} = r_{\mathcal{C}}$. When infinite copies of the codeword are transmitted, $r_{\text{tx}} \to 0$.

The procedure $\mathcal{P}$ is characterized by the following rules:

- Stopping: Should the procedure observe another sample?

[1]We will use the terms 'observation cost', 'receive cost' and 'sampling cost' interchangeably.

- Experimentation: If so, what sample should be observed next?
- Decision: If not, what result should be declared on stopping?

We use the r.v. N to denote the number of samples observed prior to stopping. The *receive rate* is defined analogously to the transmit rate, $r_{\text{rx}} = k/\mathrm{E}[\mathsf{N}]$. Again, $r_{\text{rx}} = r_{\text{tx}} = r_{\mathcal{C}}$ in a conventional system. In our more general setting, $r_{\text{rx}} \geq r_{\text{tx}}$, while r_{rx} and $r_{\mathcal{C}}$ obey no mutual constraint.

The observation or receive cost per bit, $c_{\text{bit}} = \mathrm{SNR}/(2r_{\text{rx}})$, is defined analogously to E_b/N_0. For conventional systems, $c_{\text{bit}} = E_b/N_0$ allowing easy comparison of receive costs. Note that while stating problem (1), we equated cost with the number of samples observed per bit. Our definition here is broader because it allows comparison of schemes operating at different SNRs.

Problem 2. Minimize SNR (equivalently c_{bit}) for a specified r_{rx} and p_{b}.

This is equivalent to problem (1) in the sense that the solution space for both is characterized by the admissible $(\mathrm{SNR}, r_{\text{rx}}, p_{\text{b}})$ tuples.

A. *Fundamental limits and relationship to feedback*

It follows from Shannon's capacity theorem that,

$$\frac{E_b}{N_0} \geq \frac{2^{2r} - 1}{2r}$$

for reliable communication, where r is the information rate in bits/channel use. This limit also applies to receive costs as defined above,

$$c_{\text{bit}} \geq \frac{2^{2r_{\text{rx}}} - 1}{2r_{\text{rx}}} \tag{1}$$

Every system can achieve an E_b/N_0 equal to c_{bit} via feedback[2]. Since feedback does not increase capacity, the limit on receive cost follows.

These arguments illustrate the close connection of our problem to that of coding with feedback. Feedback schemes can dramatically increase reliability over traditional coding for identical blocklength. Some such schemes are directly applicable to our problem. For instance, decision feedback schemes, which involve re-transmission of codewords based on the decoder output, provide achievable upper-bounds for receive costs. Also, a variety of *hybrid Automatic-Repeat-Request* (hybrid ARQ) schemes employed in communications networks have receive cost counterparts, and we will use such schemes as benchmarks for our proposed solutions.

Despite these connections, it is important to note that feedback coding and receive cost coding are not equivalent. While every system with feedback can (trivially) achieve a transmit cost equal to any achievable receive cost, the converse is not true. Restated, there exist feedback coding schemes that do not translate to receive cost schemes. This is because feedback encoders can exploit perfect knowledge of *both* channel outputs and the codeword. For instance, they can send termination hints to the receiver (in addition to repeating the codeword), significantly improving performance over decision feedback. Such hints are clearly not possible in our setup.

[2] In our work, 'feedback', when used without qualification, implies a transmitter that knows channel outputs exactly.

B. *Reducing costs via traditional coding*

Consider the setup in figure 2 operating at a SNR of 10.5 dB. A (2,1) repetition code is required to achieve a BER of 10^{-6}. We can achieve the same performance via a $(4,3)$ parity-check code. Hence, the number of symbols required for 3 information bits may be reduced from 6 to 4 – a savings, or coding gain $\gamma_{\mathcal{C}}(10^{-6})$ of 1.5x (1.7 dB).

However, coding gain does not always translate to lower receive costs. Continuing our example, a (64,6) orthogonal code, i.e. 64-PPM (figure 3), achieves a coding gain $\gamma_{\mathcal{C}}(10^{-6})$=3 dB. However, the receiver must observe 64 6-symbol slots[3] versus 12 symbols for a repetition code, *increasing* receive costs by 32x!

This highlights two differences from conventional coding. First, conventional cost metrics do not have meaningful receive cost analogues when M-ary modulation is used. Second, any conventional coded system that achieves minimum receive cost is necessarily capacity achieving, but the converse is not true, as illustrated by orthogonal codes.

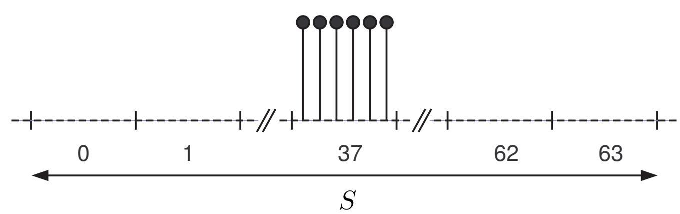


Fig. 3: An example (64,6) orthogonal codeword, S.

A valid question at this juncture is – why not use a capacity achieving Turbo or LDPC code to minimize receive cost? The reason is that the receive problem is governed by a performance versus complexity landscape that is dramatically more favorable when compared with conventional coding. In other words, we can employ codes that are much simpler than traditional ones while achieving identical receive costs.

IV. REDUCING COSTS VIA THE SPRT

Fixed-length repetition codes offer no receive cost improvements. However, substantial reduction in observation is possible if we repeat a bit indefinitely and use a decoder

[3] That each slot has 6 symbols is a consequence of the desired BER and SNR, and unrelated to k=6!

with variable stopping times. Wald's sequential probability ratio test (SPRT) is a scheme that does this optimally [4].

Definition (SPRT)**.** Consider that $n \geq 1$ observations have been made thus far in a binary hypothesis testing problem. The SPRT is defined by the following rule,

$$\ell_{1,2}(\boldsymbol{y}_n) = \sum_{i=1}^{n} \ln \frac{p(y_i \mid \theta_1)}{p(y_i \mid \theta_2)} \begin{cases} \geq A & \text{Accept } \theta_1 \\ \leq B & \text{Accept } \theta_2 \\ \in [B, A] & \text{Continue sampling} \end{cases}$$

where A, B are thresholds derived from the desired error probabilities.

The SPRT can be trivially implemented for our binary AWGN channel (figure 4).

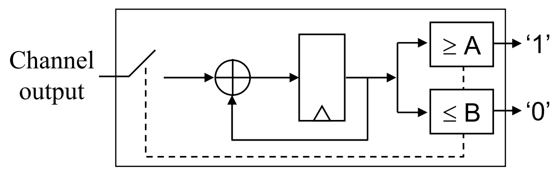


Fig. 4: The SPRT decoder for bit repetition over the AWGN channel.

Figure 5 plots the receive costs of the r_{rx}=1/2 SPRT and the popular, fixed-length [133 171], constraint length $K = 7$, rate 1/2 convolutional code used in e.g., OFDM-based 802.11 WLANs. The SPRT achieves a gain of about 4.5 dB, about a dB short of the (2,1,7) code. Stated in receive cost terms, the SPRT and (2,1,7) code reduce observation to roughly a third and a quarter, respectively, compared with a conventional repetition code. However, the SPRT has a hardware complexity that is several orders of magnitude lower than the (2,1,7) Viterbi decoder – tens of gates compared to tens of thousands. This example illustrates the dramatic difference in performance versus complexity when coding for receive rather than transmit costs. These improvements occur for the same underlying reason as in feedback systems – we exploit atypical channel behavior to achieve much greater reliability (equivalently, much lower observation costs).

Evidently, infinite transmit energy is required to realize the SPRT's benefit. The reality is not as dire. Wald's work on truncation suggests that the SPRT scheme must spend about a dB *more* in transmit costs (E_b/N_0) than a fixed-length system in order to retain most of its benefit. We are in the process of characterizing the impact of non-zero transmit rates on the performance of receive cost techniques.

We can compute the SPRT's asymptotic *sequential* gain, γ_{s}, i.e. gain as $r_{\text{rx}} \to 0$, by comparing its error exponent with that of a fixed-length repetition code. For the binary AWGN channel, we have,

$$\gamma_{\text{s}} = \frac{\mathcal{D}(\mathcal{N}(-a, \sigma^2) \parallel \mathcal{N}(a, \sigma^2))}{\mathcal{D}(\mathcal{N}(0, \sigma^2) \parallel \mathcal{N}(a, \sigma^2))} = 4 \equiv 6 \text{ dB} \qquad (2)$$

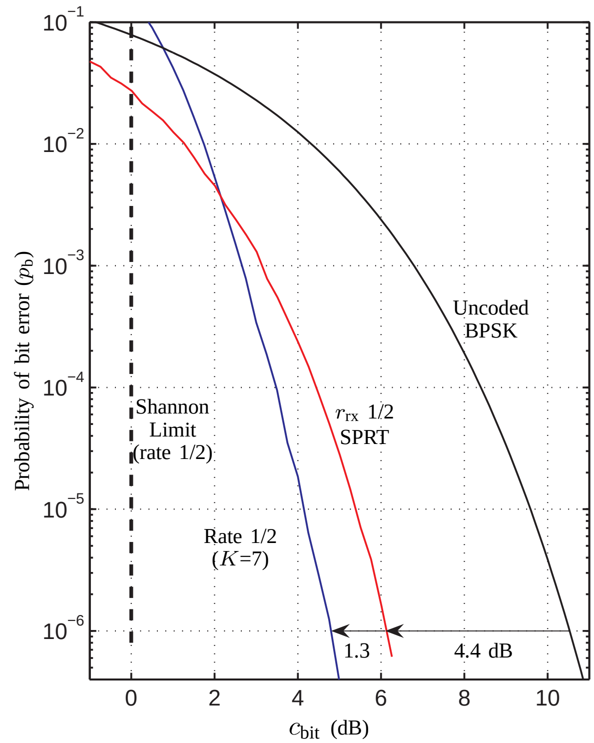


Fig. 5: Receive costs of the rate 1/2 SPRT and K=7, [133 171] convolutional code.

where $\mathcal{D}(\cdot \parallel \cdot)$ denotes the K-L distance. Hence, the most we can expect from the SPRT is a 4x reduction in observation costs.

V. Procedures for General Block Codes

The natural extension of the SPRT idea is sequential sampling of encoded data.

Problem 3 (Decoding block codes under observation constraints)**.** Consider a transmitter that sends infinite copies of a codeword drawn from $\mathcal{C}$, a (n, k) block code, each of which is corrupted by independent noise at the receiver. What procedure minimizes c_{bit} under a r_{rx} and p_{b} constraint?

The problem of choosing from several available coded bits to infer the transmitted codeword is one of *optimum experiment design* i.e. minimizing sampling when several types of observations can be made ('experiments') to infer the state of nature. We begin with an asymptotically optimum procedure (i.e. as $r_{\text{rx}} \to 0$) and then consider the non-asymptotic case.

A. Asymptotic case

In what follows, the set of n experiments is denoted by $\mathscr{E} = \{e\}$. These are 'pure' experiments that form the basis for randomized experiments, whose set we denote by $\mathscr{E}^*$. Likelihoods must be conditioned on not only their underlying hypothesis, but also the experiment. We will use

$\mathrm{P}_{\theta;e} \triangleq p(y \mid \theta; e)$. The distance between two hypotheses under an experiment e will be denoted by $\mathcal{D}_e(\theta_i, \theta_j) \triangleq \mathcal{D}(\mathrm{P}_{\theta_i;e} \parallel \mathrm{P}_{\theta_j;e})$.

Definition (Chernoff's Procedure A)**.** Suppose, without any loss of generality, that the ML estimate is $\hat{\Theta} = \theta_0$ after m observations, $\boldsymbol{y}_m$, have been made. Then,

I. Pick $e(m+1) = \arg \sup_{e \in \mathscr{E}^*} \mu_e$, where, $\mu_e = \inf_{\theta \neq \theta_0} \mathcal{D}_e(\theta_0, \theta)$

II. Terminate if $\min_{j \neq 0} \ell_{0,j}(\boldsymbol{y}_m) > a$.

where a is a positive threshold derived from the desired reliability.

Thus, at every step Procedure A picks the experiment that, roughly speaking, maximizes the minimum K-L distance between the most likely hypothesis and the remaining ones. The termination condition is similar to the SPRT. Chernoff proved that this procedure is asymptotically optimum i.e. no other procedure can reduce the expected number of observations by an order-of-magnitude compared with Procedure A as $\max\ \varepsilon_\theta \to 0$ [5]. Applying Procedure A to the decoding problem (3) suggests the following result.

Conjecture 1 (Asymptotic optimality of uniform sampling)**.** *Given a 'good'* (n, k, d) *linear code* $\mathcal{C}$*, uniformly sampling coded bits is an optimum sequential strategy as* $\max\ \varepsilon_\theta \to 0$ *(equivalently* $r_{\text{rx}} \to 0$*).*

A 'good' code is one whose coding gain cannot be improved via coordinate deletions. Note that the expected number of observations is made large for a *fixed* block size. This is in contrast to information theoretic results where block size is made large. One can prove that uniformly sampling a code reduces sampling costs by $\gamma_{\mathcal{C}} \cdot \gamma_{\text{s}}$ asymptotically i.e. the gain is a product of the coding and sequential gain (as defined in (2)).

B. Non-asymptotic case

Procedure A suffers from two key drawbacks when the expected number of observations is small (i.e. r_{rx} is not close to zero). First, its sampling choice relies on the ML estimate, which can be very unreliable in the initial phases of observation. Second, it ignores the *a posteriori* probabilities of hypotheses in its formulation of the distance metric μ_e.

Blot and Meeter proposed Procedure B which tackles the second problem by incorporating reliability information in distance calculations [6],

$$e(m+1) = \arg\max_{e \in \mathscr{E}} \mu_e, \text{ where } \mu_e = \sum_{\theta} \Pi_\theta \mathcal{D}_e(\hat{\Theta}, \theta)$$

where Π_θ is the *a posteriori* probability of the hypothesis. Box and Hill proposed a metric that eliminates reliance on the ML estimate altogether [7],

$$\mu_e = \sum_{\theta'} \sum_{\theta} \Pi_{\theta'} \Pi_\theta \left[\mathcal{D}_e(\theta', \theta) + \mathcal{D}_e(\theta, \theta')\right]$$

For the decoding problem, we denote the distance metric for the experiment entailing observation of bit j by μ_j. If the posterior probability of the ML codeword exceeds 1/2, Procedure B yields,

$$\mu_j^{(\mathrm{B})} = \min(\pi_j, 1 - \pi_j)$$

where π_j is the *a posteriori* probability of the bit being 1. Hence, the procedure picks the bit with the maximum *a posteriori* entropy. This matches our intuition that the most uncertain bit yields the most information. The more symmetric distance metric of Box-Hill yields,

$$\mu_j^{(\mathrm{BH})} = 2\pi_j(1 - \pi_j)$$

Thus, Box-Hill also prescribes maximum entropy (ME) sampling.

VI. Results

We begin with the simple example of ME sampling the (7,3,4) dual Hamming code with r_{rx}=1/2 (figure 6).

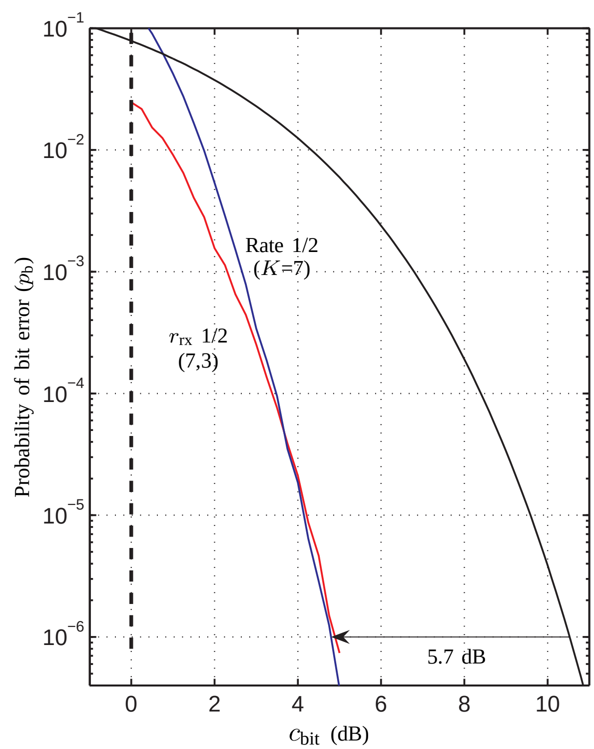
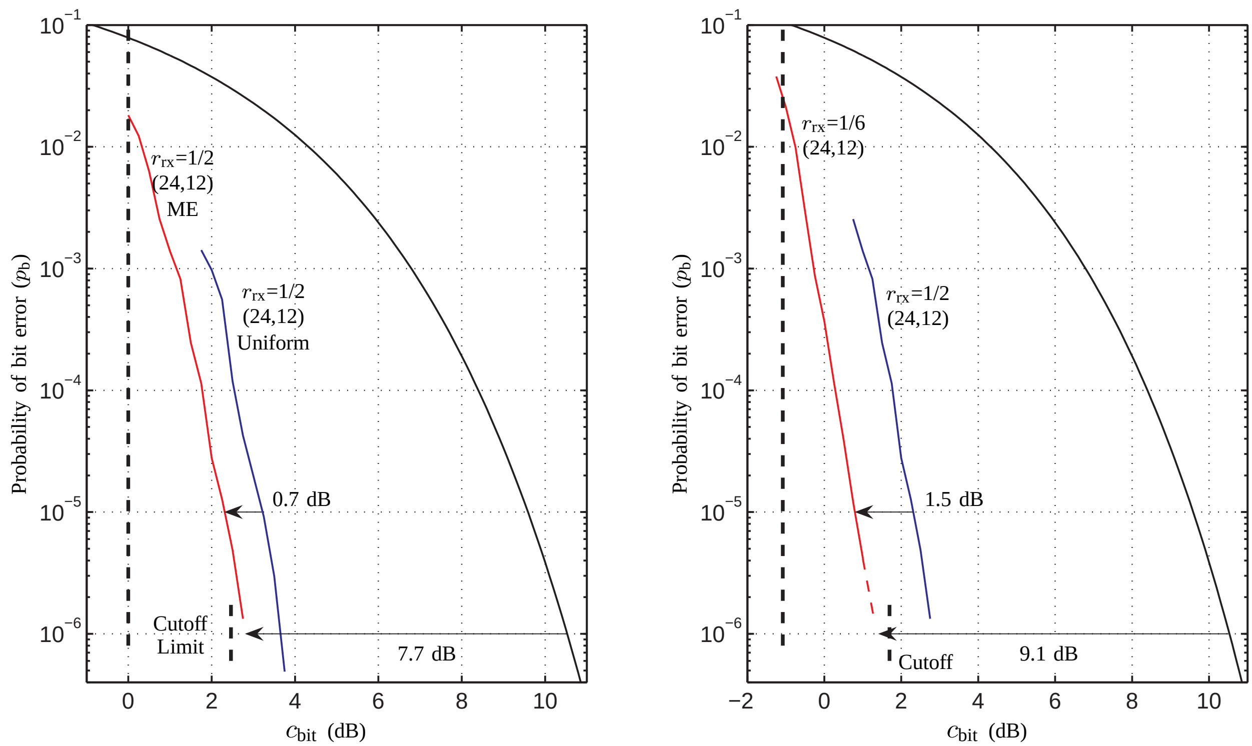


Fig. 6: Receive costs of the (7,3) code using ME sampling and the K=7 conv. code.

The key observation is that a (7,3) code achieves the same receive cost as the much stronger, but fixed-length, K=7 convolutional code. Note that the *transmit* costs of a (7,3) $r_{\text{rx}} = 1/2$ are significantly higher than that of the (2,1,7) code (optimistically, at least 7 dB) and the ME decoder produces its output after a variable delay. Also, while a (7,3) ME



(a) (b)

Fig. 7: Performance of the Golay code (a) Using uniform and ME sampling, and (b) At two different rates using ME sampling (plotted limits refer to r_{rx}=1/6).

decoder is orders of magnitude simpler than a (2,1,7) Viterbi decoder, it is not computationally trivial (more on this later).

Next we consider the (24,12,8) Golay code, chosen for its remarkable coding gain at a small blocklength. When decoded using the ME criterion, it achieves a cost that is 0.3 dB from the cutoff limit (figure 7a). Also, the ME criterion outperforms uniform sampling by about 0.7 dB, which is significant in regimes around the cutoff limit.

Conventional codes that achieve performance comparable to the r_{rx}=1/2 Golay code do so at significantly larger blocklengths. Examples include a rate 0.4 concatenated code with a RS_8(160,128) outer code and a (2,1,K=7) inner code, and various Turbo and LDPC codes with blocklengths on the order of 100s of bits. The significant reduction in blocklength holds even when contemporary feedback techniques are considered. Hagenauer's rate-compatible, punctured convolutional (RCPC) codes enable an ARQ (feedback) scheme where disjoint subsets of coded output bits are successively transmitted until the codeword is successfully decoded [8]. Rowitch and Milstein have generalized this to rate-compatible punctured Turbo (RCPT) codes [9]. Both these schemes require $n \geq 256$ to achieve performance comparable to our scheme at $r_{rx} = 1/2$.

The advantage of ME sampling is even more pronounced for lower receive rates (figure 7b)[4]. This is because increasing the expected number of observations expands the class of atypical channel behavior that sequential schemes may exploit. For r_{rx} =1/6, the Golay code's cost is 0.3 dB lower than the rate 1/6 cutoff limit. It outperforms rate 1/6 RCPC and RCPT codes by about 2 dB, which is rather dramatic considering the operating regime.

The ME algorithm can be implemented by calculating the bit-APPs via the BCJR algorithm running on a minimal trellis. This yields a computational and storage complexity proportional to the number of trellis edges. E.g., ME decoding of the (7,3) code would require on the order of 100 operations (multiplications/additions) per information bit. The comparable number for the Golay code is 10,000! Thus, while ME sampling can reduce blocklengths by orders-of-magnitude, it can incur a prohibitive computational burden in regimes close to capacity, which could negate the energy savings due to reduced sampling. We are currently investigating new codes and various approximations to the ME algorithm to reduce its computational complexity.

[4] Due to computational limitations, part of the curve was extrapolated.

VII. SUMMARY

We turn the classic information theoretic problem on its head and ask what coding techniques work best when cost is dominated by sampling the noisy received signal. This formulation is of interest in short-range, low-data-rate UWB communications devices that must be extremely energy efficient. We propose a sequential scheme that samples the bit with the maximum *a posteriori* entropy at every step and show that it allows reduction of blocklengths by an order-of-magnitude over traditional coding schemes, including those with feedback.

ACKNOWLEDGMENTS

This work was funded in part by an IBM Research Fellowship and grants from the NSF and DARPA HI-MEMS (contract FA8650-07-C-7704). Fred Lee introduced the author to the minimum observation cost synchronization problem. Thanks to Lizhong Zheng and Anant Sahai for suggesting entropy as a criteria for optimum design.

REFERENCES

[1] F. S. Lee and A. Chandrakasan, "A 2.5nJ/b 0.65V 3-to-5GHz Subbanded UWB Receiver in 90nm CMOS," in *International Solid State Circuits Conference*, February 2007, pp. 116–117.

[2] D. Wentzloff and A. Chandrakasan, "A 47pJ/pulse 3.1-to-5GHz All-Digital UWB Transmitter in 90nm CMOS," in *International Solid State Circuits Conference*, February 2007, pp. 118–119.

[3] C. Howland and A. Blanksby, "A 220mW 1Gb/s 1024-Bit Rate-1/2 Low Density Parity Check Code Decoder," in *IEEE 2001 Custom Integrated Circuits Conference (CICC'01)*, 2001, pp. 293–296.

[4] A. Wald, "Sequential method of sampling for deciding between two courses of action," *Journal of the American Statistical Association*, vol. 40, no. 231, pp. 277–306, 1945.

[5] H. Chernoff, "Sequential design of experiments," *Annals of Mathematical Statistics*, vol. 30, pp. pp. 755–770, 1959.

[6] W. Blot and D. Meeter, "Sequential experimental design procedures," *Journal of the American Statistical Association*, vol. 68, no. 343, pp. 586–593, 1973.

[7] G. E. P. Box and W. J. Hill, "Discrimination among mechanistic models," *Technometrics*, vol. 9, no. 1, pp. 57–71, 1967.

[8] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Transactions on Communications*, vol. 36, no. 4, pp. 389–400, 1988.

[9] D. N. Rowitch and L. B. Milstein, "On the performance of hybrid FEC/ARQ systems using rate compatible punctured turbo (RCPT) codes," *IEEE Transactions on Communications*, vol. 48, no. 6, pp. 948–959, 2000.