An Information-Theoretic Characterization of Channels That Die

Lav R. Varshney, *Member, IEEE*, Sanjoy K. Mitter, *Life Fellow, IEEE*, and Vivek K Goyal, *Senior Member, IEEE*

***Abstract*—Given the possibility of communication systems failing catastrophically, we investigate limits to communicating over channels that fail at random times. These channels are finite-state semi-Markov channels. We show that communication with arbitrarily small probability of error is not possible. Making use of results in finite blocklength channel coding, we determine sequences of blocklengths that optimize transmission volume communicated at fixed maximum message error probabilities. We provide a partial ordering of communication channels. A dynamic programming formulation is used to show the structural result that channel state feedback does not improve performance.**

***Index Terms*—Channel coding, communication channels, dynamic programming, finite blocklength regime, reliability theory.**

"a communication channel… might be inoperative because of an amplifier failure, a broken or cut telephone wire,…"

—I. M. Jacobs

I. Introduction

PHYSICAL systems have a tendency to fail at random times. This is true whether considering communication systems embedded in sensor networks that may run out of energy [3], synthetic communication systems embedded in biological cells that may die [4], communication systems embedded in spacecraft that may enter black holes [5], or communication systems embedded in oceans with undersea cables that may be cut [6]. In these scenarios and beyond, failure of the communication system may be modeled as communication channel death.

As such, it is of interest to study information-theoretic limits on communicating over channels that die at random times. This paper gives results on the fundamental limits of what is possible and what is impossible when communicating over channels that die. Communication with arbitrarily small probability of error (*Shannon reliability*) is not possible for any positive communication volume; however, a suitably defined notion of η-reliability is possible. Schemes that optimize communication volume for a given level of η-reliability are developed herein.

The central tradeoff in communicating over channels that die is in the lengths of codeword blocks. Longer blocks improve communication performance as classically known, whereas shorter blocks have a smaller probability of being prematurely terminated due to channel death. In several settings, a simple greedy algorithm for determining the sequence of blocklengths yields a certifiably optimal solution. We also develop a dynamic programming formulation to optimize the ordered integer partition that determines the sequence of blocklengths. Besides algorithmic utility, solving the dynamic program (DP) demonstrates the structural result that channel state feedback does not improve performance.

The optimization of codeword blocklengths is reminiscent of frame size control in wireless networks [7]–[10], however, such techniques are used in conjunction with automatic repeat request protocols and are motivated by amortizing protocol information. Moreover, the results demonstrate the benefit of adapting to either channel state or decision feedback. Contrarily, we show that adaptation to channel state provides no benefit for channels that die.

Limits on channel coding with finite blocklength [11]–[18] are central to our development. Indeed, channels that die bring the notion of finite blocklength to the fore and provide a concrete physical reason to step back from infinity.[1] Notions of outage in wireless communication [19], [20] and lost letters in postal channels [21] are similar to channel death, except that neither outage nor lost letters are permanent conditions. Therefore, blocklength asymptotics are useful to study those channel models but are not useful for channels that die. Recent work that has similar motivations as this paper provides the outage capacity of a wireless channel [22].

The remainder of this paper is organized as follows. Section II defines discrete memoryless channels (DMCs) that die and shows that these channels have zero Shannon capacity. Section III states the communication system model and also fixes our novel performance criteria. Section IV shows that our notion of Shannon reliability is not achievable, strengthening the result of zero Shannon capacity and then provides a communication scheme and determines its performance. Section V optimizes performance for several death distributions using

Manuscript received June 10, 2011; revised May 03, 2012; accepted May 03, 2012. Date of publication May 11, 2012; date of current version August 14, 2012. This material is based upon work supported in part by the National Science Foundation under a Graduate Research Fellowship and Grants 0729069, 0836720, and 0801549, and in part by a Siemens Corporate Research Grant, "Advanced Control Methods for Complex Networked Systems." This paper was presented in part at the 47th Annual Allerton Conference on Communication, Control, and Computing, 2009.

L. R. Varshney was with the Department of Electrical Engineering and Computer Science, the Research Laboratory of Electronics, and the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. He is now with the IBM Thomas J. Watson Research Center, Hawthorne, NY 10532 USA (e-mail: varshney@alum.mit.edu).

S. K. Mitter is with the Department of Electrical Engineering and Computer Science, the Engineering Systems Division, and the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: mitter@mit.edu).

V. K Goyal is with the Department of Electrical Engineering and Computer Science and the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: vgoyal@mit.edu).

Communicated by I. Kontoyiannis, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2012.2199078

[1]The phrase "back from infinity" is borrowed from J. Ziv's 1997 Shannon Lecture.

either a greedy algorithm or a dynamic programming algorithm. Optimization demonstrates that channel state feedback does not improve performance. Section VI discusses the partial ordering of channels. Section VII suggests several extensions to this work.

II. Channel Model

Consider a channel with finite input alphabet $\mathcal{X}$ and finite output alphabet $\mathcal{Y}$. It has an *alive* state $s = a$ when it acts like a noisy DMC and a *dead* state $s = d$ when it erases the input.[2] Assume throughout the paper that the DMC from the alive state has zero error capacity [24] equal to zero.[3]

For example, if the channel acts like a binary symmetric channel (BSC) with crossover probability $0 < \varepsilon < 1$ in the alive state, with $\mathcal{X} = \{0,1\}$, and $\mathcal{Y} = \{0,1,?\}$, then the transmission matrix in the alive state is

$$p(y|x, s = a) = p_a(y|x) = \begin{bmatrix} 1-\varepsilon & \varepsilon & 0 \\ \varepsilon & 1-\varepsilon & 0 \end{bmatrix} \quad (1)$$

and the transmission matrix in the dead state is

$$p(y|x, s = d) = p_d(y|x) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}. \quad (2)$$

The channel starts in state $s = a$ and then transitions to $s = d$ at some random time T, where it remains for all time thereafter. That is, the channel is in state a for times $n = 1, 2, \ldots, T$ and in state d for times $n = T+1, T+2, \ldots$. The death time distribution is denoted $p_T(t)$. Note that there is always a finite $t^\dagger$ such that $p_T(t^\dagger) > 0$.

A. Finite-State Semi-Markov Channel

Channels that die can be classified as finite-state channels (FSCs) [27, Sec. 4.6].

Proposition 1: A channel that dies $(\mathcal{X}, p_a(y|x), p_d(y|x), p_T(t), \mathcal{Y})$ is an FSC.

Proof: Follows by definition, since the channel has two states. ∎

Channels that die have semi-Markovian [28, Sec. 4.8], [29, Sec. 5.7] properties.

Definition 1: A semi-Markov process changes state according to a Markov chain but takes a random amount of time between changes. More specifically, it is a stochastic process with states from a discrete alphabet $\mathcal{S}$, such that whenever it enters state s, $s \in \mathcal{S}$:

1) The next state it will enter is state r with probability that depends only on $s, r \in \mathcal{S}$.
2) Given that the next state to be entered is state r, the time until the transition from s to r occurs has distribution that depends only on $s, r \in \mathcal{S}$.

[2]Our results can be extended to cover cases where the channel acts as other channels, such as Gaussian or Gilbert-Elliott channels [15], [23], in the alive state.

[3]If the channel is noiseless in the alive state, the problem is similar to settings where fountain codes [25] are used in the point-to-point case and growth codes [26] are used in the network case.

Definition 2: The Markovian sequence of states of a semi-Markov process is called the embedded Markov chain of the semi-Markov process.

Definition 3: A semi-Markov process is irreducible if its embedded Markov chain is irreducible.

Proposition 2: A channel that dies $(\mathcal{X}, p_a(y|x), p_d(y|x), p_T(t), \mathcal{Y})$ has a channel state sequence that is a nonirreducible semi-Markov process.

Proof: When in state a, the next state is d with probability 1 and given that the next state is to be d, the time until the transition from a to d has distribution $p_T(t)$. When in state d, the next state is d with probability 1. Thus, the channel state sequence is a semi-Markov process.

The semi-Markov state process is not irreducible because the a state of the embedded Markov chain is transient. ∎

Note that when T is a geometric random variable, the channel state process forms a Markov chain, with transient state a and recurrent, absorbing state d.

There are further special classes of FSCs.

Definition 4: An FSC is a finite-state semi-Markov channel (FSSMC) if its state sequence forms a semi-Markov process.

Definition 5: An FSC is a finite-state Markov channel (FSMC) if its state sequence forms a Markov chain.

Proposition 3: A channel that dies $(\mathcal{X}, p_a(y|x), p_d(y|x), p_T(t), \mathcal{Y})$ is an FSSMC and is an FSMC when T is geometrically distributed.

Proof: Follows from Propositions 1 and 2. ∎

FSMCs have been widely studied in the literature [27], [30], [31], particularly the panic button/child's toy channel of Gallager [30, p. 26], [27, p. 103] and the Gilbert-Elliott channel and its extensions [32], [33].

Contrarily, FSSMCs seem not to have been specifically studied in information theory. There are a few works [34]–[36] that give semi-Markov channel models for wireless communications systems but do not provide information-theoretic characterizations.

B. Capacity is Zero

A channel that dies has Shannon capacity equal to zero. To show this, first notice that if the initial state of a channel that dies were not fixed, then it would be an indecomposable FSC [27, Sec. 4.6], where the effect of the initial state dies away.

Proposition 4: If the initial state of a channel that dies $(\mathcal{X}, p_a(y|x), p_d(y|x), p_T(t), \mathcal{Y})$ is not fixed, then it is an indecomposable FSC.

Proof: The embedded Markov chain for a channel that dies has a unique absorbing state d. ∎

Indecomposable FSCs have the property that the upper capacity, defined in [27, eqs. (4.6.6) and (4.6.7)], and lower capacity, defined in [27, eqs. (4.6.3) and (4.6.4)], are identical [27, Th. 4.6.4]. This can be used to show that the capacity of a channel that dies is zero.

Proposition 5: The Shannon capacity, C, of a channel that dies $(\mathcal{X}, p_a(y|x), p_d(y|x), p_T(t), \mathcal{Y})$ is zero.

Proof: Although the initial state is $s_1 = a$ here, temporarily suppose that s_1 may be either a or d. Then the channel is indecomposable by Proposition 4.

The lower capacity $\underline{C}$ equals the upper capacity $\overline{C}$, for indecomposable channels by [27, Th. 4.6.4]. The information rate of a memoryless $p_d(y|x)$ "dead" channel is clearly zero for any input distribution, so the lower capacity $\underline{C} = 0$. Thus, the Shannon capacity for a channel that dies with initial alive state is $C = \overline{C} = 0$. ■

III. Communication System

In order to provide an information-theoretic characterization of a channel that dies, a communication system that contains the channel is described. We start with a general formulation and then impose some restrictions to define performance criteria.

A. General System

We have an information stream (like i.i.d. equiprobable bits) that is to be transmitted to a receiver over a channel that dies. An encoder maps the information stream into a codeword X_1^n and a noisy version of this codeword is received Y_1^n, with received symbols ? after channel death. Due to channel death, the transmitted codeword can be thought of as randomly truncated and must be decoded at that truncation length. Thus, the decoder is a sequence of finite blocklength decoders, one for each possible truncation length, charged with recovering X_1^n or a function of X_1^n.

Note that the encoder mapping may be a tree code, a block code, or some other style of code. Going forward, we restrict attention to constructing codeword X_1^n as a concatenation of shorter codewords that each represent individual messages.

B. Sequences of Messages

Now consider that the information stream can be grouped into a sequence of k messages, $(W_1, W_2, \ldots, W_k)$. Each message W_i is drawn from a message set $\mathcal{W}_i = \{1, 2, \ldots, M_i\}$. Each message W_i is encoded into a channel input codeword $X_1^{n_i}(W_i)$ and these codewords $(X_1^{n_1}(W_1), X_1^{n_2}(W_2), \ldots, X_1^{n_k}(W_k))$ are transmitted in sequence over the channel. A noisy version of this codeword sequence is received, $Y_1^{n_1+n_2+\cdots+n_k}(W_1, W_2, \ldots, W_k)$. The receiver then guesses the sequence of messages using an appropriate decoding rule g, to produce $(\hat{W}_1, \hat{W}_2, \ldots, \hat{W}_k) = g(Y_1^{n_1+n_2+\cdots+n_k})$. The $\hat{W}_i$s are drawn from alphabets $\mathcal{W}_i^{\ominus} = \mathcal{W}_i \cup \ominus$, where the $\ominus$ message indicates the decoder declaring an erasure. The receiver makes an error on message i if $\hat{W}_i \neq W_i$ and $\hat{W}_i \neq \ominus$.

Block coding results are typically expressed with the concern of sending one message rather than k messages as here.[4]

System definitions can be formalized as follows.

Definition 6: An (M_i, n_i) individual message code for a channel that dies $(\mathcal{X}, p_a(y|x), p_d(y|x), p_T(t), \mathcal{Y})$ consists of:

1) an individual message index set $\{1, 2, \ldots, M_i\}$; and
2) an individual message encoding function $f_i : \{1, 2, \ldots, M_i\} \mapsto \mathcal{X}^{n_i}$.

The individual message index set $\{1, 2, \ldots, M_i\}$ is denoted $\mathcal{W}_i$, and the set of individual message codewords $\{f_i(1), f_i(2), \ldots, f_i(M_i)\}$ is called the individual message codebook.

Definition 7: An $(M_i, n_i)_{i=1}^k$ code for a channel that dies $(\mathcal{X}, p_a(y|x), p_d(y|x), p_T(t), \mathcal{Y})$ is a sequence of k individual message codes, $(M_i, n_i)_{i=1}^k$, in the sense of comprising:

1) a sequence of individual message index sets $\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_k$;
2) a sequence of individual message encoding functions $f = (f_1, f_2, \ldots, f_k)$; and
3) a decoding function $g : \mathcal{Y}^{\sum_{i=1}^k n_i} \mapsto \mathcal{W}_1^{\ominus} \times \mathcal{W}_2^{\ominus} \times \cdots \times \mathcal{W}_k^{\ominus}$.

There is no essential loss of generality by assuming that the decoding function g is decomposed into a sequence of individual message decoding functions $g = (g_1, g_2, \ldots, g_n)$ where $g_i : \mathcal{Y}^{n_i} \mapsto \mathcal{W}_i^{\ominus}$ when individual messages are chosen independently, due to this independence and the conditional memorylessness of the channel.

To define performance measures, we assume that the decoder operates on an individual message basis. That is, when applying the communication system, let $\hat{W}_1 = g_1(Y_1^{n_1})$, $\hat{W}_2 = g_2(Y_{n_1+1}^{n_1+n_2})$, and so on.

For the sequel, we make a further assumption on the operation of the decoder. This assumption corresponds to the physical properties of a communication system where the decoder fails catastrophically. Once the decoder fails, it cannot perform any decoding operations, and so the ? symbols in the channel model of system failure must be ignored.

Assumption 1: If all n_i channel output symbols used by individual message decoder g_i are not ?, then the range of g_i is $\mathcal{W}_i$. If any of the n_i channel output symbols used by individual message decoder g_i are ?, then g_i maps to $\ominus$.

C. Performance Measures

We formally write the notion of error for the communication system as follows.

Definition 8: For all $1 \leq w \leq M_i$, let

$$\lambda_w(i) = \Pr[\hat{W}_i \neq w | W_i = w, \hat{W}_i \neq \ominus]$$

be the conditional message probability of error given that the ith individual message is w.

Definition 9: The maximal probability of error for an (M_i, n_i) individual message code is

$$\lambda_{\max}(i) = \max_{w \in \mathcal{W}_i} \lambda_w(i).$$

Definition 10: The maximal probability of error for an $(M_i, n_i)_{i=1}^k$ code is

$$\lambda_{\max} = \max_{i \in \{1,\ldots,k\}} \lambda_{\max}(i).$$

[4]An alternate formulation of communicating over channels that die using tree codes [37, Ch. 10] with random truncation would also be interesting. In fact, communicating over channels that die using convolutional codes with sequential decoding would be very natural, but would require performance criteria different from the ones developed herein.

Performance criteria weaker than traditional in information theory are defined, since the Shannon capacity of a channel that dies is zero (Proposition 5). In particular, we define formal notions of how much information is transmitted using a code and how long it takes.

Definition 11: The transmission time of an $(M_i, n_i)_{i=1}^k$ code is $N = \sum_{i=1}^k n_i$.

Definition 12: The expected transmission volume of an $(M_i, n_i)_{i=1}^k$ code is

$$V = \mathrm{E}_T \left\{ \sum_{i \in \{1,\ldots,k | \hat{W}_i \neq \ominus\}} \log M_i \right\}.$$

Notice that although declared erasures do not lead to errors, they do not contribute transmission volume either.

The several performance criteria for a code may be combined together.

Definition 13: Given $0 \leq \eta < 1$, a pair of numbers (N_0, V_0) (where N_0 is a positive integer and V_0 is nonnegative) is said to be an achievable transmission time-volume at η-reliability if there exists, for some k, an $(M_i, n_i)_{i=1}^k$ code for the channel that dies $(\mathcal{X}, p_a(y|x), p_d(y|x), p_T(t), \mathcal{Y})$ such that

$$\lambda_{\max} \leq \eta, \quad (3)$$
$$N \leq N_0, \text{ and} \quad (4)$$
$$V \geq V_0. \quad (5)$$

Moreover, (N_0, V_0) is said to be an achievable transmission time-volume at Shannon reliability if it is an achievable transmission time-volume at η-reliability for all $0 < \eta < 1$.

IV. Limits on Communication

Having defined the notion of achievable transmission time-volume at various levels of reliability, the goal of this study is to demarcate what is achievable.

A. Shannon Reliability is not Achievable

Not only is the Shannon capacity of a channel that dies zero, but also there is no $V > 0$ such that (N, V) is an achievable transmission time-volume at Shannon reliability. A coding scheme that always declares erasures would achieve zero error probability (and therefore Shannon reliability) but would not provide positive transmission volume; this is also not allowed under Assumption 1.

Lemmas are stated and proved after the proof of the main proposition. For brevity, the proof is limited to the alive-BSC case, but can be extended to general alive-DMCs by choosing the two most distant letters in $\mathcal{Y}$ for constructing the repetition code, among other things.

Proposition 6: For a channel that dies $(\mathcal{X}, p_a(y|x), p_d(y|x), p_T(t), \mathcal{Y})$, there is no $V > 0$ such that (N, V) is an achievable transmission time-volume at Shannon reliability.

Proof: From the error probability viewpoint, transmitting longer codes is not harder than transmitting shorter codes (Lemma 1) and transmitting smaller codes is not harder than transmitting larger codes (Lemma 2). Hence, the desired result follows by showing that even the longest and smallest code that has positive expected transmission volume cannot achieve Shannon reliability.

Clearly, the longest and smallest code uses a single individual message code of length $n_1 \to \infty$ and size $M_1 = 2$. Among such codes, transmitting the binary repetition code is not harder than transmitting any other code (Lemma 3). Hence, showing that the binary repetition code cannot achieve Shannon reliability yields the desired result.

Consider transmitting a single $(M_1 = 2, n_1)$ individual message code that is simply a binary repetition code over a channel that dies $(\mathcal{X}, p_a(y|x), p_d(y|x), p_T(t), \mathcal{Y})$.

Let $\mathcal{W}_1 = \{00000\ldots, 11111\ldots\}$, where the two codewords are of length n_1. Assume that the all-zeros codeword and the all-ones codeword are each transmitted with probability 1/2 and measure average probability of error, since average error probability lower bounds $\lambda_{\max}(1)$ [27, Problem 5.32]. The transmission time $N = n_1$ and let $N \to \infty$. The expected transmission volume is $\log 2 > 0$.

Under equiprobable signaling over a BSC, the minimum error probability decoder is the maximum likelihood decoder, which in turn is the minimum distance decoder [38, Problem 2.13].

The scenario corresponds to binary hypothesis testing over a $\mathrm{BSC}(\varepsilon)$ with T observations (since after the channel dies, the output symbols do not help with hypothesis testing). Since there is a finite $t^\dagger$ such that $p_T(t^\dagger) > 0$, there is a fixed constant K such that $\lambda_{\max} > K > 0$ for any realization $T = t$.

Thus, Shannon reliability is not achievable. ■

Lemma 1: When transmitting over the alive state's memoryless channel $p_a(y|x)$, let the maximal probability of error $\lambda_{\max}(i)$ for an optimal (M_i, n_i) individual message code and minimum probability of error individual decoder g_i be $\lambda_{\max}(i; n_i)$. Then, $\lambda_{\max}(i; n_i + 1) \leq \lambda_{\max}(i; n_i)$.

Proof: Consider the optimal block-length-n_i individual message code/decoder, which achieves $\lambda_{\max}(i; n_i)$. Use it to construct an $n_i + 1$ individual message code that appends a dummy symbol to each codeword and an associated decoder that operates by ignoring this last symbol. The error performance of this (suboptimal) code/decoder is clearly $\lambda_{\max}(i; n_i)$, and so the optimal performance can only be better: $\lambda_{\max}(i; n_i + 1) \leq \lambda_{\max}(i; n_i)$. ■

Lemma 2: When transmitting over the alive state's memoryless channel $p_a(y|x)$, let the maximal probability of error $P_e^{\max}(i)$ for an optimal (M_i, n_i) individual message code and minimum probability of error individual decoder $f_D^{(i)}$ be $P_e^{\max}(i; M_i)$. Then $P_e^{\max}(i; M_i) \leq P_e^{\max}(i; M_i + 1)$.

Proof: Follows from sphere-packing principles. ■

Lemma 3: When transmitting over the alive state's memoryless channel $p_a(y|x)$, the optimal $(M_i = 2, n_i)$ individual message code can be taken as a binary repetition code.

Proof: Under minimum distance decoding (which yields the minimum error probability [38, Problem 2.13]) for a code transmitted over a BSC, increasing the distance between codewords can only reduce error probability. The repetition code has maximum Hamming distance between codewords. ■

Notice that Proposition 6 also directly implies Proposition 5, providing an alternative proof.

Corollary 1: The Shannon capacity of a channel that dies $(\mathcal{X}, p_a(y|x), p_d(y|x), p_T(t), \mathcal{Y})$ is zero.

B. Finite Blocklength Channel Coding

Before developing an optimal scheme for η-reliable communication over a channel that dies, finite blocklength channel coding is reviewed.

Under our definitions, traditional channel coding results [11], [13]–[18] provide information about individual message codes, determining the achievable trios $(n_i, M_i, \lambda_{\max}(i))$. In particular, the largest possible M_i for a given n_i and $\lambda_{\max}(i)$ is denoted $M^*(n_i, \lambda_{\max}(i))$.

The purpose of this study is not to improve upper and lower bounds on finite blocklength channel coding, but to use existing results to study channels that die. In fact, for the sequel, simply assume that the function $M^*(n_i, \lambda_{\max}(i))$ is known, as are codes/decoders that achieve this value. In principle, optimal individual message codes may be found through exhaustive search [13], [39]. Although algebraic notions of code quality do not directly imply error probability quality [40], perfect codes such as the Hamming or Golay codes may also be optimal in certain limited cases.

Recent results comparing upper and lower bounds around Strassen's normal approximation to $\log M^*(n_i, \lambda_{\max}(i))$ [41] have demonstrated that the approximation is quite good [15].

Remark 1: We assume that optimal $M^*(n_i, \eta)$-achieving individual message codes are known. Exact upper and lower bounds to $\log M^*(n_i, \eta)$ can be substituted to make our results precise. For numerical demonstrations, we will further assume that optimal codes have performance given by Strassen's approximation.

The following expression for $\log M^*(n_i, \eta)$ that first appeared in [41] is also given in [15, eq. (54)].

Lemma 4: Let $M^*(n_i, \eta)$ be the largest size of an individual message code with blocklength n_i and maximal error probability upper bounded by $\lambda_{\max}(i) < \eta$. Then, for any DMC with capacity C and $0 < \eta \le 1/2$

$$\log M^*(n_i, \eta) = n_i C - \sqrt{n_i \rho} Q^{-1}(\eta) + O(\log n_i)$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\, dt$$

$$\rho = \min_{p_X : D(p_X p_{Y|X} \| p_X p_Y) = C} \mathrm{var}\left[\log \frac{p_{Y|X}(y|x)}{p_Y(y)}\right]$$

and standard asymptotic notation is used.

For the $\mathrm{BSC}(\varepsilon)$, the approximation (ignoring the $O(\log n_i)$ term above) is

$$\log M^* \approx n_i(1 - h_2(\varepsilon)) - \sqrt{n_i \varepsilon (1-\varepsilon)} Q^{-1}(\eta) \log_2 \tfrac{1-\varepsilon}{\varepsilon} \quad (6)$$

where $h_2(\cdot)$ is the binary entropy function to the base 2. This BSC expression first appeared in [42].

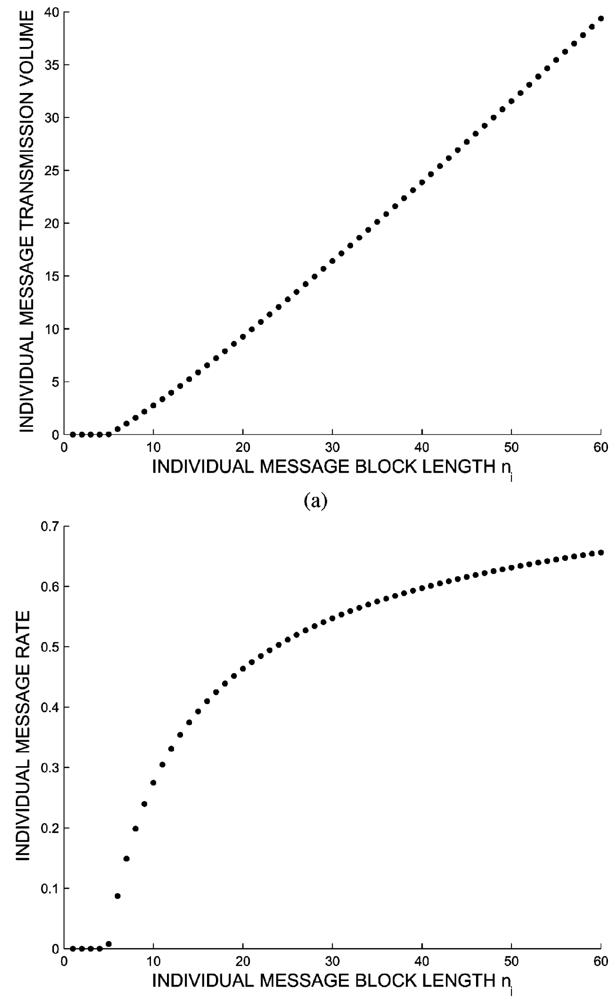

Fig. 1. (a) Expression (6) for $\varepsilon = 0.01$ and $\eta = 0.001$. (b) Normalized version, $(\log M^*(n_i, \eta))/n_i$, for $\varepsilon = 0.01$ and $\eta = 0.001$. The capacity of a $\mathrm{BSC}(\varepsilon)$ is $1 - h_2(\varepsilon) = 0.92$.

For intuition, we plot the approximate $\log M^*(n_i, \eta)$ function for a $\mathrm{BSC}(\varepsilon)$ in Fig. 1(a). Notice that $\log M^*$ is zero for small n_i since no code can achieve the target error probability η. Also notice that $\log M^*$ is a monotonically increasing function of n_i. Moreover, notice in Fig. 1(b) that even when normalized, $(\log M^*)/n_i$, is a monotonically increasing function of n_i. Therefore longer blocks provide more "bang for the buck." The curve in Fig. 1(b) asymptotically approaches capacity.

C. η-Reliable Communication

We now describe a coding scheme that achieves positive expected transmission volume at η-reliability. Survival probability of the channel plays a key role in measuring performance.

Definition 14: The survival function of a channel that dies $(\mathcal{X}, p_a(y|x), p_d(y|x), p_T(t), \mathcal{Y})$ is $\Pr[T > t]$, is denoted $R_T(t)$, and satisfies

$$R_T(t) = \Pr[T > t] = 1 - \sum_{\tau=1}^{t} p_T(\tau) = 1 - F_T(t)$$

where F_T is the cumulative distribution function.

$R_T(t)$ is a nonincreasing function.

Proposition 7: The transmission time-volume

$$\left(N = \sum_{i=1}^{k} n_i, V = \sum_{i=1}^{k} R_T(e_i) \log M^*(n_i, \eta)\right)$$

is achievable at η-reliability for any sequence $(n_i)_{i=1}^k$ of individual message codeword lengths, where $e_0 = 0, e_1 = n_1, e_2 = n_1 + n_2, \ldots, e_k = \sum_{i=1}^k n_i$.

Proof: Code Design: A target error probability η and a sequence $(n_i)_{i=1}^k$ of individual message codeword lengths are fixed. Construct a length-k sequence of (M_i, n_i) individual message codes and individual decoding functions $(\mathcal{W}_i, f_i, g_i)$ that achieve optimal performance. The size of $\mathcal{W}_i$ is $|\mathcal{W}_i| = \log M^*(n_i, \eta)$. Note that individual decoding functions g_i have range $\mathcal{W}_i$ rather than $\mathcal{W}_i^{\ominus}$.

Encoding: A codeword $W_1 = w_1$ is selected uniformly at random from the codebook $\mathcal{W}_1$. The mapping of this codeword into n_1 channel input letters, $X_{e_0+1}^{e_1} = f_1(w_1)$, is transmitted in channel usage times $n = e_0 + 1, e_0 + 2, \ldots, e_1$.

Then, a codeword $W_2 = w_2$ is selected uniformly at random from the codebook $\mathcal{W}_2$. The mapping of this codeword into n_2 channel input letters, $X_{e_1+1}^{e_2} = f_2(w_2)$, is transmitted in channel usage times $n = e_1 + 1, e_1 + 2, \ldots, e_2$.

This procedure continues until the last individual message code in the code is transmitted. That is, a codeword $W_k = w_k$ is selected uniformly at random from the codebook $\mathcal{W}_k$. The mapping of this codeword into n_k channel input letters, $X_{e_{k-1}+1}^{e_k} = f_k(w_k)$, is transmitted in channel usage times $n = e_{k-1} + 1, e_{k-1} + 2, \ldots, e_k$.

We refer to channel usage times $n \in \{e_{i-1} + 1, e_{i-1} + 2, \ldots, e_i\}$ as the ith transmission epoch.

Decoding: For decoding, the channel output symbols for each epoch are processed separately. If any of the channel output symbols in an epoch are erasure symbols ?, then a decoding erasure $\ominus$ is declared for the message in that epoch, i.e., $\hat{W}_i = \ominus$. Otherwise, the individual message decoding function $g_i : \mathcal{Y}^{n_i} \to \mathcal{W}_i$ is applied to obtain $\hat{W}_i = g_i(Y_{e_{i-1}+1}^{e_i})$.

Performance Analysis: Having defined the communication scheme, we measure the error probability, transmission time, and expected transmission volume.

The decoder will either produce an erasure $\ominus$ or use an individual message decoder g_i. When g_i is used, the maximal error probability of individual message code error is bounded as $\lambda_{\max}(i) < \eta$ by construction. Since declared erasures $\ominus$ do not lead to error, and since all $\lambda_{\max}(i) < \eta$, it follows that

$$\lambda_{\max} < \eta.$$

The transmission time is simply $N = \sum n_i$.

Recall the definition of expected transmission volume:

$$\mathrm{E}\left\{ \sum_{i \in \{1,\ldots,k | \hat{W}_i \neq \ominus\}} \log M_i \right\} = \sum_{i \in \{1,\ldots,k | \hat{W}_i \neq \ominus\}} \mathrm{E}\left\{\log M_i\right\}$$

and the fact that the channel produces the erasure symbol ? for all channel usage times after death, $n > T$, but not before. Combining this with the length of an optimal code, $\log M^*(n_i, \eta)$, leads to the expression

$$\sum_{i=1}^{k} \Pr[T > e_i] \log M^*(n_i, \eta)$$

since all individual message codewords that are received in their entirety before the channel dies are decoded using g_i whereas any individual message codewords that are even partially cut off are declared $\ominus$.

Recalling the definition of the survival function, the expected transmission volume of the communication scheme is

$$\sum_{i=1}^{k} R_T(e_i) \log M^*(n_i, \eta)$$

as desired. ■

Proposition 7 is valid for any choice of $(n_i)_{i=1}^k$. Since $(\log M^*)/n_i$ is monotonically increasing, it is better to use individual message codes that are as long as possible. With longer individual message codes, however, there is a greater chance of many channel usages being wasted if the channel dies in the middle of transmission. The basic tradeoff is captured in picking the set of values $\{n_1, n_2, \ldots, n_k\}$. For fixed and finite N, this involves picking an ordered integer partition $n_1 + n_2 + \cdots + n_k = N$. We optimize this choice in Section V.

D. Converse Arguments

Since we simply have operational expressions and no informational expressions in our development, as per Remark 1, and since optimal individual message codes and individual message decoders are assumed to be used, it seems that converse arguments are not required. Indeed due to Assumption 1, Proposition 7 gives the best performance possible.

In particular, optimality follows from two facts stemming from Assumption 1. First, that the last partially erased message block cannot be decoded (due to the physical modeling of system failure). Second, that errors-and-erasures decoding [43] by the g_i for codewords that are received before channel death is not allowed.

One might wonder whether the possibility of errors-and-erasures decoding by the individual message decoders need be explicitly restricted.

Let $M^*(n_i, \xi_i, \eta)$ be the maximum individual message codebook size under erasure probability ξ_i and maximum error probability η. Then, at the level of Strassen's approximation (up to the $\log n$ term), $\log M^*(n_i, \xi_i, \eta)$ and $\log M^*(n_i, \eta)$ are the same [44, Th. 47]. Hence, there is little, if any, benefit to errors-and-erasures decoding. A precise characterization, however, requires precise knowledge of specific optimal codes obtainable through large-scale combinatorial enumeration.

V. Optimizing the Communication Scheme

In Section IV-C, we had not optimized the lengths of the individual message codes; we do so here. For fixed η and N, we maximize the expected transmission volume V over the choice of the ordered integer partition $n_1 + n_2 + \cdots + n_k = N$:

$$\max_{(n_i)_{i=1}^k : \sum n_i = N} \sum_{i=1}^{k} R_T(e_i) \log M^*(n_i, \eta). \tag{7}$$

For finite N, this optimization can be carried out by an exhaustive search over all 2^{N-1} ordered integer partitions. If the death distribution $p_T(t)$ has finite support, there is no loss of generality in considering only finite N. Since exhaustive search has exponential complexity, however, there is value in trying to use a simplified algorithm. A dynamic programming formulation for the finite horizon case is developed in Section V-C. Section V-A develops a greedy algorithm which is applicable to both the finite and infinite horizon cases and yields the optimal solution for certain problems.

A. Greedy Algorithm

To try to solve the optimization problem (7), we propose a greedy algorithm that optimizes blocklengths n_i one by one.

Algorithm 1:

1) Maximize $R_T(n_1) \log M^*(n_1, \eta)$ through the choice of n_1 independently of any other n_i.
2) Maximize $R_T(e_2) \log M^*(n_2, \eta)$ after fixing $e_1 = n_1$, but independently of later n_i.
3) Maximize $R_T(e_3) \log M^*(n_3, \eta)$ after fixing e_2, but independently of later n_i.
4) Continue in the same manner for all subsequent n_i.

Sometimes the algorithm produces the correct solution.

Proposition 8: The solution produced by the greedy algorithm, (n_i), is locally optimal if

$$\frac{R_T(e_i) \log M^*(n_i, \eta) - R_T(e_i - 1) \log M^*(n_i - 1, \eta)}{R_T(e_{i+1}) \left[\log M^*(n_{i+1} + 1, \eta) - \log M^*(n_{i+1}, \eta)\right]} \geq 1 \tag{8}$$

for each i.

Proof: The solution of the greedy algorithm partitions time using a set of epoch boundaries (e_i). The proof proceeds by testing whether local perturbation of an arbitrary epoch boundary can improve performance. There are two possible perturbations: a shift to the left or a shift to the right.

First consider shifting an arbitrary epoch boundary e_i to the right by one. This makes the left epoch longer and the right epoch shorter. Lengthening the left epoch does not improve performance due to the greedy optimization of the algorithm. Shortening the right epoch does not improve performance since $R_T(e_i)$ remains unchanged whereas $\log M^*(n_i, \eta)$ does not increase since $\log M^*$ is a nondecreasing function of n_i.

Now, consider shifting an arbitrary epoch boundary e_i to the left by one. This makes the left epoch shorter and the right epoch longer. Reducing the left epoch will not improve performance due to greediness, but enlarging the right epoch might improve performance, so the gain and loss must be balanced.

The loss in performance (a positive quantity) for the left epoch is

$$\Delta_l = R_T(e_i) \log M^*(n_i, \eta) - R_T(e_i - 1) \log M^*(n_i - 1, \eta)$$

whereas the gain in performance (a positive quantity) for the right epoch is

$$\Delta_r = R_T(e_{i+1}) \left[\log M^*(n_{i+1} + 1, \eta) - \log M^*(n_{i+1}, \eta)\right].$$

If $\Delta_l \geq \Delta_r$, then perturbation will not improve performance. The condition may be rearranged as

$$\frac{R_T(e_i) \log M^*(n_i, \eta) - R_T(e_i - 1) \log M^*(n_i - 1, \eta)}{R_T(e_{i+1}) \left[\log M^*(n_{i+1} + 1, \eta) - \log M^*(n_{i+1}, \eta)\right]} \geq 1.$$

This is condition (8), so the left-perturbation does not improve performance. Hence, the solution produced by the greedy algorithm is locally optimal. ■

Proposition 9: The solution produced by the greedy algorithm, (n_i), is globally optimal if

$$\frac{R_T(e_i) \log M^*(n_i, \eta) - R_T(e_i - K_i) \log M^*(n_i - K_i, \eta)}{R_T(e_{i+1}) \left[\log M^*(n_{i+1} + K_i, \eta) - \log M^*(n_{i+1}, \eta)\right]} \geq 1 \tag{9}$$

for each i, and any nonnegative integers $K_i \leq n_i$.

Proof: The result follows by repeating the argument for local optimality in Proposition 8 for shifts of any admissible size K_i. ■

There is an easily checked special case of global optimality condition (9) under Strassen's approximation, given in the forthcoming Proposition 10.

Lemma 5: The function $\log M_S^*(z, \eta) - \log M_S^*(z - K, \eta)$ is a nondecreasing function of z for any $0 \leq K \leq z$, where

$$\log M_S^*(z, \eta) = zC - \sqrt{z\rho} Q^{-1}(\eta) \tag{10}$$

is Strassen's approximation.

Proof: Essentially follows from the fact that $\sqrt{z}$ is a concave ∩ function in z. More specifically $\sqrt{z}$ satisfies

$$-\sqrt{z} + \sqrt{z - K} \leq -\sqrt{z+1} + \sqrt{z + 1 - K}$$

for $K \leq z$. This implies that

$$\begin{aligned} &- \sqrt{z}\sqrt{\rho} Q^{-1}(\eta) + \sqrt{z - K}\sqrt{\rho} Q^{-1}(\eta) \\ &\quad \leq -\sqrt{z+1}\sqrt{\rho} Q^{-1}(\eta) + \sqrt{z+1-K}\sqrt{\rho} Q^{-1}(\eta). \end{aligned}$$

Adding the positive constant KC to both sides, in the form $zC - zC + KC$ on the left and in the form $(z+1)C - (z+1)C + KC$ on the right yields

$$\begin{aligned} &zC - \sqrt{z\rho} Q^{-1}(\eta) - (z - K)C + \sqrt{z - K}\sqrt{\rho} Q^{-1}(\eta) \\ &\quad \leq (z+1)C - \sqrt{z+1}\sqrt{\rho} Q^{-1}(\eta) - (z + 1 - K)C \\ &\qquad + \sqrt{z + 1 - K}\sqrt{\rho} Q^{-1}(\eta) \end{aligned}$$

and so

$$[\log M_S^*(z,\eta) - \log M_S^*(z-K,\eta)] \\ \leq [\log M_S^*(z+1,\eta) - \log M_S^*(z+1-K,\eta)].$$

■

Proposition 10: If the solution produced by the greedy algorithm using Strassen's approximation (10) satisfies $n_1 \geq n_2 \geq \cdots \geq n_k$, then condition (9) for global optimality is satisfied.

Proof: Since $R_T(\cdot)$ is a nonincreasing survival function

$$R_T(e_i - K) \geq R_T(e_{i+1}) \tag{11}$$

for the nonnegative integer K. Since the function $[\log M_S^*(z,\eta) - \log M_S^*(z-K,\eta)]$ is a nondecreasing function of z by Lemma 5, and since the n_i are in nonincreasing order

$$\log M_S^*(n_i,\eta) - \log M_S^*(n_i - K,\eta) \\ \geq \log M_S^*(n_{i+1}+K,\eta) - \log M_S^*(n_{i+1},\eta). \tag{12}$$

Taking products of (11) and (12) and rearranging yields the condition

$$\frac{R_T(e_i - K)\,[\log M_S^*(n_i,\eta) - \log M_S^*(n_i-K,\eta)]}{R_T(e_{i+1})\,[\log M_S^*(n_{i+1}+K,\eta) - \log M_S^*(n_{i+1},\eta)]} \geq 1.$$

Since $R_T(\cdot)$ is a nonincreasing survival function

$$R_T(e_i - K) \geq R_T(e_i) \geq R_T(e_{i+1}).$$

Therefore, the global optimality condition (9) is also satisfied, by substituting $R_T(e_i)$ for $R_T(e_i - K)$ in one place. ■

B. Geometric Death Distribution

A common failure mode for systems that do not age is a geometric death time T [45]

$$p_T(t) = \alpha(1-\alpha)^{(t-1)}$$

and

$$R_T(t) = (1-\alpha)^t$$

where $\alpha \in (0,1)$ is the death time parameter.

Proposition 11: When T is geometrically distributed, the solution to (7) under Strassen's approximation yields equal epoch sizes. This optimal size is given by

$$\arg\max_{\nu} R_T(\nu) \log M^*(\nu,\eta).$$

Proof: Begin by showing that Algorithm 1 will produce a solution with equal epoch sizes. Recall that the survival function of a geometric random variable with parameter $0 < \alpha \leq 1$ is $R_T(t) = (1-\alpha)^t$. Therefore, the first step of the algorithm will choose n_1 as

$$n_1 = \arg\max_{\nu} (1-\alpha)^{\nu} \log M^*(\nu,\eta).$$

The second step of the algorithm will choose

$$n_2 = \arg\max_{\nu} (1-\alpha)^{n_1}(1-\alpha)^{\nu} \log M^*(\nu,\eta) \\ = \arg\max_{\nu} (1-\alpha)^{\nu} \log M^*(\nu,\eta)$$

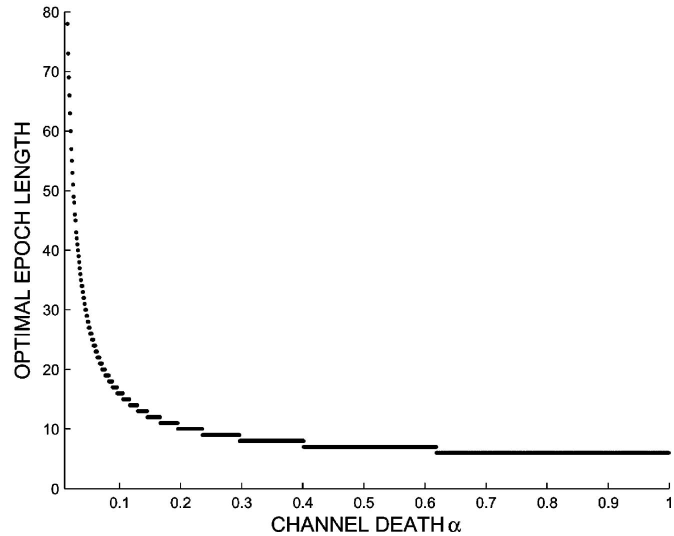

Fig. 2. Optimal epoch lengths under Strassen's approximation for an (ε,α) BSC-geometric channel that dies for $\varepsilon = 0.01$ and $\eta = 0.001$.

which is the same as n_1. In general

$$n_i = \arg\max_{\nu} (1-\alpha)^{e_{i-1}}(1-\alpha)^{\nu} \log M^*(\nu,\eta) \\ = \arg\max_{\nu} (1-\alpha)^{\nu} \log M^*(\nu,\eta)$$

so $n_1 = n_2 = \cdots$.

Such a solution satisfies $n_1 \geq n_2 \geq \cdots$ and so it is optimal by Proposition 10. ■

The optimal epoch size for geometric death under Strassen's approximation can be found analytically, [46, Sec. 6.4.2]. Consider the setting when the alive state corresponds to a $\mathrm{BSC}(\varepsilon)$. For fixed crossover probability ε and target error probability η, the optimal epoch size is plotted as a function of α in Fig. 2. The less likely the channel is to die early, the longer the optimal epoch length.

Alternatively, rather than fixing η, one might fix the number of bits to be communicated and find the best level of reliability that is possible. Fig. 3 shows the best $\lambda_{\max} = \eta$ that is possible when communicating 5 bits over a $\mathrm{BSC}(\varepsilon)$-$\mathrm{geometric}(\alpha)$ channel that dies.

Notice that the geometric death time distribution forms a boundary case for Proposition 10. One can consider discrete Weibull death time distributions [47] to see what happens with heavier tails

$$p_T(t) = (1-\alpha)^{(t-1)^{\beta}} - (1-\alpha)^{t^{\beta}}$$

and

$$R_T(t) = (1-\alpha)^{t^{\beta}}$$

where β is the shape parameter. When $\beta > 1$, the tail is lighter than geometric, and when $\beta < 1$, the tail is heavier than geometric.

With heavy-tailed death distributions, the greedy algorithm gives epoch sizes that are nonincreasing: $n_1 \geq n_2 \geq \cdots$, and therefore optimal; it is better to send long blocks first and then send shorter ones.

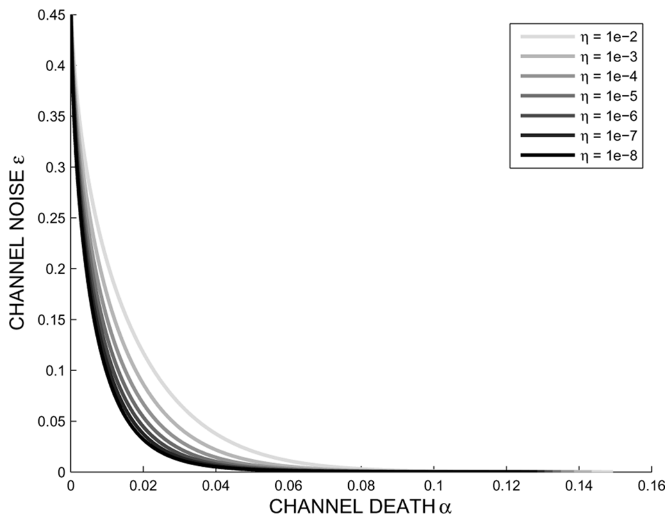

Fig. 3. Achievable η-reliability in sending 5 bits over (ε, α) BSC-geometric channel that dies.

C. Dynamic Programming

The greedy algorithm of the previous section solves (7) under certain conditions. For finite N, a DP may be used to solve (7) under any conditions. To develop the DP formulation [48], we assume that channel state feedback (whether the channel output is ? or whether it is some other symbol) is available to the transmitter; however, solving the DP will show that channel state feedback is not required.

System Dynamics:

$$\begin{bmatrix} \zeta_n \\ \omega_n \end{bmatrix} = \begin{bmatrix} (\zeta_{n-1}+1)\hat{s}_{n-1} \\ \omega_{n-1}\kappa_{n-1} \end{bmatrix} \tag{13}$$

for $n = 1, 2, \ldots, N+1$. The following state variables, disturbances, and controls are used:

1) $\zeta_n \in \mathbb{Z}^*$ is a state variable that counts the location in the current transmission epoch;
2) $\omega_n \in \{0,1\}$ is a state variable that indicates whether the channel is alive (1) or dead (0);
3) $\kappa_n \in \{0,1\} \sim \mathrm{Bern}\,(R_T(n))$ is a disturbance that kills (0) or revives (1) the channel in the next time step; and
4) $\hat{s}_n \in \{0,1\}$ is a control input that starts (0) or continues (1) a transmission epoch in the next time step.

Initial State: Since the channel starts alive (note that $R_T(1) = 1$) and since the first transmission epoch starts at the beginning of time

$$\begin{bmatrix} \zeta_1 \\ \omega_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{14}$$

Additive Cost: Transmission volume $\log M^*(\zeta_n + 1, \eta)$ is credited if the channel is alive (i.e., $\omega_n = 1$) and the transmission epoch is to be restarted in the next time step (i.e., $1 - \hat{s}_n = 1$). This implies a cost function

$$c_n(\zeta_n, \omega_n, \hat{s}_n) = -(1-\hat{s}_n)\omega_n \log M^*(\zeta_n + 1, \eta). \tag{15}$$

This is negative so smaller is better. Note that the 0-1-valued variable $(1 - \hat{s}_n)$ ensures contribution to the cost only when an epoch ends.

Terminal Cost: There is no terminal cost: $c_{N+1} = 0$.

Cost-to-go: From time n to time $N+1$ is

$$\begin{aligned} &\mathrm{E}_{\vec{\kappa}}\left\{\sum_{i=n}^{N} c_i(\zeta_i, \omega_i, \hat{s}_i)\right\} \\ &\quad = -\mathrm{E}_{\vec{\kappa}}\left\{\sum_{i=n}^{N}(1-\hat{s}_i)\omega_i \log M^*(\zeta_i+1, \eta)\right\}. \end{aligned}$$

Notice that the state variable ζ_n which counts epoch time is known to the transmitter and is determinable by the receiver through transmitter simulation. The state variable ω_n indicates the channel state and is known to the receiver by observing the channel output. It may be communicated to the transmitter through the channel state feedback. The following result follows directly.

Proposition 12: A communication scheme that follows the dynamics (13) and additive cost (15) achieves the transmission time-volume

$$\left(N, V = -\mathrm{E}\left[\sum_{n=1}^{N} c_n\right]\right)$$

at η-reliability.

Proof: The proof is trivial: it can directly be verified that the constructed system dynamics, initial state, additive cost, terminal cost, and cost-to-go yield the stated transmission time-volume. ■

DP may be used to find the optimal control policy $(\hat{s}_n)$.

Proposition 13: The optimal $-V$ for the initial state (14), dynamics (13), additive cost (15), and no terminal cost is equal to the cost of the solution produced by the dynamic programming algorithm.

Proof: The system described by initial state (14), dynamics (13), and additive cost (15) is in the form of the *basic problem* of dynamic programming [48, Sec. 1.2]. Thus, the result follows from [48, Prop. 1.3.1]. ■

The DP optimization computations are now carried out; standard J notation is used for cost [48]. The base case at time $N+1$ is

$$J_{N+1}(\zeta_{N+1}, \omega_{N+1}) = c_{N+1} = 0.$$

In proceeding backwards from time N to time 1

$$\begin{aligned} &J_n(\zeta_n, \omega_n) \\ &\quad = \min_{\hat{s}_n \in \{0,1\}} \mathrm{E}_{\kappa_n}\{c_n(\zeta_n, \omega_n, \hat{s}_n) + J_{n+1}\left(f_n(\zeta_n, \omega_n, \hat{s}_n, \kappa_n)\right)\} \end{aligned}$$

for $n = 1, 2, \ldots, N$, where

$$\begin{aligned} f_n(\zeta_n, \omega_n, \hat{s}_n, \kappa_n) &= [\,\zeta_{n+1} \quad \omega_{n+1}\,]^T \\ &= [\,(\zeta_n+1)\hat{s}_n \quad \omega_n\kappa_n\,]^T. \end{aligned}$$

Substituting our additive cost function yields

$$\begin{aligned} &J_n(\zeta_n, \omega_n) \\ &= \min_{\hat{s}_n \in \{0,1\}} -\mathrm{E}_{\kappa_n}\left\{(1-\hat{s}_n)\omega_n \log M^*(\zeta_n+1,\eta)\right\} \\ &\quad + \mathrm{E}_{\kappa_n}\{J_{n+1}\} \\ &= \min_{\hat{s}_n \in \{0,1\}} -(1-\hat{s}_n) R_T(n) \log M^*(\zeta_n+1,\eta) + \mathrm{E}_{\kappa_n}\{J_{n+1}\}. \end{aligned} \tag{16}$$

Notice that the state variable ω_n dropped out of the first term when we took the expectation with respect to the disturbance κ_n. This is true for each stage in the DP.

Proposition 14: For a channel that dies $(\mathcal{X}, p_a(y|x), p_d(y|x), p_T(t), \mathcal{Y})$, channel state feedback does not improve performance.

Proof: By repeating the expectation calculation in (16) for each stage n in the stage-by-stage DP algorithm, it is verified that state variable ω does not enter into the stage optimization problem. Hence, the transmitter does not require channel state feedback to determine the optimal signaling strategy. ■

D. Dynamic Programming Example

To provide some intuition on the choice of epoch lengths, we present a short example. Consider the channel that dies with $\mathcal{X} = \{0,1\}$, $\mathcal{Y} = \{0,1,?\}$, $p_a(y|x)$ given by (1) with $\varepsilon = 0.01$, $p_d(y|x)$ given by (2), and $p_T(t)$ that is uniform over a finite horizon of length 40 (disallowing death in the first time step):

$$p_T(t) = \begin{cases} 1/39, & t = 2, \ldots, 40 \\ 0, & \text{otherwise.} \end{cases}$$

Our goal is to communicate with η-reliability, $\eta = 0.001$.

Since the death distribution has finite support, there is no benefit to transmitting after death is guaranteed. Suppose some sequence of n_is is chosen arbitrarily: $(n_1 = 13,\ n_2 = 13,\ n_3 = 13,\ n_4 = 1)$. This has expected transmission volume (under Strassen's approximation)

$$\begin{aligned} V &= \sum_{i=1}^{4} R_T(e_i) \log M^*(n_i, \eta) \\ &\overset{(a)}{=} \log M^*(13, 0.001) \sum_{i=1}^{3} R_T(e_i) \\ &\overset{(b)}{=} \log M^*(13, 0.001) \sum_{i=1}^{3} \frac{40 - e_i}{39} \\ &= \log M^*(13, 0.001)[R_T(13) + R_T(26) + R_T(39)] \\ &\overset{(c)}{\approx} 4.600[9/13 + 14/39 + 1/39] = 4.954 \text{ bits.} \end{aligned}$$

where (a) removes the fourth epoch since uncoded transmission cannot achieve η-reliability, (b) substitutes the expression for the survival function, and (c) uses the numerical approximation (6) for $\log_M^* \approx 4.600$ when $n_i = 13$, $\eta = 0.001$ and $\varepsilon = 0.01$.

If we run the DP algorithm to optimize the ordered integer partition, we get the result $(n_1 = 20, n_2 = 12, n_3 = 6, n_4 = 2)$.[5] Notice that since the solution is in order, the greedy algorithm would also have succeeded. The expected transmission volume for this strategy (under approximation (6)) is

$$\begin{aligned} V &= R_T(20) \log M^*(20, 0.001) + R_T(32) \log M^*(12, 0.001) \\ &\quad + R_T(38) \log M^*(6, 0.001) \\ &\approx (20/39) \cdot 9.2683 + (8/39) \cdot 3.9694 + (2/39) \cdot 0.5223 \\ &= 5.594 \text{ bits.} \end{aligned}$$

[5]Equivalently $(n_1 = 20, n_2 = 12, n_3 = 6, n_4 = 1, n_5 = 1)$, since the last two channel usages are wasted [see Fig. 1(a)] to hedge against channel death.

E. Precise Solution

It has been assumed that optimal finite blocklength codes are known and used. Moreover, Strassen's approximation has been used for certain computations. It is, however, also of interest to determine precisely which code should be used over a channel that dies. This section gives an example where a sequence of length-23 binary Golay codes [49] are optimal. Similar examples may be developed for other perfect codes.[6]

Before presenting the example, the sphere-packing upper bound on $\log M^*(n_i, \eta)$ for a $\mathrm{BSC}(\varepsilon)$ is derived. Recall the notion of decoding radius [50] and let $\rho(\varepsilon, \eta)$ be the largest integer such that

$$\sum_{s=0}^{\rho} \binom{n_i}{s} \varepsilon^s (1-\varepsilon)^{n_i - s} \le 1 - \eta.$$

The sphere-packing bound follows from counting how many decoding regions of radius ρ could conceivably fit in the Hamming space 2^{n_i} disjointly. Let $D_{s,m}$ be the number of channel output sequences that are decoded into message w_m and have distance s from the mth codeword. By the nature of Hamming space

$$D_{s,m} \le \binom{n_i}{s}$$

and due to the volume constraint

$$\sum_{m=1}^{M} \sum_{s=0}^{\rho} D_{s,m} \le 2^{n_i}.$$

Hence, the maximal codebook size $M^*(n_i, \eta)$ is upper-bounded as

$$\begin{aligned} M^*(n_i, \eta) &\le \frac{2^{n_i}}{\sum_{s=0}^{\rho} D_{s,m}} \\ &\le \frac{2^{n_i}}{\sum_{s=0}^{\rho(\varepsilon,\eta)} \binom{n_i}{s}}. \end{aligned}$$

Thus, the sphere-packing upper bound on $\log M^*(n_i, \eta)$ is

$$\log M^*(n_i, \eta) \le n_i - \log \left[\sum_{s=0}^{\rho(\varepsilon,\eta)} \binom{n_i}{s} \right] \triangleq \log M_{sp}(n_i, \eta).$$

Perfect codes such as the binary Golay code of length 23 achieve the sphere-packing bound with equality if the decoding radius $\rho(\varepsilon, \eta)$ matches the distance between codewords in the code.

[6]A perfect code is one for which there are equal-radius spheres centered at the codewords that are disjoint and that completely fill $\mathcal{X}^{n_i}$. Note that perfect codes other than repetition codes do not exist for most combinations of dimension and alphabet size.

Consider an (ε, α) BSC-geometric channel that dies, with $\varepsilon = 0.01$ and $\alpha = 0.05$. The target error probability is fixed at $\eta = 2.9 \times 10^{-6}$. For these values of ε and η, the decoding radius $\rho(\varepsilon, \eta) = 1$ for $2 \le n_i \le 3$. It is $\rho(\varepsilon, \eta) = 2$ for $4 \le n_i \le 10$; $\rho(\varepsilon, \eta) = 3$ for $11 \le n_i \le 23$; $\rho(\varepsilon, \eta) = 4$ for $24 \le n_i \le 40$; and so on.

Moreover, one can note that the $(n = 23, M = 4096)$ binary Golay code has a decoding radius of 3; thus it meets the BSC sphere-packing bound

$$M_{sp}(23, 2.9 \times 10^{-6}) = \frac{2^{23}}{1 + 23 + 253 + 1771} = 4096$$

with equality.

Now to bring channel death into the picture: If one proceeds greedily, following Algorithm 1, but using the sphere-packing bound $\log M_{sp}(n_i, \eta)$ rather than the optimal $\log M^*(n_i, \eta)$

$$\begin{aligned} n_1(\varepsilon = 0.01, \alpha = 0.05, \eta = 2.9 \times 10^{-6}) \\ = \arg\max_{\nu} \bar{\alpha}^{\nu} \log_2 \frac{2^{\nu}}{\sum_{s=0}^{\rho(\varepsilon,\eta)}} = 23. \end{aligned}$$

By the memorylessness argument of Proposition 11, it follows that running Algorithm 1 with the sphere-packing bound will yield $23 = n_1 = n_2 = \cdots$.

It remains to show that Algorithm 1 actually gives the true solution. Had Strassen's approximation been used rather than the sphere-packing bound, the result would follow directly from Proposition 11. Instead, the global optimality condition (9) can be verified exhaustively for all 23 possible shift sizes K for the first epoch

$$\frac{\bar{\alpha}^{23} \log M_{sp}(23, \eta) - \bar{\alpha}^{23-K} \log M_{sp}(23 - K, \eta)}{\bar{\alpha}^{46} \log M_{sp}(23 + K) - \bar{\alpha}^{46} \log M_{sp}(23, \eta)} \ge 1.$$

Then, the same exhaustive verification is performed for all 23 possible shifts for the second epoch

$$\begin{aligned} \frac{\bar{\alpha}^{46} \log M_{sp}(23, \eta) - \bar{\alpha}^{46-K} \log M_{sp}(23 - K, \eta)}{\bar{\alpha}^{69} \log M_{sp}(23 + K) - \bar{\alpha}^{69} \log M_{sp}(23, \eta)} &\ge 1 \\ \frac{\bar{\alpha}^{23}\left[\bar{\alpha}^{23} \log M_{sp}(23, \eta) - \bar{\alpha}^{23-K} \log M_{sp}(23 - K, \eta)\right]}{\bar{\alpha}^{23}\left[\bar{\alpha}^{46} \log M_{sp}(23 + K) - \bar{\alpha}^{46} \log M_{sp}(23, \eta)\right]} &\ge 1 \\ \frac{\bar{\alpha}^{23} \log M_{sp}(23, \eta) - \bar{\alpha}^{23-K} \log M_{sp}(23 - K, \eta)}{\bar{\alpha}^{46} \log M_{sp}(23 + K) - \bar{\alpha}^{46} \log M_{sp}(23, \eta)} &\ge 1. \end{aligned}$$

For each future epoch beyond the second, a nearly identical exhaustive verification procedure can be carried out to show that using the length-23 binary Golay code is optimal for that epoch.

F. Practical Codes and Empirical Death Distributions

It should be noted that the algorithms developed for optimizing communication schemes over channels that die work with arbitrary death distributions, even empirically measured ones, e.g., the experimentally characterized death properties of a synthetic biology communication system [4, Fig. 3: Reliability].

Further, rather than considering the $\log M^*(n_i, \eta)$ function for optimal finite blocklength codes, the code optimization procedures would work just as well if a collection of finite blocklength codes was provided. Such a limited set of codes might be selected for decoding complexity or other practical reasons. As an example, consider the collection $\mathcal{C}$ of 9191 binary minimum distance codes of lengths between 6 and 16 given in [39, DVD supplement].[7] We run the optimization over the example in Section V-D but restricting to $\mathcal{C}$.

The solution obtained from optimization has epoch sizes $(n_1 = 15,\ n_2 = 15,\ n_3 = 9,\ n_4 = 1)$. When measuring expected transmission volume using Strassen's approximation for this set of epoch sizes, the result is 5.344 bits. This is less than the 5.594 bits under the optimal epoch sizes under Strassen's approximation. Strassen's approximation is, however, only an approximation and the exact expected transmission volume achieved with the epoch sizes optimized for $\mathcal{C}$ is 7.246 bits. The two minimum distance codes used are the $(n = 15,\ M = 256,\ d = 5)$ code and the $(n = 9,\ M = 6,\ d = 3)$ code. It remains to be seen whether the restriction to the collection of minimum distance codes is actually suboptimal.

VI. Partial Ordering of Channels

It is of interest to order channels that die by quality. The partial ordering of DMCs was studied by Shannon [51], and as a first step, we can slightly extend his result to order channels that die having common death distributions.

Definition 15: Let $p(i, j)$ be the transition probabilities for a DMC C_1 and let $q(k, l)$ be the transition probabilities for a DMC C_2. Then, C_1 is said to include C_2, $C_1 \supseteq C_2$, if there exist two sets of valid transition probabilities $r_\gamma(k, i)$ and $t_\gamma(j, l)$, and there exists a vector g: $g_\gamma \ge 0$ and $\sum_\gamma g_\gamma = 1$, such that

$$\sum_{\gamma, i, j} g_\gamma r_\gamma(k, i) p(i, j) t_\gamma(j, l) = q(k, l).$$

Proposition 15: Consider two channels that die with identical death distributions: $(\mathcal{X}_1, p_a, p_d, p_T(t), \mathcal{Y}_1)$ and $(\mathcal{X}_2, q_a, q_d, p_T(t), \mathcal{Y}_2)$. Let DMC C_1 correspond to p_a and let DMC C_2 correspond to q_a and moreover suppose that $C_1 \supseteq C_2$. Fix a transmission time N and an expected transmission volume V. Let η_1 be the best level of reliability for the first channel and η_2 be the best level of reliability for the second channel, under (N, V). Then, $\eta_1 \le \eta_2$.

Proof: The main theorem of [51] proves that the average error probability when transmitting an individual message code over C_1 is less than or equal to the average error probability when transmitting the same individual message code over C_2.

Shannon's proof [51] holds *mutatis mutandis* for maximum error probability, replacing "average error probability" by "maximum error probability."

The desired result follows by concatenating individual message codes into a code. ■

We can also order channels that die having common alive state transition probabilities.

Definition 16: Consider two random variables T and U with survival functions $R_T(\cdot)$ and $R_U(\cdot)$, respectively. Then, U is

[7] By a minimum distance code, we mean a code that has maximum cardinality among all codes that have a given length and a given minimum distance among codewords.

said to stochastically dominate T, $U \geq_{\rm st} T$, if $R_T(t) \leq R_U(t)$ for all t.

Proposition 16: Consider two channels that die with identical state properties: $(\mathcal{X}, p_a(y|x), p_d(y|x), p_T, \mathcal{Y})$ and $(\mathcal{X}, p_a(y|x), p_d(y|x), q_U, \mathcal{Y})$. Let death random variable T correspond to p_T and let death random variable U correspond to q_U and moreover suppose that $U \geq_{\rm st} T$. Fix a transmission time N and a level of reliability η. Let V_1 be the best expected transmission volume for the first channel and V_2 be the best expected transmission volume for the second channel, under (N, η). Then, $V_2 \geq V_1$.

Proof: Recall the expected transmission volume expression (7) for the first channel

$$\max_{(n_i):\sum n_i = N} \sum_i R_T(e_i) \log M^*(n_i, \eta)$$

and for the second channel

$$\max_{(\nu_i):\sum \nu_i = N} \sum_i R_U(\iota_i) \log M^*(\nu_i, \eta).$$

Since $R_T(t) \leq R_U(t)$ for all t, the result follows directly. ■

These two results give individual ordering principles in the two dimensions essentially depicted in Fig. 3. Putting them together provides a partial order on all channels that die: if one channel is better than another channel in both dimensions, then it is better overall.

Proposition 17: Consider two channels that die: $(\mathcal{X}_1, p_a, p_d, p_T, \mathcal{Y}_1)$ and $(\mathcal{X}_2, q_a, q_d, q_U, \mathcal{Y}_2)$. Let DMC C_1 correspond to p_a and let DMC C_2 correspond to q_a and moreover suppose that $C_2 \supseteq C_1$. Let death random variable T correspond to p_T and let death random variable U correspond to q_U and moreover suppose that $U \geq_{\rm st} T$. Fix a transmission time N and a level of reliability η. Let V_1 be the best expected transmission volume for the first channel and V_2 be the best expected transmission volume for the second channel, under (N, η). Then, $V_2 \geq V_1$.

VII. Conclusion and Future Work

We have formulated the problem of communication over channels that die and have shown how to maximize expected transmission volume at a given level of error probability reliability.

There are several extensions to the basic formulation studied in this paper that one might consider; we list a few.

1) Inspired by synthetic biology [4], rather than thinking of death time as independent of the signaling scheme X_1^n, one might consider channels that die because they lose fitness as a consequence of operation: T would be dependent on X_1^n. This would be similar to Gallager's panic button/child's toy channel, and would have intersymbol interference [27], [30]. There would also be strong connections to channels that heat up [52] and communication with a dynamic cost [53, Ch. 3].
2) In the emerging attention economy [54], agents faced with information overload [55] may permanently stop listening to certain communication media received over noisy channels. This setting is exactly modeled by channels that die. The impact of communication over channels that die on the productivity and efficiency of human organizations may be determined by building on the results herein.
3) Since channel death is indicated by the symbol ?, the receiver unequivocally knows death time. Other channel models might not have a distinct output letter for death and would need to detect death, perhaps using the theory of estimating stopping times [56].
4) Inspired by communication terminals that randomly lie within communication range, e.g., in vehicular communication, one might also consider a channel that is born at a random time and then dies at a random time. One would suspect that channel state feedback would be beneficial. Networks of birth–death channels are also of interest and would have connections to percolation-style work [2].
5) This study has simply considered the channel coding problem; however, there are several formulations of end-to-end information transmission problems over channels that die, which are of interest in many application areas. There is no reason to suspect a separation principle.

Randomly stepping back from infinity leads to some new understanding of the fundamental limits of communication in the presence of noise and unreliability.

Acknowledgment

We thank Barry Canton (Ginkgo BioWorks) and Drew Endy (Stanford University) for discussions on synthetic biology that initially inspired this work. Discussions with Baris Nakiboğlu and Yury Polyanskiy are also appreciated.

References

[1] L. R. Varshney, S. K. Mitter, and V. K. Goyal, "Channels that die," in *Proc. 47th Annu. Allerton Conf. Commun. Control Comput.*, Sep.–Oct. 2009, pp. 566–573.
[2] I. M. Jacobs, "Connectivity in probabilistic graphs: An abstract study of reliable communications in systems containing unreliable components," Sc.D. thesis, Massachusetts Inst. Technol., Cambridge, MA, Aug. 1959.
[3] I. Dietrich and F. Dressler, "On the lifetime of wireless sensor networks," *ACM Trans. Sensor Netw.*, vol. 5, no. 1, article no. 5, Feb. 2009.
[4] B. Canton, A. Labno, and D. Endy, "Refinement and standardization of synthetic biological parts and devices," *Nat. Biotechnol.*, vol. 26, no. 7, pp. 787–793, Jul. 2008.
[5] J. D. Bekenstein, "The limits of information," *Stud. Hist. Philos. Mod. Phys.*, vol. 32, no. 4, pp. 511–524, Dec. 2001.
[6] D. R. Headrick, *The Invisible Weapon: Telecommunications and International Politics, 1851–1945*. New York: Oxford Univ. Press, 1991.
[7] S. Hara, A. Ogino, M. Araki, M. Okada, and N. Morinaga, "Throughput performance of SAW-ARQ protocol with adaptive packet length in mobile packet data transmission," *IEEE Trans. Veh. Technol.*, vol. 45, no. 3, pp. 561–569, Aug. 1996.
[8] E. Modiano, "An adaptive algorithm for optimizing the packet size used in wireless ARQ protocols," *Wireless Netw.*, vol. 5, no. 4, pp. 279–286, Jul. 1999.
[9] P. Lettieri and M. B. Srivastava, "Adaptive frame length control for improving wireless link throughput, range, and energy efficiency," in *Proc. 17th Annu. Joint Conf. IEEE Comput. Commun. Soc.*, Mar. 1998, vol. 2, pp. 564–571.
[10] S. Ci, H. Sharif, and K. Nuli, "Study of an adaptive frame size predictor to enhance energy conservation in wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 283–292, Feb. 2005.
[11] D. Slepian, "Bounds on communication," *Bell Syst. Tech. J.*, vol. 42, no. 3, pp. 681–707, May 1963.
[12] S. S. L. Chang, B. Harris, and J. J. Metzner, "Optimum message transmission in a finite time," *IRE Trans. Inf. Theory*, vol. IT-8, no. 5, pp. 215–224, Sep. 1962.

[13] S. J. MacMullan and O. M. Collins, "A comparison of known codes, random codes, and the best codes," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 3009–3022, Nov. 1998.

[14] J. N. Laneman, "On the distribution of mutual information," presented at the Inf. Theory Appl. Inaugural Workshop, Feb. 2006.

[15] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[16] D. Buckingham and M. C. Valenti, "The information-outage probability of finite-length codes over AWGN channels," in *Proc. 42nd Annu. Conf. Inf. Sci. Syst.*, Mar. 2008, pp. 390–395.

[17] G. Wiechman and I. Sason, "An improved sphere-packing bound for finite-length codes over symmetric memoryless channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1962–1990, May 2008.

[18] Y. Altŭg and A. B. Wagner, "Refinement of the sphere packing bound for symmetric channels," in *Proc. 49th Annu. Allerton Conf. Commun. Control Comput.*, Sep. 2011, pp. 30–37.

[19] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 359–378, May 1994.

[20] A. Goldsmith, *Wireless Communications*. New York: Cambridge Univ. Press, 2005.

[21] J. K. Wolf, A. D. Wyner, and J. Ziv, "The channel capacity of the postal channel," *Inf. Control*, vol. 16, no. 2, pp. 167–172, Apr. 1970.

[22] M. Zeng, R. Zhang, and S. Cui, "On the outage capacity of a dying channel," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2008, pp. 1–5.

[23] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Dispersion of the Gilbert-Elliott channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1829–1848, Apr. 2011.

[24] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. IT-2, no. 3, pp. 8–19, Sep. 1956.

[25] S. Sanghavi, "Intermediate performance of rateless codes," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2007, pp. 478–482.

[26] A. Kamra, V. Misra, J. Feldman, and D. Rubenstein, "Growth codes: Maximizing sensor network data persistence," in *Proc. Conf. Appl. Technol. Archit. Protocols Comput. Commun.*, Sep. 2006, pp. 255–266.

[27] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[28] S. M. Ross, *Stochastic Processes*. New York: Wiley, 1996.

[29] R. G. Gallager, *Discrete Stochastic Processes*. Boston, MA: Kluwer, 1996.

[30] R. Gallager, *Information Theory and Reliable Communication*, ser. International Centre for Mechanical Sciences, Courses and Lectures. Vienna, Austria: Springer-Verlag, 1972.

[31] S. Tatikonda and S. Mitter, "The capacity of channels with feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 323–349, Jan. 2009.

[32] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert-Elliott channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 1211–1290, Nov. 1989.

[33] A. J. Goldsmith and P. P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 868–886, May 1996.

[34] L. E. Braten and T. Tjelta, "Semi-Markov multistate modeling of the land mobile propagation channel for geostationary satellites," *IEEE Trans. Antennas Propag.*, vol. 50, no. 12, pp. 1795–1802, Dec. 2002.

[35] J. Wang, J. Cai, and A. S. Alfa, "New channel model for wireless communications: Finite-state phase-type semi-Markov channel model," in *Proc. IEEE Int. Conf. Commun.*, May 2008, pp. 4461–4465.

[36] S. Wang and J.-T. Park, "Modeling and analysis of multi-type failures in wireless body area networks with semi-Markov model," *IEEE Commun. Lett.*, vol. 14, no. 1, pp. 6–8, Jan. 2010.

[37] F. Jelinek, *Probabilistic Information Theory: Discrete and Memoryless Models*. New York: McGraw-Hill, 1968.

[38] R. J. McEliece, *The Theory of Information and Coding*. Cambridge, U.K.: Cambridge Univ. Press, 2002.

[39] P. Kaski and P. R. J. Östergård, *Classification Algorithms for Codes and Designs*. Berlin, Germany: Springer, 2006.

[40] A. Barg and A. McGregor, "Distance distribution of binary codes and the error probability of decoding," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4237–4246, Dec. 2005.

[41] V. Strassen, "Asymptotische abschätzungen in Shannons informationstheorie," in *Proc. Trans. 3rd Prague Conf. Inf. Theory, Statist. Decis. Funct., Random Process.*, Prague, Czech Republic, 1962, pp. 689–723.

[42] L. Weiss, "On the strong converse of the coding theorem for symmetric channels without memory," *Q. Appl. Math.*, vol. 18, no. 3, pp. 209–214, Oct. 1960.

[43] G. D. Forney Jr., "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 2, pp. 206–220, Mar. 1968.

[44] Y. Polyanskiy, "Channel coding: Nonasymptotic fundamental limits," Ph.D. dissertation, Princeton Univ., Princeton, NJ, Nov. 2010.

[45] D. J. Davis, "An analysis of some failure data," *J. Am. Statist. Assoc.*, vol. 47, no. 258, pp. 113–150, Jun. 1952.

[46] L. R. Varshney, "Unreliable and Resource-Constrained Decoding," Ph.D. thesis, Massachusetts Inst. Technol., Cambridge, MA, Jun. 2010.

[47] M. S. A. Khan, A. Khalique, and A. M. Abouammoh, "On estimating parameters in a discrete Weibull distribution," *IEEE Trans. Rel.*, vol. 38, no. 3, pp. 348–350, Aug. 1989.

[48] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Belmont, MA: Athena Scientific, 2005, vol. 1.

[49] M. J. E. Golay, "Notes on digital coding," *Proc. IRE*, vol. 37, no. 6, p. 657, Jun. 1949.

[50] R. E. Blahut, *Theory and Practice of Error Control Codes*. Reading, MA: Addison-Wesley, 1983.

[51] C. E. Shannon, "A note on a partial ordering for communication channels," *Inf. Control*, vol. 1, no. 4, pp. 390–397, Dec. 1958.

[52] T. Koch, A. Lapidoth, and P. P. Sotiriadis, "Channels that heat up," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3594–3612, Aug. 2009.

[53] K. Eswaran, "Communication and third parties: Costs, cues, and confidentiality," Ph.D. dissertation, Univ. California, Berkeley, Berkeley, 2009.

[54] T. H. Davenport and J. C. Beck, *The Attention Economy: Understanding the New Currency of Business*. Boston: Harvard Business School Press, 2001.

[55] T. Van Zandt, "Information overload in a network of targeted communication," *Rand J. Econ.*, vol. 35, no. 3, pp. 542–560, Autumn 2004.

[56] U. Niesen and A. Tchamkerten, "Tracking stopping times through noisy observations," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 422–432, Jan. 2009.

Lav R. Varshney (S'00–M'10) received the B.S. degree with honors in electrical and computer engineering (magna cum laude) from Cornell University, Ithaca, NY, in 2004. He received the S.M., E.E., and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2006, 2008, and 2010, respectively.

He is a research staff member at the IBM Thomas J. Watson Research Center, Hawthorne, NY. He has been a postdoctoral associate, National Science Foundation graduate research fellow, and held various research and teaching positions at MIT. He has also held visiting positions at École Polytechnique Fédérale de Lausanne and at Cold Spring Harbor Laboratory. His current research interests include information theory, data analytics, sociotechnical systems, and neuroscience.

Dr. Varshney is a member of Tau Beta Pi, Eta Kappa Nu, and Sigma Xi. He received the J.-A. Kong Award Honorable Mention for Electrical Engineering doctoral thesis and the E. A. Guillemin Thesis Award for Outstanding Electrical Engineering S. M. Thesis.

Sanjoy K. Mitter (M'68–SM'77–F'79) received his Ph.D. degree from the Imperial College of Science and Technology in 1965. He taught at Case Western Reserve University from 1965 to 1969. He joined MIT in 1969 where he has been a Professor of Electrical Engineering since 1973. He was the Director of the MIT Laboratory for Information and Decision Systems from 1981 to 1999. He has also been a Professor of Mathematics at the Scuola Normale, Pisa, Italy from 1986 to 1996. He has held visiting positions at Imperial College, London; University of Groningen, Holland; INRIA, France; Tata Institute of Fundamental Research, India and ETH, Zürich, Switzerland; and several American universities.

Professor Mitter will be the Ulam Scholar at Los Alamos National Laboratories in April 2012 and the John von Neumann Visiting Professor in Mathematics at the Technical University of Munich, Germany from May-June 2012. He was awarded the AACC Richard E. Bellman Control Heritage Award for 2007. He was the McKay Professor at the University of California, Berkeley in March 2000, and held the Russell-Severance-Springer Chair in Fall 2003. He is the winner of the 2000 IEEE Control Systems Award. He was elected a Foreign Member of Istituto Veneto di Scienze, Lettere ed Arti in 2003. In 1988, he was elected to the National Academy of Engineering.

His current research interests are communication and control in a networked environment, the relationship of statistical and quantum physics to information theory and control, and autonomy and adaptiveness for integrative organization.

Vivek K Goyal (S'92–M'98–SM'03) received the B.S. degree in mathematics and the B.S.E. degree in electrical engineering from the University of Iowa, Iowa City, where he received the John Briggs Memorial Award for the top undergraduate across all colleges. He received the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, where he received the Eliahu Jury Award for outstanding achievement in systems, communications, control, or signal processing.

He was a Member of Technical Staff in the Mathematics of Communications Research Department of Bell Laboratories, Lucent Technologies, 1998–2001; and a Senior Research Engineer for Digital Fountain, Inc., 2001–2003. He is currently Esther and Harold E. Edgerton Associate Professor of electrical engineering at the Massachusetts Institute of Technology. His research interests include computational imaging, sampling, quantization, and source coding theory.

Dr. Goyal is a member of Phi Beta Kappa, Tau Beta Pi, Sigma Xi, Eta Kappa Nu and SIAM. He was awarded the 2002 IEEE Signal Processing Society Magazine Award and an NSF CAREER Award. He served on the IEEE Signal Processing Society's Image and Multiple Dimensional Signal Processing Technical Committee. He is a Technical Program Committee Co-chair of IEEE ICIP 2016 and a permanent Conference Co-chair of the SPIE Wavelets and Sparsity conference series.