Error Exponents for Variable-length block codes with feedback and cost constraints

Barış Nakiboğlu Robert G. Gallager Moe Z. Win
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology, Cambridge, MA, 02139
Email: {nakib, gallager, moewin }@mit.edu

***Abstract*— Variable-length block-coding schemes are investigated for discrete memoryless channels (DMC) with perfect feedback under cost constraints. Upper and lower bounds are found for the minimum achievable probability of decoding error $P_{e,\min}$ as a function of transmission rate R, cost constraint $\mathcal{P}$, and expected block length $\overline{\tau}$. For given $\mathcal{P}$ and R, the lower and upper bounds to the exponent $-(\ln P_{e,\min})/\overline{\tau}$ are asymptotically equal as $\overline{\tau} \to \infty$. The reliability function, $\lim_{\overline{\tau}\to\infty}(-\ln P_{e,\min})/\overline{\tau}$, as a function of $\mathcal{P}$ and R, is concave in the pair $(\mathcal{P}, R)$ and generalizes the linear reliability function of Burnashev [1] to include cost constraints.**

I. Introduction

A variable-length block code is a code in which each message must be encoded, transmitted, and permanently decoded before the next message enters the encoder. In an ordinary block-coding scheme with feedback, the codewords all have a predetermined length, but the codeword symbols can depend on previous channel outputs as well as the message. For variable-length block coding, the disjointness of the time intervals for subsequent messages is preserved, but the receiver dynamically determines the decoding time based on the received symbols up to that point. We start with a brief overview of previous work on fixed and variable-length block codes with feedback.

A. Outline of previous work

A widely accepted quality-of-service criterion for fixed-length block codes is the error exponent, $-(\ln P_e)/\tau$, where τ is the block length. Dobrushin [3] showed that the sphere-packing exponent is an asymptotic upper bound for the error exponent for fixed-length block codes on symmetric DMC's with feedback. Haroutunian [5] derived an upper bound for arbitrary DMC's, but it has been long conjectured that the sphere packing bound also applies in this case.

For discrete-time additive-white-Gaussian-noise (AWGN) channels, Pinsker [7] showed that the sphere-packing exponent is an asymptotic upper bound for the error exponent with feedback under the added constraint that the total energy, for each message and noise realization, is at most the average power constraint times the block length. Using only an average power constraint, without the above added constraint, Schalkwijk and Kailath [9], [8], showed that P_e can be made to decay as a two-fold exponential in block length. Kramer [6] later showed that an n-fold exponential decay can be achieved for any $n > 0$; no lower bound to $P_{e,\min}$ is known in this case.

Variable-length block coding on a DMC allows the decoding to be delayed under unusually severe noise (just as additional energy can be used in the AWGN case). Burnashev [1] developed upper and lower bounds to $P_{e,\min}$ for variable-length block coding schemes for the DMC. For given R, his lower and upper bounds to $(-\ln P_{e,\min})/\overline{\tau}$ are asymptotically equal as $\overline{\tau} \to \infty$. The resulting reliability function, $\lim_{\overline{\tau}\to\infty}(-\ln P_{e,\min})/\overline{\tau}$, as a function of R decreases linearly from some constant $\mathbf{D}^*$ at $R = 0$ to 0 at the channel capacity $\mathbf{C}^*$. Our main contribution here is to find the reliability function of variable-length block-coding schemes on DMC's with average cost constraints. We show that for each average cost $\mathcal{P}$, the reliability function is a concave function of R.

B. Forward channel, feedback channel, and cost constraint

The forward channel is assumed to be a DMC with input alphabet $\{1, \ldots, |\mathcal{X}|\}$ and output alphabet $\{1, \ldots, |\mathcal{Y}|\}$. The input and output at time n are denoted by X_n and Y_n; the n-tuples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ are denoted by X^n and Y^n. The feedback channel is discrete and noiseless with an arbitrarily large alphabet size $|\mathcal{Z}|$ (although $|\mathcal{Z}| = |\mathcal{Y}|$ is sufficient). The symbol Z_n sent from the receiver at time n can depend on Y^n and is received without error at the transmitter after X_n and before X_{n+1} is sent. Z^n denotes $Z_1, \ldots, Z_n$.

The forward DMC is defined by the $|\mathcal{X}|$ by $|\mathcal{Y}|$ transition matrix $\{P_{kj}\}$ where, for each time n, $\mathbf{P}[Y_n = j | X_n = k] = P_{kj}$. The channel is memoryless in the sense that

$$\mathbf{P}\left[Y_n | X^n, Y^{n-1}, Z^{n-1}\right] = \mathbf{P}\left[Y_n | X_n\right].$$

We assume throughout that each pair of rows are different. This causes no loss of generality since inputs i and k could be considered the same if $P_{ij} = P_{kj}$ for all $j \in \mathcal{Y}$. With the exception of the concluding section, we also assume that $P_{kj} > 0$ for all k, j,

For each input letter $k \in \mathcal{X}$, there is a transmission cost ρ_k. The cost S_τ of transmitting a codeword of length τ is the sum of the costs of the τ symbols in the codeword. The cost constraint $\mathcal{P}$ means that $\mathbf{E}[\mathcal{S}_\tau] \leq \mathcal{P}\mathbf{E}[\tau]$. We usually refer to $\mathcal{P}$ as a *power constraint* and to S_τ as *energy*. With this definition of power constraint, $\mathcal{P}$ can be seen to upper bound the long-term time-average cost per symbol over a long string of successive message transmissions.

We assume that $\min_k \rho_k = 0$. This causes no loss of generality, since otherwise $\min_k \rho_k$ could be subtracted from

1-4244-0504-1/06/$20.00 ©2006 IEEE

$\mathcal{P}$ and from each ρ_k.

II. Achievability: Asymptotically optimum codes

In this section, we describe a class of coding schemes that are adaptations of the Yamamoto and Itoh [10] codes, adapted to account for the power constraint. We derive the relationship between rate, power constraint, and error exponent for these codes.

We begin with a slightly simpler problem, finding fixed-length block codes for an error-or-erasure decoder, i.e., a decoder which can either decode the message or produce an erasure symbol. The objective will be to minimize (or approximately minimize) the error probability while keeping the erasure probability relatively small.

A. Fixed-length block codes with error-or-erasure decoding

Consider a code of fixed-length ℓ containing two phases of length ℓ_1 and ℓ_2 respectively. The first phase uses a power constraint $\mathcal{P}_1$ and the second $\mathcal{P}_2$. To meet an overall power constraint $\mathcal{P}$, we require $\ell\mathcal{P} = \ell_1\mathcal{P}_1 + \ell_2\mathcal{P}_2$. Define η as ℓ_1/ℓ, so that this power constraint becomes

$$\mathcal{P} = \eta\mathcal{P}_1 + (1-\eta)\mathcal{P}_2.$$

Phase 1 consists of a conventional block code without feedback, operating incrementally close to the capacity $\mathbf{C}(\mathcal{P}_1)$ of the channel subject to constraint $\mathcal{P}_1$,

$$\mathbf{C}(\mathcal{P}_1) = \max_{\phi:\, \sum_k \phi_k\rho_k \le \mathcal{P}_1} \sum_{k,j} \phi_k P_{kj} \ln \frac{P_{kj}}{\sum_i \phi_i P_{ij}}. \quad (1)$$

Here and throughout, ϕ is assumed to be a probability assignment, i.e., $\phi_k \ge 0$ for each k and $\sum_k \phi_k = 1$. The conventional non-feedback coding theorem[1] is as follows: for any $\delta_1 > 0$, there is an $\epsilon_1(\delta_1) > 0$ such that, for all large enough ℓ_1, codes exist with $M \ge e^{\ell_1[\mathbf{C}(\mathcal{P}_1)-\delta_1]}$ code words, each of energy at most $\ell_1\mathcal{P}_1$ and each with error probability upper bounded by

$$P_{e1} \le \exp -\ell_1\epsilon_1(\delta_1).$$

Using such a code in phase 1, the decoder makes a tentative decision at the end of phase 1. The transmitter (knowing the decision via feedback) then sends a binary codeword, $\mathbf{x}_A$ for 'accept' and $\mathbf{x}_R$ for 'reject' in phase 2. Let P_{RA} be the probability that the receiver decodes $\mathbf{x}_A$ given that $\mathbf{x}_R$ is sent. Similarly, P_{AR} is the probability of decoding $\mathbf{x}_R$ given $\mathbf{x}_A$.

If $\mathbf{x}_A$ is decoded, the receiver gives its tentative decision to the user and the overall probability of error P_e satisfies $P_e \le P_{RA}$. If $\mathbf{x}_R$ is decoded, an erasure is released and the probability of erasure P_r satisfies $P_r \le P_{AR} + P_{e1}$. Assume for now that the power constraint may be violated by an incrementally small amount. Thus we choose $\mathbf{x}_A$ to satisfy the constraint, and choose $\mathbf{x}_R$ arbitrarily since it is rarely used. We bound $-\ln P_{RA}$ by the divergence of the output distribution conditional on $\mathbf{x}_A$ relative to that conditional on $\mathbf{x}_R$.

To be more explicit, define the single letter divergence for the input letter k as

$$D_k = \max_i \sum_j P_{kj} \ln \frac{P_{kj}}{P_{ij}}.$$

For each k, let i_k be an input letter achieving the above maximum. If $\mathbf{x}_a$ contains $\phi_k\ell_2$ occurrences of letter k and $\mathbf{x}_r$ is chosen to contain the letter i_k whenever $\mathbf{x}_a$ contains k, then the following result holds[2]: for any $\delta_2 > 0$, there is an $\epsilon_2(\delta_2) > 0$ such that

$$\begin{aligned} P_{RA} &\le \exp\left[\sum_k -\ell_2\phi_k D_k + \ell_2\delta_2\right], && (2)\\ P_{AR} &\le \exp\left[-\ell_2\epsilon_2(\delta_2)\right]. && (3)\end{aligned}$$

From (2), we want to choose $\mathbf{x}_A$ to maximize $\sum_k D_k\phi_k$ subject to the power constraint. Thus, for a power constraint $\mathcal{P}_2$ in phase 2, define $\mathbf{D}(\mathcal{P}_2)$ as

$$\mathbf{D}(\mathcal{P}_2) = \max_{\phi:\, \sum_k \phi_k\rho_k \le \mathcal{P}_2} \sum_k D_k\phi_k. \quad (4)$$

The function $\mathbf{D}(\mathcal{P})$ in (4) is the maximum of a linear function over linear constraints. As illustrated in Figure 1, $\mathbf{D}(\mathcal{P})$ is piecewise linear, non-decreasing, and concave in its region of definition, $\mathcal{P} \ge 0$.

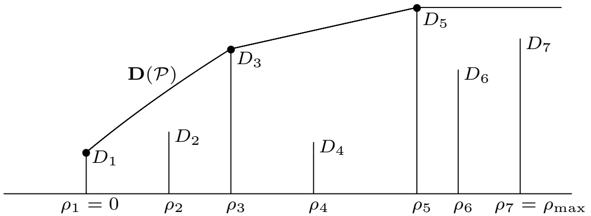

Fig. 1. Calculation of $\mathbf{D}(\mathcal{P})$ as a function of $\mathcal{P}$. The single letter divergences D_k are also shown. For convenience, the inputs are ordered in terms of cost. Note that ϕ_k need be positive for at most 2 values of k.

Choosing the codewords $\mathbf{x}_A$ and $\mathbf{x}_R$ according to this maximization, (2) becomes

$$P_{RA} \le \exp\left[-\ell_2\mathbf{D}(\mathcal{P}_2) + \ell_2\delta_2\right]. \quad (5)$$

The power constraint $\mathcal{P}_2$ is then satisfied by $\mathbf{x}_A$ and The power in $\mathbf{x}_R$ can be upper bounded by $\rho_{\max}$. The preceding results are summarized in the following lemma.

Lemma 1: For all $\mathcal{P}_1 \ge 0$, $\mathcal{P}_2 \ge 0$, and for all $0 < \eta < 1$, all positive δ_1 and δ_2, and all sufficiently large ℓ, there is an error and erasure code with $M \ge \exp\{\eta\ell[\mathbf{C}(\mathcal{P}_1) - \delta_1]\}$ such that for each codeword, the probability of error P_e, the probability of erasure P_r, and the expected energy $\mathbf{E}[\mathcal{S}]$ satisfy

$$\begin{aligned} P_e &\le \exp\{-(1-\eta)\ell\,[\mathbf{D}(\mathcal{P}_2) - \delta_2]\}, && (6)\\ P_r &\le e^{-\eta\ell\,\epsilon_1(\delta_1)} + e^{-(1-\eta)\ell\,\epsilon_2(\delta_2)}, && (7)\\ \mathbf{E}[\mathcal{S}] &\le \ell[\eta\mathcal{P}_1 + (1-\eta)\mathcal{P}_2 + \rho_{\max}e^{-\eta\ell\,\epsilon_1(\delta_1)}]. && (8)\end{aligned}$$

[1] See, for example, Theorem 7.3.2 in [4]

[2] This can be derived, for example, by starting with Theorem 5 in [2] and specializing to the case of asymptotically small s.

B. Variable-length block codes

The above error-or-erasure code is now converted into a variable-length block code. As in Yamamoto and Itoh [10], the transmitter observes each erasure via the feedback and repeats the fixed length codeword until it is accepted. Since an error occurs independently after each repetition of the fixed length codeword,

$$P_e \leq \frac{1}{1-P_r} \exp\left\{(1-\eta)\ell[\mathbf{D}(\mathcal{P}_2)+\delta_2]\right\}. \qquad (9)$$

The duration τ of a block is ℓ times the number of error-or-erasure tries until acceptance, so $\mathbf{E}[\tau] \leq \ell/(1-P_r)$. Similarly the expected energy $\mathbf{E}[\mathcal{S}_\tau]$ over the entire transmission satisfies $\mathbf{E}[\mathcal{S}_\tau] \leq \mathbf{E}[\mathcal{S}]/(1-P_r)$. Finally, using (8),

$$\frac{\mathbf{E}[\mathcal{S}_\tau]}{\mathbf{E}[\tau]} \leq \ell[\eta\mathcal{P}_1 + (1-\eta)\mathcal{P}_2 + \rho_{\max}e^{-\eta\ell\,\epsilon_1(\delta_1)}].$$

C. Optimization of the bound

The above lemma can be interpreted as providing a nominal rate of transmission, $R = \eta\mathbf{C}(\mathcal{P}_1)$, a nominal power constraint, $\mathcal{P} = \eta\mathcal{P}_1 + (1-\eta)\mathcal{P}_2$, and a nominal exponent of error probability, $E(R,\mathcal{P}) = (1-\eta)\mathbf{D}(\mathcal{P}_2)$. We have shown that codes exist simultaneously approaching each of these parameters arbitrarily closely as ℓ becomes large.

For given $\mathcal{P} > 0$ and $R < \mathbf{C}(\mathcal{P})$, we now maximize the exponent $(1-\eta)\mathbf{D}(\mathcal{P}_2)$ over $\eta, \mathcal{P}_1$, and $\mathcal{P}_2$, subject to the constraints $R = \eta\mathbf{C}(\mathcal{P}_1)$ and $\mathcal{P} = \eta\mathcal{P}_1 + (1-\eta)\mathcal{P}_2$. The first constraint can be satisfied for any $\eta \geq R/\mathbf{C}^*$ and the corresponding $\mathcal{P}_1$ can be restricted without loss of generality to satisfy $\mathcal{P}_1 \leq P^*$ where $\mathcal{P}^*$ is the smallest power that achieves the unconstrained capacity $\mathbf{C}^*$. In this range, $\mathbf{C}(\mathcal{P}_1)$ has an inverse $\mathbf{C}^{-1}$ and $\mathcal{P}_1 = \mathbf{C}^{-1}(R/\eta)$. The second constraint implies that $\eta\mathcal{P}_1 = \eta\mathbf{C}^{-1}(R/\eta) \leq \mathcal{P}$. The function $\eta\mathbf{C}^{-1}(R/\eta)$ is decreasing in η, so this constraint is satisfied for $\eta \geq \eta^*(R,\mathcal{P})$ where η^* satisfies $\eta\mathbf{C}^{-1}(R/\eta) = \mathcal{P}$. This constraint also ensures that $\eta \geq R/\mathbf{C}^*$, so the optimized error exponent can be expressed as

$$E(R,\mathcal{P}) = \sup_{\eta^*(R,\mathcal{P}) \leq \eta < 1} (1-\eta)\mathbf{D}\left(\frac{\mathcal{P} - \eta\mathbf{C}^{-1}(R/\eta)}{1-\eta}\right) \qquad (10)$$

The constraint $\eta\mathbf{C}^{-1}(R/\eta) \leq \mathcal{P}$ is equivalent to $\eta \geq \frac{\mathcal{P}}{\mathcal{E}_{\mathbf{c}}{}^{-1}(\frac{\mathcal{P}}{R})}$, where $\mathcal{E}_{\mathbf{c}}{}^{-1}(\cdot)$, is the inverse of the function $\mathcal{E}_{\mathbf{c}}(x) = \frac{x}{\mathbf{C}(x)}$. By calculating the function $\mathcal{E}_{\mathbf{c}}{}^{-1}(\cdot)$ once, we can avoid checking the condition for $\eta\mathbf{C}^{-1}(R/\eta) \leq \mathcal{P}$ for each $(\mathcal{P}, R)$ pair.[3]

Using the concavity of $\mathbf{D}$ and the convexity of $\mathbf{C}^{-1}$, it is not difficult to show that $E(R,\mathcal{P})$ is concave in the pair $(R,\mathcal{P})$. Figure 2 illustrates this exponent rate function for a given $\mathcal{P}$.

We can now substitute this optimized result into (9), getting

$$P_e \leq \exp -\ell[E(R,\mathcal{P}) - \delta].$$

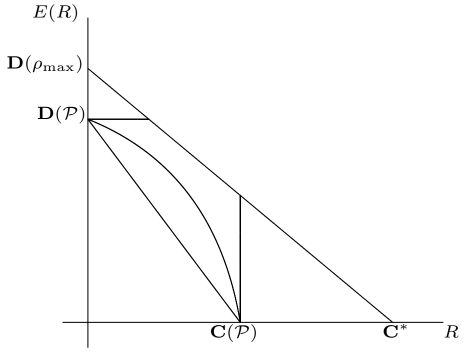

Fig. 2. A typical $E(R,\mathcal{P})$ curve.

From (7), the factor $1/(1-P_r)$ has been absorbed into the arbitrary term δ. The block length ℓ of the error-or-erasure code can similarly be replaced by the expected length of the variable-length block code, getting the following theorem:

Theorem 1: For all $\mathcal{P} \geq 0$, all positive δ and all sufficiently large $\overline{\tau}$, there is a variable-length block code with $M \geq \exp[\overline{\tau}(R-\delta_1)]$ such that for each codeword, the probability of error P_e and the expected energy $\mathbf{E}[\mathcal{S}_\tau]$ for each message satisfy

$$\begin{aligned} P_e &\leq \exp\{-\overline{\tau}[E(R,\mathcal{P}) - \delta]\}, & (11) \\ \mathbf{E}[\mathcal{S}_\tau] &\leq \left(\mathcal{P} + \rho_{\max}e^{-\overline{\tau}\epsilon(\delta)}\right)\overline{\tau}, & (12) \end{aligned}$$

where $\epsilon(\delta) > 0$ for each $\delta > 0$.

For $\mathcal{P} > 0$, the function $E(R,\mathcal{P})$ is continuous and the term $\rho_{\max}e^{-\overline{\tau}\epsilon(\delta)}$ in (12) can be absorbed into the δ in (11). Thus, for large enough $\overline{\tau}$, (12) can be replaced by $\mathbf{E}[\mathcal{S}_\tau] \leq \mathcal{P}\overline{\tau}$, meaning that each codeword satisfies the power constraint. For $\mathcal{P} = 0$, E(R, P) may not be achieved if the constraint is $\mathbf{E}[\mathcal{S}_\tau] \leq 0$; thus (12) must be used as is.

III. The converse: relating $\overline{\tau}$ and P_e

We have established an upper bound on P_e for given rate R, power $\mathcal{P}$, and expected block length $\overline{\tau}$ by developing and analyzing a particular class of algorithms. Here we want to develop a lower bound to P_e. It turns out to be more convenient to establish a lower bound on the expected decoding time $\overline{\tau}$ for a given required error probability P_e, number of messages M, and power constraint $\mathcal{P}$. In this lower bound, we can still use the idea of a two phase analysis, although this does not restrict the coding or decoding.

The analysis is a simplification and generalization of Burnashev [1] and is based on the evolution at each time n of the conditional message entropy, conditioned on the observations at the receiver. The first phase is the interval until this conditional entropy drops from $\ln M$ to some fixed intermediate value, taken here to be 1. The second phase is the interval until this conditional entropy further drops to meet the constraint on error probability; Fano's inequality is used to link the conditional entropy to the error probability. In the first phase we create a stochastic sequence related to the decrease in conditional entropy at each instant n, and in the second phase we create a stochastic sequence related to the decrease in the logarithm of the conditional entropy.

[3] If $\mathbf{C}(\cdot)$ is linear, say $\mathbf{C} = \beta x$ over an interval $(0, x_0)$, then $\mathcal{E}_{\mathbf{c}}(x) = 1/\beta$, in that interval. Then $\mathcal{E}_{\mathbf{c}}{}^{-1}(1/\beta) = x_0$ and $\mathcal{E}_{\mathbf{c}}{}^{-1}$ is undefined for smaller arguments.

Establishing this lower bound to $\overline{\tau}$ is more involved than the upper bound to P_e, since the lower bound must apply to *all* variable-length block codes. We start with a more precise definition of variable-length block coding and decoding algorithms. Then we state some technical lemmas about probability of error and the above stochastic sequences. These are used to outline the proof of a theorem lower bounding the expected decoding time. Finally, this is converted to an upper bound on the reliability function which agrees with the lower bound in section II.

A. Variable length coding and decoding algorithms

In a variable-length block-coding scheme, the transmitter initially receives one of M equiprobable messages from the set $\mathcal{M} = \{1, \dots, M\}$. It transmits successive channel symbols about that message, say message θ, until the receiver makes a decision and releases the decoded message to the user. The time of this decision is a random variable denoted by τ.

Given noiseless feedback, we can restrict attention to coding algorithms in which each input symbol X_n is a deterministic function of message and feedback.[4]

$$X_n = \mathbf{C}_n(\theta, Z^{n-1}) \qquad \forall Z^{n-1}, \forall \theta. \tag{13}$$

The entire observation of the receiver up to time n, including Y^n and any additional random choices, can be summarized by the σ-field $\mathcal{F}_n$ generated by these random variables. The nested sequence of $\mathcal{F}_n$'s is called a filtration $\mathcal{F}$.

A decoding criterion is a decision rule about continuing or stopping the communication, depending on the observations up to that time, i.e., it is a Markov stopping time with respect to the filtration $\mathcal{F}$. The message is also decoded at the stopping time.

At each time n, depending on the realization $\mathfrak{f}_n$ of σ-field $\mathcal{F}_n$, the receiver has an a posteriori probability $p_i(\mathfrak{f}_n)$ for each i in $\mathcal{M}$. Consequently the conditional entropy of the message, given $\mathcal{F}_n$, is a random variable $\mathcal{H}_{\mathcal{F}_n}$, measurable in $\mathcal{F}_n$. Its sample value for any realization $\mathfrak{f}_n \in \mathcal{F}_n$, is given by:

$$\mathcal{H}_{\mathfrak{f}_n} = H(\theta \mid \mathcal{F}_n = \mathfrak{f}_n) = -\sum_{i=1}^{M} p_i(\mathfrak{f}_n) \ln p_i(\mathfrak{f}_n).$$

Fano's inequality can be extended to variable decoding time systems by upper bounding the expected values of the these conditional entropies at the decoding time τ. The result is that for $\mathbf{P}[\tau < \infty] = 1$,

$$\mathbf{E}[\mathcal{H}_{\mathcal{F}_\tau}] \le \mathfrak{h}(P_e) + P_e \ln(M-1), \tag{14}$$

where $\mathfrak{h}(x) = -x\ln(x) - (1-x)\ln(1-x)$.

This suggests that the conditional entropy is usually very small at the decoding time, which motivates focusing on how fast the logarithm of the entropy changes in the second phase of the analysis below.

B. Bounds on the change in conditional entropy

Let

$$V_n^{\mathcal{P}} = \mathcal{H}_{\mathcal{F}_n} + \gamma_{\mathbf{C}}^{\mathcal{P}}(\mathbf{E}[\mathcal{S}_n | \mathcal{F}_n] - n\mathcal{P}) \tag{15}$$

$$W_n^{\mathcal{P}} = \ln \mathcal{H}_{\mathcal{F}_n} + \gamma_{\mathbf{D}}^{\mathcal{P}}(\mathbf{E}[\mathcal{S}_n | \mathcal{F}_n] - n\mathcal{P}). \tag{16}$$

where $\gamma_{\mathbf{C}}^{\mathcal{P}}$ and $\gamma_{\mathbf{D}}^{\mathcal{P}}$ are the Lagrange multipliers for the cost constraints in the optimization problems given in (1) and (4) respectively.

$V_n^{\mathcal{P}}$ will be used to keep track of changes in entropy and cost with time for phase 1, whereas $W_n^{\mathcal{P}}$ will be used to keep track of changes in the logarithm of entropy for phase 2. Using (1) and (4), one can derive the following bounds on the expected change of these random variables in one time unit.

Lemma 2: $\forall n \ge 0$, and $\forall \mathcal{P} \ge 0$,

$$\mathbf{E}\left[V_n^{\mathcal{P}} - V_{n+1}^{\mathcal{P}} \middle| \mathcal{F}_n\right] \le \mathbf{C}(\mathcal{P}) \tag{17}$$

$$\mathbf{E}\left[W_n^{\mathcal{P}} - W_{n+1}^{\mathcal{P}} \middle| \mathcal{F}_n\right] \le \mathbf{D}(\mathcal{P}) \tag{18}$$

The following result relates expected stopping times for a stochastic sequence to expected changes over time:

Lemma 3: Let $\{\Gamma_i\}$ be a stochastic sequence measurable in the filtration $\mathcal{F}$ and let τ_i and τ_f be stopping times with respect to the filtration $\mathcal{F}$ such that $\mathbf{E}[\tau_f] < \infty$ and $\tau_i(w) \le \tau_f(w)$ $\forall w \in \mathcal{F}$. If $\exists K, R \in \Re$ such that

$$\mathbf{E}[|\Gamma_n - \Gamma_{n+1}| | \mathcal{F}_n] < K \quad \text{and} \quad \mathbf{E}[\Gamma_n - \Gamma_{n+1} | \mathcal{F}_n] \le R,$$

then

$$R\mathbf{E}[\tau_f - \tau_i | \mathcal{F}_0] \ge \mathbf{E}\left[\Gamma_{\tau_i} - \Gamma_{\tau_f} \middle| \mathcal{F}_0\right].$$

Evidently these conditions are satisfied by both $V_n^{\mathcal{P}}$ and $W_n^{\mathcal{P}}$, provided that we can uniformly bound the change of $W_n^{\mathcal{P}}$ in one time unit. One can prove that for any $Y_{n+1} = j$,

$$|\ln \mathcal{H}_{\mathcal{F}_n} - \ln \mathcal{H}_{\mathcal{F}_{n+1}}| \le \max_{i,k} \ln \frac{P_{kj}}{P_{ij}} \le \max_{i,k,j} \ln \frac{P_{kj}}{P_{ij}} = \mathbf{F} \tag{19}$$

Consequently we get the following corollary.

Corollary 1: For any pair of stopping times (τ_i, τ_f), and any coding algorithm; if $\tau_f \ge \tau_i$ and $\mathbf{E}\left[\mathcal{S}_{\tau_f} - \mathcal{S}_{\tau_i}\right] \le \mathcal{P}\mathbf{E}[\tau_f - \tau_i]$ then

$$\mathbf{E}\left[\mathcal{H}_{\mathcal{F}_{\tau_i}} - \mathcal{H}_{\mathcal{F}_{\tau_f}}\right] \le \mathbf{C}(\mathcal{P})\mathbf{E}[\tau_f - \tau_i] \tag{20}$$

$$\mathbf{E}\left[\ln \mathcal{H}_{\mathcal{F}_{\tau_i}} - \ln \mathcal{H}_{\mathcal{F}_{\tau_f}}\right] \le \mathbf{D}(\mathcal{P})\mathbf{E}[\tau_f - \tau_i] \tag{21}$$

C. Lower bound for the expected decoding time

The following theorem uses the previous lemmas to lower bound the expected decoding time $\overline{\tau}$. The parameter η below is essentially the fraction of overall decoding time required for $\mathcal{H}_{\mathcal{F}_\tau}$ to first reach 1. Intuitively, we can view $\mathcal{V}_1 \approx \ln M$ and $\mathcal{V}_2 \approx -\ln P_e$ in the sense that $\mathcal{V}_1 / \ln M \to 1$ and $\mathcal{V}_2 / (-\ln P_e) \to 1$ as $\overline{\tau} \to \infty$.

Theorem 2: Given any DMC with feedback, consider a variable-length block code with $M > 2$, $P_e > 0$, and $\mathcal{P} > 0$. If $\mathbf{E}[\mathcal{S}_\tau] \le \mathcal{P}\mathbf{E}[\tau]$, then the expected number of observations $\mathbf{E}[\tau]$ satisfies the inequality

$$\mathbf{E}[\tau] \ge \min_{0<\eta<1; \mathcal{P}_1} \max \left\{ \frac{\mathcal{V}_1}{\eta \mathbf{C}(\mathcal{P}_1)}, \frac{\mathcal{V}_2}{(1-\eta)\mathbf{D}\left(\frac{\mathcal{P}-\eta\mathcal{P}_1}{1-\eta}\right)} \right\}, \tag{22}$$

[4] This allows the receiver to feed back not only the channel outputs but also some random choices. Random choices at the transmitter provide no added generality since those choices (for all possible θ) could be made earlier at the receiver with no loss of performance.

where

$$\mathcal{V}_1 = \ln M \left(1 - P_e(\ln M - \ln P_e + 1) - \frac{1}{\ln M}\right),$$
$$\mathcal{V}_2 = -\ln P_e \left(1 - \frac{\mathbf{F}}{-\ln P_e} - \frac{\ln(\ln M - \ln P_e + 1)}{-\ln P_e}\right).$$

Outline of Proof: Define the stopping time t_1 as the smallest n for which $\mathcal{H}_{\mathcal{F}_n} \leq 1$. Then $\tau_1 = \min(\tau, t_1)$ is also a Markov stopping time and $\tau_1(w) \leq \tau(w)\ \forall w$. Let us define η and $\mathcal{P}_1$, as

$$\eta = \frac{\mathbf{E}[\tau_1]}{\mathbf{E}[\tau]}, \qquad \mathcal{P}_1 = \frac{\mathbf{E}[\mathcal{S}_{\tau_1}]}{\mathbf{E}[\tau_1]}. \tag{23}$$

The main steps of the proof are as follows:

- Use the Markov inequality to upper bound $\mathbf{P}[\mathcal{H}_{\mathcal{F}_\tau} > 1]$ in terms of $\mathbf{E}[\mathcal{H}_{\mathcal{F}_\tau}]$.
- Use this bound and the fact that $\mathcal{H}_{\mathcal{F}_n} \leq \ln M$ to upper bound $\mathbf{E}\left[\mathcal{H}_{\mathcal{F}_{\tau_1}}\right]$.
- Use (23), the upper bound on $\mathbf{E}\left[\mathcal{H}_{\mathcal{F}_{\tau_1}}\right]$, $\mathcal{H}_{\mathcal{F}_0} = \ln M$, and (20) to show that $\mathbf{E}[\tau] \geq \frac{\mathcal{V}_1}{\eta \mathbf{C}(\mathcal{P}_1)}$.
- Use (19) and the definition of τ_1 to lower bound $\mathbf{E}\left[\ln \mathcal{H}_{\mathcal{F}_{\tau_1}}\right]$.
- Use Jensen's inequality and (14) to upper bound $\mathbf{E}[\ln \mathcal{H}_{\mathcal{F}_\tau}]$.
- Use (23), the upper bound on $\mathbf{E}[\ln \mathcal{H}_{\mathcal{F}_\tau}]$, the lower bound on $\mathbf{E}\left[\ln \mathcal{H}_{\mathcal{F}_{\tau_1}}\right]$ and (21) to get $\mathbf{E}[\tau] \geq \frac{\mathcal{V}_2}{(1-\eta)\mathbf{D}\left(\frac{\mathcal{P}-\eta\mathcal{P}_2}{1-\eta}\right)}$.
- The maximum of the above lower bounds on $\mathbf{E}[\tau]$ will be a lower bound on $\mathbf{E}[\tau]$ in terms of η,$\mathcal{P}_1$, and $\mathcal{P}$.
- Minimize the bound over η and $\mathcal{P}_1$ to remove the dependence on η and $\mathcal{P}_1$.

Now we can find a lower bound on $-\ln P_e/\bar{\tau}$, for codes of rate $R = (\ln M)/\bar{\tau}$, using the above result.

Theorem 3: For any DMC with all $P_{jk} > 0$, assume $\mathcal{P} \geq 0$,$R \leq \mathbf{C}(\mathcal{P})$, and $\delta > 0$. Then for all sufficiently large $\mathbf{E}[\tau]$,

$$P_e \geq e^{-\mathbf{E}[\tau](E(R,\mathcal{P})+\delta)}.$$

Proof: Assume that the above statement is not true. Then for some $\delta > 0$, there exists a sequence of codes such that $\lim_{i\to\infty} \mathbf{E}\left[\tau^i\right] = \infty$ and ${P_e}^i < e^{-\mathbf{E}\left[\tau^i\right](E(R,\mathcal{P})+\delta)}$ for all i. Then

$$\lim_{i\to\infty} \frac{{\mathcal{V}_1}^i}{-\ln M^i} = 1 \quad \text{and} \quad \lim_{i\to\infty} \frac{{\mathcal{V}_2}^i}{-\ln {P_e}^i} = 1.$$

Thus for any $1 > \delta_1 > 0$ and for large enough $\mathbf{E}[\tau]$, theorem 2 implies that

$$\mathbf{E}[\tau] \geq \min_{0<\eta<1;\mathcal{P}_1} \max\left\{\frac{(1-\delta_1)\ln M}{\eta \mathbf{C}(\mathcal{P}_1)}, \frac{-(1-\delta_1)\ln P_e}{(1-\eta)\mathbf{D}\left(\frac{\mathcal{P}-\eta\mathcal{P}_1}{1-\eta}\right)}\right\},$$

$$1 \geq \min_{0<\eta<1;\mathcal{P}_1} \max\left\{\frac{(1-\delta_1)R}{\eta \mathbf{C}(\mathcal{P}_1)}, \frac{(1-\delta_1)\frac{-\ln P_e}{\mathbf{E}[\tau]}}{(1-\eta)\mathbf{D}\left(\frac{\mathcal{P}-\eta\mathcal{P}_1}{1-\eta}\right)}\right\},$$

$$(1-\delta_1)\frac{-\ln P_e}{\mathbf{E}[\tau]} \leq E(R(1-\delta),\mathcal{P}).$$

Using the fact that $E(R,P)$ is a decreasing function of R, one can argue that for any $\delta' > 0$ and for large enough $\mathbf{E}[\tau]$,

$$\frac{-\ln P_e}{\mathbf{E}[\tau]} \leq E(R,\mathcal{P}) + \delta'$$

Choose $\delta' = \delta$, $P_e \geq e^{-\mathbf{E}[\tau](E(R,\mathcal{P})+\delta)}$, which contradicts our initial assumption.

IV. The zero error case

The reliability function above relies heavily on the assumption that $P_{ij} > 0$ for all i,j. Here assume $P_{ij} = 0$ for some i,j and $P_{kj} > 0$ for some k and that same j. In this case $D_k = \infty$. Suppose that the 'accept' codeword of section 2 uses all k's, the 'reject' message all i's, and the receiver decodes 'accept' only if it receives one or more j's. In this case, no errors can ever occur for the corresponding variable-length block code. Asymptotically, phase 2 can occupy a negligible portion of the block, so that $\mathbf{C}(\mathcal{P})$ is the zero-error cost constrained capacity of the channel for variable-length block coding. This result is implicitly contained in Burnashev [1] for the DMC without cost constraints.

References

[1] M. V. Burnashev. Data transmission over a discrete channel with feedback, random transmission time. *Problemy Perdachi Informatsii*, 12, No. 4:10–30, 1976.

[2] E.R. Berlekamp C.E. Shannon, R.G. Gallager. Lower bounds to error probability for coding on discrete memoryless channels. *Information and Control*, 10, No. 1:65–103, 1967.

[3] R. L. Dobrushin. An asymptotic bound for the probability error of information transmission through a channel without memory using the feedback. *Problemy Kibernetiki*, vol 8:161–168, 1962.

[4] Robert G. Gallager. *Information Theory and Reliable Communication*. John Wiley & Sons, Inc., New York, NY, USA, 1968.

[5] E. A. Haroutunian. A lower bound of the probability of error for channels with feedback. *Problemy Peredachi Informatsii*, vol 13:36–44, 1977.

[6] A. Kramer. Improving communication reliability by use of an intermittent feedback channel. *Information Theory, IEEE Transactions on*, Vol.15, Iss.1:52–60, 1969.

[7] M. S. Pinsker. The probability of error in block transmission in a memoryless gaussian channel with feedback. *Problemy Perdachi Informatsii*, 4(4):1–4, 1968.

[8] J. P. M. Schalkwijk. A coding scheme for additive noise channels with feedback–ii: Band-limited signals. *Information Theory, IEEE Transactions on*, Vol.12, Iss.2:183–189, 1966.

[9] J. P. M. Schalkwijk and T. Kailath. A coding scheme for additive noise channels with feedback–i: No bandwidth constraint. *Information Theory, IEEE Transactions on*, Vol.12, Iss.2:172–182, 1966.

[10] H. Yamamoto and K. Itoh. Asymptotic performance of a modified schalkwijk-barron scheme for channels with noiseless feedback. *IEEE Transactions on Information Theory*, Vol.25, Iss.6:729–733, 1979.