MIT Open Access Articles

Variations on a Theme by Schalkwijk and Kailath

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation	Gallager, R.G., and B. Nakiboglu. "Variations on a Theme by Schalkwijk and Kailath." Information Theory, IEEE Transactions on 56.1 (2010): 6-17.
As Published	http://dx.doi.org/10.1109/TIT.2009.2034896
Publisher	Institute of Electrical and Electronics Engineers
Version	Final published version
Citable link	http://hdl.handle.net/1721.1/61702
Terms of Use	Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

Massachusetts Institute of Technology

DSpace@MIT

Variations on a Theme by Schalkwijk and Kailath

Robert G. Gallager, *Life Fellow, IEEE*, and Barış Nakiboğlu

***Abstract*—Schalkwijk and Kailath (1966) developed a class of block codes for Gaussian channels with ideal feedback for which the probability of decoding error decreases as a second-order exponent in block length for rates below capacity. This well-known but surprising result is explained and simply derived here in terms of a result by Elias (1956) concerning the minimum mean-square distortion achievable in transmitting a single Gaussian random variable over multiple uses of the same Gaussian channel. A simple modification of the Schalkwijk–Kailath scheme is then shown to have an error probability that decreases with an exponential *order* which is linearly increasing with block length. In the infinite bandwidth limit, this scheme produces zero error probability using bounded expected energy at all rates below capacity. A lower bound on error probability for the finite bandwidth case is then derived in which the error probability decreases with an exponential order which is linearly increasing in block length at the same rate as the upper bound.**

***Index Terms*—Additive memoryless Gaussian noise channel, block codes, error probability, feedback, reliability, Schalkwijk–Kailath encoding scheme.**

I. Introduction

THIS paper describes coding and decoding strategies for discrete-time additive memoryless Gaussian-noise (DAMGN) channels with ideal feedback. It was shown by Shannon [14] in 1961 that feedback does not increase the capacity of memoryless channels, and was shown by Pinsker [10] in 1968 that fixed-length block codes on Gaussian-noise channels with feedback cannot exceed the sphere-packing bound if the energy per codeword is bounded independently of the noise realization. It is clear, however, that reliable communication can be simplified by the use of feedback, as illustrated by standard automatic repeat strategies at the data link control layer. There is a substantial literature (for example, [11], [3], [9]) on using variable-length strategies to substantially improve the rate of exponential decay of error probability with *expected* coding constraint length. These strategies essentially use the feedback to coordinate postponement of the final decision when the noise would otherwise cause errors. Thus, small error probabilities can be achieved through the use of occasional long delays, while keeping the expected delay small.

For DAMGN channels an additional mechanism for using feedback exists whereby the transmitter can transmit unusually large amplitude signals when it observes that the receiver is in danger of making a decoding error. The power (i.e., the expected squared amplitude) can be kept small because these large amplitude signals are rarely required. In 1966, Schalkwijk and Kailath [13] used this mechanism in a fixed-length block-coding scheme for infinite bandwidth Gaussian noise channels with ideal feedback. They demonstrated the surprising result that the resulting probability of decoding error decreases as a second-order exponential[1] in the code constraint length at all transmission rates less than capacity. Schalkwijk [12] extended this result to the finite bandwidth case, i.e., DAMGN channels. Later, Kramer [8] (for the infinite bandwidth case) and Zigangirov [15] (for the finite bandwidth case) showed that the above doubly exponential bounds could be replaced by kth-order exponential bounds for any $k > 2$ in the limit of arbitrarily large block lengths. Later encoding schemes inspired by the Schalkwijk and Kailath approach have been developed for multiuser communication with DAMGN [16]–[20], secure communication with DAMGN [21], and point-to-point communication for Gaussian noise channels with memory [22].

The purpose of this paper is threefold. First, the existing results for DAMGN channels with ideal feedback are made more transparent by expressing them in terms of a 1956 paper by Elias on transmitting a single signal from a Gaussian source via multiple uses of a DAMGN channel with feedback. Second, using an approach similar to that of Zigangirov in [15], we strengthen the results of [8] and [15], showing that error probability can be made to decrease with block length n at least with an exponential order $an - b$ for given coefficients $a > 0$ and $b > 0$. Third, a lower bound is derived. This lower bound decreases with an exponential order in n equal to $an + b'(n)$ where a is the same as in the upper bound and $b'(n)$ is a sublinear function[2] of the block length n.

Neither this paper nor the earlier results in [12], [13], [8], and [15] are intended to be practical. Indeed, these second- and higher order exponents require unbounded amplitudes (see [10], [2], [9]). Also, Kim *et al.* [7] have recently shown that if the feedback is ideal except for additive Gaussian noise, then the error probability decreases only as a single exponential in block length, although the exponent increases with increasing signal-to-noise ratio (SNR) in the feedback channel. Thus, our purpose here is simply to provide increased understanding of the ideal conditions assumed.

We first review the Elias result [4], and use it to get an almost trivial derivation of the Schalkwijk and Kailath results. The derivation yields an exact expression for error probability, optimized over a class of algorithms including those in [12], [13]. The linear processing inherent in that class of algorithms is then relaxed to obtain error probabilities that decrease with block length n at a rate much faster than an exponential order

Manuscript received December 15, 2008; revised August 09, 2009. Current version published December 23, 2009.

The authors are with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: gallager@mit.edu; nakib@mit.edu).

Communicated by G. Kramer, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2009.2034896

[1]For integer $k \geq 1$, the kth-order exponent function $g_k(x)$ is defined as $g_k(x) = \exp(\exp(\cdots(\exp(x))\cdots))$ with k repetitions of $\exp$. A function $f(x) \geq 0$ is said to decrease as a kth-order exponential if for some constant $A > 0$ and all sufficiently large x, $f(x) \leq 1/g_k(Ax)$.

[2]i.e., $\lim_{n\to\infty} \frac{b'(n)}{n} = 0$.

0018-9448/$26.00 © 2009 IEEE

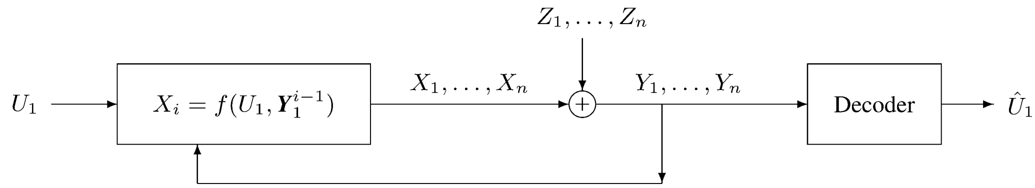

Fig. 1. The setup for n channel uses per source use with ideal feedback.

of 2. Finally, a lower bound to the probability of decoding error is derived. This lower bound is first derived for the case of two codewords and is then generalized to arbitrary rates less than capacity.

II. The Feedback Channel and the Elias Result

Let $X_1, \ldots, X_n = \boldsymbol{X}_1^n$ represent $n > 1$ successive inputs to a DAMGN channel with ideal feedback. That is, the channel outputs $Y_1, \ldots, Y_n = \boldsymbol{Y}_1^n$ satisfy $\boldsymbol{Y}_1^n = \boldsymbol{X}_1^n + \boldsymbol{Z}_1^n$ where $\boldsymbol{Z}_1^n$ is an n-tuple of statistically independent Gaussian random variables, each with zero mean and variance σ_Z^2, denoted $\mathcal{N}(0, \sigma_Z^2)$. The channel inputs are constrained to some given average power constraint S in the sense that the inputs must satisfy the second-moment constraint

$$\frac{1}{n}\sum_{i=1}^{n} S_i \leq S, \quad \text{where } S_i = \mathsf{E}\left[X_i^2\right]. \tag{1}$$

Without loss of generality, we take $\sigma_Z^2 = 1$. Thus, S is both a power constraint and an SNR constraint.

A discrete-time channel is said to have ideal feedback if each output Y_i, $1 \leq i \leq n$, is made known to the transmitter in time to generate input X_{i+1} (see Fig. 1). Let U_1 be the random source symbol to be communicated via this n-tuple of channel uses. Then each channel input X_i is some function $f(U_1, \boldsymbol{Y}_1^{i-1})$ of the source and previous outputs. Assume (as usual) that U_1 is statistically independent of $\boldsymbol{Z}_1^n$.

Elias [4], was interested in the situation where $U_1 \sim \mathcal{N}(0, \sigma_1^2)$ is a Gaussian random variable rather than a discrete message. For $n = 1$, the rate–distortion bound (with a mean-square distortion measure) is achieved without coding or feedback. For $n > 1$, attempts to map U_1 into an n-dimensional channel input in the absence of feedback involve nonlinear or twisted modulation techniques that are ugly at best. Using the ideal feedback, however, Elias constructed a simple and elegant procedure for using the n channel symbols to send U_1 in such a way as to meet the rate–distortion bound with equality.

Let $S_i = \mathsf{E}[X_i^2]$ be an arbitrary choice of energy, i.e., second moment, for each i, $1 \leq i \leq n$. It will be shown shortly that the optimal choice for $S_1, \ldots, S_n$, subject to (1), is $S_i = S$ for $1 \leq i \leq n$. Elias's strategy starts by choosing the first transmitted signal X_1 to be a linear scaling of the source variable U_1, scaled to meet the second-moment constraint, i.e.,

$$X_1 = \frac{\sqrt{S_1}U_1}{\sigma_1}.$$

At the receiver, the minimum mean-square error (MMSE) estimate of X_1 is $\mathsf{E}[X_1|Y_1] = \frac{S_1 Y_1}{1+S_1}$, and the error in that estimate is $\mathcal{N}(0, \frac{S_1}{1+S_1})$. It is more convenient to keep track of the MMSE estimate of U_1 and the error U_2 in that estimate. Since U_1 and X_1 are the same except for the scale factor $\sigma_1/\sqrt{S_1}$, these are given by

$$\mathsf{E}[U_1|Y_1] = \frac{\sigma_1\sqrt{S_1}\,Y_1}{1+S_1} \tag{2}$$

$$U_2 = U_1 - \mathsf{E}[U_1|Y_1] \tag{3}$$

where $U_2 \sim \mathcal{N}(0, \sigma_2^2)$ and $\sigma_2^2 = \frac{\sigma_1^2}{1+S_1}$.

Using the feedback, the transmitter can calculate the error term U_2 at time 2. Elias's strategy is to use U_2 as the source signal (without a second-moment constraint) for the second transmission. This unconstrained signal U_2 is then linearly scaled to meet the second-moment constraint S_2 for the second transmission. Thus, the second transmitted signal X_2 is given by

$$X_2 = \frac{\sqrt{S_2}U_2}{\sigma_2}.$$

We use this notational device throughout, referring to the unconstrained source signal to be sent at time i by U_i and to the linear scaling of U_i, scaled to meet the second moment constraint S_i, as X_i.

The receiver calculates the MMSE estimate $\mathsf{E}[U_2|Y_2] = \frac{\sigma_2\sqrt{S_2}\,Y_2}{1+S_2}$ and the transmitter then calculates the error in this estimate, $U_3 = U_2 - \mathsf{E}[U_2|Y_2]$. Note that

$$\begin{aligned} U_1 &= U_2 + \mathsf{E}[U_1|Y_1] \\ &= U_3 + \mathsf{E}[U_2|Y_2] + \mathsf{E}[U_1|Y_1]. \end{aligned}$$

Thus, U_3 can be viewed as the error arising from estimating U_1 by $\mathsf{E}[U_1|Y_1] + \mathsf{E}[U_2|Y_2]$. The receiver continues to update its estimate of U_1 on subsequent channel uses, and the transmitter continues to transmit linearly scaled versions of the current estimation error. Then the general expressions are as follows:

$$X_i = \frac{\sqrt{S_i}U_i}{\sigma_i} \tag{4}$$

$$\mathsf{E}[U_i|Y_i] = \frac{\sigma_i\sqrt{S_i}\,Y_i}{1+S_i} \tag{5}$$

$$U_{i+1} = U_i - \mathsf{E}[U_i|Y_i] \tag{6}$$

where $U_{i+1} \sim \mathcal{N}(0, \sigma_{i+1}^2)$ and $\sigma_{i+1}^2 = \frac{\sigma_i^2}{1+S_i}$.

Iterating on (6) from $i = 1$ to n yields

$$U_{n+1} = U_1 - \sum_{i=1}^{n} \mathsf{E}[U_i|Y_i]. \tag{7}$$

Similarly, iterating on $\sigma_{i+1}^2 = \sigma_i^2/(1+S_i)$, we get

$$\sigma_{n+1}^2 = \frac{\sigma_1^2}{\prod_{i=1}^{n}(1+S_i)}. \tag{8}$$

This says that the error arising from estimating U_1 by $\sum_{i=1}^{n} \mathsf{E}[U_i|Y_i]$ is $\mathcal{N}(0, \sigma_{n+1}^2)$. This is valid for any (nonnegative) choice of $S_1, \ldots, S_n$, and this is minimized, subject to $\sum_{i=1}^{n} S_i = nS$, by $S_i = S$ for $1 \le i \le n$. With this optimal assignment, the mean square estimation error in U_1 after n channel uses is

$$\sigma_{n+1}^2 = \frac{\sigma_1^2}{(1+S)^n}. \tag{9}$$

We now show that this is the MMSE over all ways of using the channel. The rate–distortion function for this Gaussian source with a squared-difference distortion measure is well known to be

$$R(d) = \frac{1}{2} \ln \frac{\sigma_1^2}{d}.$$

This is the minimum mutual information, over all channels, required to achieve a mean-square error (distortion) equal to d. For $d = \sigma_1^2/(1+S)^n$, $R(d)$ is $\frac{n}{2}\ln(1+S)$, which is the capacity of this channel over n uses (it was shown by Shannon [14] that feedback does not increase the capacity of memoryless channels). Thus, the Elias scheme actually meets the rate–distortion bound with equality, and no other coding system, no matter how complex, can achieve a smaller mean-square error. Note that (9) is also valid in the degenerate case $n = 1$. What is surprising about this result is not so much that it meets the rate–distortion bound, but rather that the mean-square estimation error goes down geometrically with n. It is this property that leads directly to the doubly exponential error probability of the Schalkwijk–Kailath scheme.

III. The Schalkwijk–Kailath Scheme

The Schalkwijk and Kailath (SK) scheme will now be defined in terms of the Elias scheme,[3] still assuming the discrete-time channel model of Fig. 1 and the power constraint of (1). The source is a set of M equiprobable symbols, denoted by $\{1, 2, \ldots, M\}$. The channel uses will now be numbered from 0 to $n-1$, since the use at time 0 will be quite distinct from the others. The source signal, U_0 is a standard M-PAM modulation of the source symbol. That is, for each symbol m, $1 \le m \le M$, from the source alphabet, m is mapped into the signal a_m where $a_m = m - (M+1)/2$. Thus, the M signals in U_0 are symmetric around 0 with unit spacing. Assuming equiprobable symbols, the second moment σ_0^2 of U_0 is $(M^2-1)/12$. The initial channel input X_0 is a linear scaling of U_0, scaled to have an energy S_0 to be determined later. Thus, X_0 is an M-PAM encoding, with signal separation $d_0 = \sqrt{S_0}/\sigma_0$.

$$X_0 = U_0 \sqrt{\frac{S_0}{\sigma_0^2}} = U_0 \sqrt{\frac{S_0}{12(M^2-1)}}. \tag{10}$$

The received signal $Y_0 = X_0 + Z_0$ is fed back to the transmitter, which, knowing X_0, determines Z_0. In the following $n-1$ channel uses, the Elias scheme is used to send the Gaussian random variable Z_0 to the receiver, thus reducing the effect of the noise on the original transmission. After the $n-1$ transmissions to convey Z_0, the receiver combines its estimate of Z_0 with Y_0 to get an estimate of X_0, from which the M-ary signal is detected.

Specifically, the transmitted and received signals for times $1 \le i \le n-1$ are given by (4), (5), and (6). At time 1, the unconstrained signal U_1 is Z_0 and $\sigma_1^2 = \mathsf{E}[U_1^2] = 1$. Thus, the transmitted signal X_1 is given by $\sqrt{S_1}U_1$, where the second moment S_1 is to be selected later. We choose $S_i = S_1$ for $1 \le i \le n-1$ for optimized use of the Elias scheme, and thus the power constraint in (1) becomes $S_0 + (n-1)S_1 = nS$. At the end of transmission $n-1$, the receiver's estimate of Z_0 from $Y_1, \ldots, Y_{n-1}$ is given by (7) as

$$\mathsf{E}\left[Z_0|\boldsymbol{Y}_1^{n-1}\right] = \sum_{i=1}^{n-1} \mathsf{E}[U_i|Y_i].$$

The error in this estimate, $U_n = Z_0 - \mathsf{E}[Z_0|\boldsymbol{Y}_1^{n-1}]$, is a zero-mean Gaussian random variable with variance σ_n^2, where σ_n^2 is given by (9) to be

$$\sigma_n^2 = \frac{1}{(1+S_1)^{n-1}}. \tag{11}$$

Since $Y_0 = X_0 + Z_0$ and $Z_0 = \mathsf{E}[Z_0|\boldsymbol{Y}_1^{n-1}] + U_n$ we have

$$Y_0 - \mathsf{E}\left[Z_0|\boldsymbol{Y}_1^{n-1}\right] = X_0 + U_n \tag{12}$$

where $U_n \sim \mathcal{N}(0, \sigma_n^2)$.

Note that $U_n \sim \mathcal{N}(0, \sigma_n^2)$ is a function of the noise vector $\boldsymbol{Z}_0^{n-1}$ and is thus statistically independent[4] of X_0. Thus, detecting X_0 from $Y_0 - \mathsf{E}[Z_0|\boldsymbol{Y}_1^{n-1}]$ (which is known at the receiver) is the simplest of classical detection problems, namely, that of detecting an M-PAM signal X_0 from the signal plus an independent Gaussian noise variable U_n. Using maximum-likelihood (ML) detection, an error occurs only if U_n exceeds half the distance between signal points, i.e., if

$$|U_n| \ge \frac{1}{2}\frac{\sqrt{S_0}}{\sigma_0} = \frac{1}{2}\sqrt{\frac{12S_0}{M^2-1}}.$$

Since the variance of U_n is $(1+S_1)^{-n+1}$, the probability of error is given by[5]

$$P_e = 2\frac{(M-1)}{M} Q(\gamma_n) \tag{13}$$

where $\gamma_n = \frac{1}{2}\sqrt{\frac{12S_0(1+S_1)^{n-1}}{M^2-1}}$ and $Q(x)$ is the complementary distribution function of $\mathcal{N}(0,1)$, i.e.,

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(\frac{-z^2}{2}\right) dz. \tag{14}$$

Choosing S_0 and S_1, subject $S_0 + (n-1)S_1 = nS$, to maximize γ_n (and thus minimize P_e), we get $S_1 = \max\{0, S - \frac{1}{n}\}$. That

[3]The analysis here is tutorial and was carried out in slightly simplified form in [5, p. 481]. A very readable further simplified analysis is in [23].

[4]Furthermore, for the given feedback strategy, Gaussian estimation theory can be used to show, first, that U_n is independent of $\mathsf{E}[Z_0|\boldsymbol{Y}_1^{n-1}]$, and, second, that $\tilde{Y} = Y_0 - \mathsf{E}[Z_0|\boldsymbol{Y}_1^{n-1}]$ is a sufficient statistic for X_0 based on $\boldsymbol{Y}_0^{n-1}$ (i.e., $\Pr[X_0|\boldsymbol{Y}_0^{n-1}] = \Pr[X_0|\tilde{Y}]$). Thus, this detection strategy is not as *ad hoc* as it might initially seem.

[5]The term $(M-1)/M$ in (13) arises because the largest and smallest signals each have only one nearest neighbor, whereas all other signals have two nearest neighbors.

is, if nS is less than 1, all the energy is used to send X_0 and the feedback is unused. We assume $nS > 1$ in what follows, since for any given $S > 0$ this holds for large enough n. In this case, S_0 is one unit larger than S_1, leading to

$$S_1 = S - \frac{1}{n}; \quad S_0 = S_1 + 1. \tag{15}$$

Substituting (15) into (13)

$$P_e = 2\frac{(M-1)}{M}Q(\gamma_n) \tag{16}$$

where $\gamma_n = \sqrt{\frac{3(1+S-\frac{1}{n})^n}{M^2-1}}$.

This is an *exact* expression for error probability, optimized over energy distribution, and using M-PAM followed by the Elias scheme and ML detection. It can be simplified as an upper bound by replacing the coefficient $\frac{M-1}{M}$ by 1. Also, since $Q(\cdot)$ is a decreasing function of its argument, P_e can be further upper-bounded by replacing M^2-1 by M^2. Thus

$$P_e \le 2Q(\gamma_n) \tag{17}$$

where $\gamma_n \ge \sqrt{3}(1 - \frac{1}{(1+S)n})^{n/2}\frac{(1+S)^{n/2}}{M}$.

For large M, which is the case of interest, the above bound is very tight and is essentially an equality, as first derived by Schalkwijk[6] in [12, eq. (12)]. Recalling that $nS \ge 1$, we can further lower-bound γ_n (thus upper-bounding P_e). Substituting $C(S) = \frac{1}{2}\ln(1+S)$ and $M = \exp(nR)$ we get

$$\gamma_n \ge \left[\sqrt{3}\left(1 - \frac{1}{1+n}\right)^{n/2}\right]\exp(n(C(S)-R)). \tag{18}$$

The term in brackets is decreasing in n. Thus

$$\left(1 - \frac{1}{1+n}\right)^{n/2} \ge \lim_{k\to\infty}\left(1 - \frac{1}{1+k}\right)^{k/2} \tag{19}$$

$$\ge e^{-1/2}, \quad \forall n \ge 1. \tag{20}$$

Using this together with (17) and (18) we get

$$P_e \le 2Q\left(\sqrt{\frac{3}{e}}\exp(n(C(S)-R))\right), \tag{21}$$

or more simply yet,

$$P_e \le 2Q(\exp[n(C(S)-R)]). \tag{22}$$

Note that for $R < C(S)$, P_e decreases as a second-order exponential in n.

In summary, then, we see that the use of standard M-PAM at time 0, followed by the Elias algorithm over the next $n-1$ transmissions, followed by ML detection, gives rise to a probability of error P_e that decreases as a second-order exponential for all $R < C(S)$. Also, P_e satisfies (21) and (22) for all $n \ge 1/S$.

Although P_e decreases as a second-order exponential with this algorithm, the algorithm does not minimize P_e over all algorithms using ideal feedback. The use of standard M-PAM at time 0 could be replaced by pulse-amplitude modulation (PAM) with nonequal spacing of the signal points for a modest reduction in P_e. Also, as shown in the next section, allowing transmissions 1 to $n-1$ to make use of the discrete nature of X_0 allows for a major reduction in P_e.[7]

The algorithm above, however, does have the property that it is optimal among schemes in which, first, standard PAM is used at time 0 and, second, for each i, $1 \le i \le n-1$, X_i is a linear function of Z_0 and Y_1^{i-1}. The reason for this is that Z_0 and Y_1^{n-1} are then jointly Gaussian and the Elias scheme minimizes the mean-square error in Z_0 and thus also minimizes P_e.

A. *Broadband Analysis*

Translating these results to a continuous time formulation where the channel is used $2W$ times per second,[8] the capacity (in nats per second) is $C_W = 2WC$. Letting $T = n/2W$ and letting $R_W = 2WR$ be the rate in nats per second, this formula becomes

$$P_e \le 2\,Q(\exp[(C_W - R_W)T]). \tag{23}$$

Let $\mathcal{P} = 2WS$ be the continuous-time power constraint, so that $C_W = W\ln(1+\mathcal{P}/2W)$. In the broadband limit as $W \to \infty$ for fixed $\mathcal{P}$, $C_W \to \mathcal{P}/2$. Since (23) applies for all $W > 0$, we can simply go to the broadband limit, $C_\infty = \mathcal{P}/2$. Since the algorithm is basically a discrete time algorithm, however, it makes more sense to view the infinite bandwidth limit as a limit in which the number of available degrees of freedom n increases faster than linearly with the constraint time T. In this case, the SNR per degree of freedom, $S = \mathcal{P}T/n$ goes to 0 with increasing T. Rewriting γ_n in (17) for this case

$$\gamma_n \ge \sqrt{3}\exp\left[\frac{n}{2}\ln\left(1 + \frac{\mathcal{P}T}{n} - \frac{1}{n}\right) - TR_\infty\right] \tag{24}$$

$$\ge \sqrt{3}\exp\left[\frac{\mathcal{P}T}{2} - \frac{1}{2} - \frac{\mathcal{P}^2T^2}{4n} - TR_\infty\right] \tag{25}$$

where the inequality $\ln(1+x) \ge x - x^2/2$ was used. Note that if n increases quadratically with T, then the term $\frac{\mathcal{P}^2T^2}{4n}$ is simply a constant which becomes negligible as the coefficient on the quadratic becomes large. For example, if $n \ge 6\mathcal{P}^2T^2$, then this term is at most $1/24$ and (25) simplifies to

$$\gamma_n \ge \exp[T(C_\infty - R_\infty)], \quad \text{for } n \ge 6\mathcal{P}^2T^2. \tag{26}$$

[6]Schalkwijk's work was independent of Elias's. He interpreted the steps in the algorithm as successive improvements in estimating X_0 rather than as estimating Z_0.

[7]Indeed, Zigangirov [15] developed an algorithm quite similar to that developed in the next section. The initial phase of that algorithm is very similar to the algorithm [12] just described, with the following differences. Instead of starting with standard M-PAM, [15] starts with a random ensemble of non-equally spaced M-PAM codes ingeniously arranged to form a Gaussian random variable. The Elias scheme is then used, starting with this Gaussian random variable. Thus, the algorithm in [15] has different constraints than those above. It turns out to have an insignificantly larger P_e (over this phase) than the algorithm here for S greater than $[(1/\ln\frac{6}{\pi}) - 1]$ and an insignificantly smaller P_e otherwise.

[8]This is usually referred to as a channel band-limited to W. This is a harmless and universally used abuse of the word bandwidth for channels without feedback, and refers to the ability to satisfy the Nyquist criterion with arbitrarily little power sent out of band. It is more problematic with feedback, since it assumes that the sum of the propagation delay, the duration of the transmit pulse, the duration of the matched filter at the receiver, and the corresponding quantities for the feedback, is at most $1/2W$. Even allowing for a small fraction of out-of-band energy, this requires considerably more than bandwidth W.

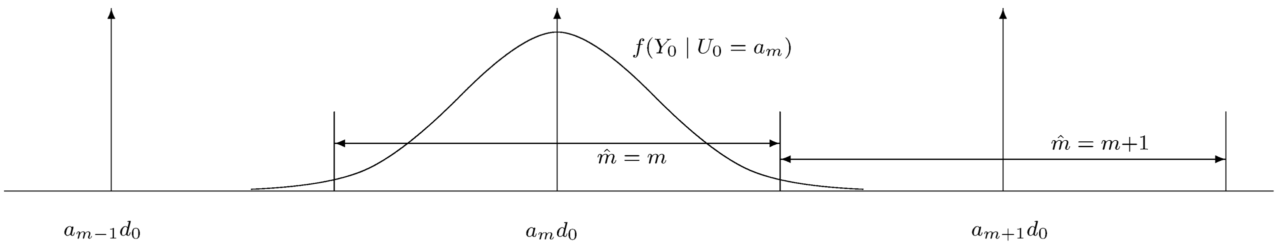

Fig. 2. Given that a_m is the sample value of the PAM source signal U_0, the sample value of X_0 is $a_m d_0$ where $d_0 = \sqrt{S_0}/\sigma_0$. The figure illustrates the probability density of Y_0 given this conditioning and shows the M-PAM signal points for X_0 that are neighbors to the sample value $X_0 = a_m d_0$. Note that this density is $\mathcal{N}(a_m d_0,\, 1)$, i.e., it is the density of Z_0, shifted to be centered at $a_m d_0$. Detection using ML at this point simply quantizes Y_0 to the nearest signal point.

This is essentially the same as the broadband SK result (see the final equation in [13]). The result in [13] used $n = e^{2TC_W}$ degrees of freedom, but chose the subsequent energy levels to be decreasing harmonically, thus slightly weakening the coefficient of the result. The broadband result is quite insensitive to the energy levels used for each degree of freedom[9], so long as S_0 is close to 1 and the other S_i are close to 0. This partly explains why the harmonic choice of energy levels in [13] comes reasonably close to the optimum result.

IV. An Alternative PAM Scheme in the High Signal-to-Noise Regime

In the previous section, Elias's scheme was used to allow the receiver to estimate the noise Z_0 originally added to the PAM signal at time 0. This gave rise to an equivalent observation, $Y_0 - \mathsf{E}[Z_0|\boldsymbol{Y}_1^{n-1}]$ with attenuated noise U_n as given in (12). The geometric attenuation of $\mathsf{E}[U_n^2]$ with n is the reason why the error probability in the Schalkwijk and Kailath (SK) [13] scheme decreases as a second-order exponential in time.

In this section, we explore an alternative strategy that is again based on the use of M-PAM at time 0, but is quite different from the SK strategy at times 1 to $n-1$. The analysis is restricted to situations in which the SNR at time 0 is so large that the distance between successive PAM signal points in X_0 is large relative to the standard deviation of the noise. In this high-SNR regime, a simpler and more effective strategy than the Elias scheme suggests itself (see Fig. 2). This new strategy is limited to the high-SNR regime, but Section V develops a two-phase scheme that uses the SK strategy for the first part of the block, and switches to this new strategy when the SNR is sufficiently large.

In this new strategy for the high-SNR regime, the receiver makes a tentative ML decision $\hat{m}_0$ at time 0. As seen in the figure, that decision is correct unless the noise exceeds half the distance $d_0 = \sqrt{S_0}/\sigma_0$ to either the signal value on the right or the left of the sample value a_m of U_0. Each of these two events has probability $Q(d_0/2)$.

The transmitter uses the feedback to calculate $\hat{m}_0$ and chooses the next signal U_1 (in the absence of a second-moment constraint) to be a shifted version of the original M-PAM signal, shifted so that $U_1 = \hat{m}_0 - m$ where m is the original message symbol being transmitted. In other words, U_1 is the integer-valued error in the receiver's tentative decision $a_{\hat{m}_0}$ of U_0. The corresponding transmitted signal X_1 is essentially given by $X_1 = U_1\sqrt{S_1/\mathsf{E}[U_1^2]}$, where S_1 is the energy allocated to X_1.

We now give an approximate explanation of why this strategy makes sense and how the subsequent transmissions are chosen. This is followed by a precise analysis. Temporarily ignoring the case where either $m = 1$ or $m = M$ (i.e., where a_m has only one neighbor), U_1 is 0 with probability $1 - 2Q(d_0/2)$. The probability that $|U_1|$ is two or more is essentially negligible, so $U_1 = \pm 1$ with a probability approximately equal to $2Q(d_0/2)$. Thus

$$E\left[U_1^2\right] \approx 2Q(d_0/2); \quad X_1 \approx \frac{U_1\sqrt{S_1}}{\sqrt{2Q(d_0/2)}}. \tag{27}$$

This means that X_1 is not only a shifted version of X_0, but (since $d_0 = \sqrt{S_0}/\sigma_0$) is also scaled up by a factor that is exponential in S_0 when S_0 is sufficiently large. Thus, the separation between adjacent signal points in X_1 is exponentially increasing with S_0.

This also means that when X_1 is transmitted, the situation is roughly the same as that in Fig. 2, except that the distance between signal points is increased by a factor exponential in S_0. Thus, a tentative decision at time 1 will have an error probability that decreases as a second-order exponential in S_0.

Repeating the same procedure at time 2 will then give rise to a third-order exponential in S_0, etc. We now turn to a precise analysis and description of the algorithm at times 1 to $n-1$.

The following lemma provides an upper bound to the second moment of U_1, which was approximated in (27).

Lemma 4.1: For any $d \geq 4$, let U be a d-quantization of a normal random variable $Z \sim \mathcal{N}(0,1)$ in the sense that for each integer ℓ, if $Z \in (d\ell - \frac{d}{2}, d\ell + \frac{d}{2}]$, then $U = \ell$. Then $\mathsf{E}[U^2]$ is upper-bounded by

$$\mathsf{E}[U^2] \leq \frac{1.6}{d}\exp\left[-\frac{d^2}{8}\right] \tag{28}$$

Note from Fig. 2 that, aside from a slight exception described below, $U_1 = \hat{m}_0 - m$ is the same as the d_0-quantization of Z_0 where $d_0 = \sqrt{S_0}/\sigma_0$. The slight exception is that $\hat{m}_0$ should always lie between 1 and M. If $Z_0 > (M - m + 1/2)$, then $U_1 = M - m$, whereas the d_0-quantization takes on a larger integer value. There is a similar limit for $Z_0 < 1 - m - 1/2$. This reduces the magnitude of U_1 in the above exceptional cases, and thus reduces the second moment. Thus, the bound in the lemma also applies to U_1. For simplicity, in what follows we avoid this complication by assuming that the receiver allows $\hat{m}_0$ to be

[9] To see this, replace $(1+S_1)^{(n-1)/2}$ in (13) by $\frac{1}{2}\exp[\sum_i \ln(1+S_i)]$, each term of which can be lower-bounded by the inequality $\ln(1+x) \geq x - x^2/2$.

larger than M or smaller than 1. This increases both the error probability and the energy over true ML tentative decisions, so the bounds also apply to the case with true ML tentative decisions.

Proof: From the definition of U, we see that $U = \ell$ if $Z \in (d\ell - \frac{d}{2}, d\ell + \frac{d}{2}]$. Thus, for $\ell \geq 1$

$$\Pr[U = \ell] = Q\left(d\ell - \frac{d}{2}\right) - Q\left(d\ell + \frac{d}{2}\right).$$

From symmetry, $\Pr[U = -\ell] = \Pr[U = \ell]$, so the second moment of U is given by

$$\begin{aligned}\mathsf{E}[U^2] &= 2\sum_{\ell=1}^{\infty} \ell^2 \left[Q\left(d\ell - \frac{d}{2}\right) - Q\left(d\ell + \frac{d}{2}\right)\right] \\ &= 2Q(d/2) + 2\sum_{\ell=2}^{\infty}[\ell^2 - (\ell-1)^2]\left[Q\left(d\ell - \frac{d}{2}\right)\right].\end{aligned}$$

Using the standard upper bound $Q(x) \leq \frac{1}{\sqrt{2\pi}\,x}\exp[-x^2/2]$ for $x > 0$, and recognizing that $\ell^2 - (\ell-1)^2 = 2\ell - 1$, this becomes

$$\begin{aligned}\mathsf{E}[U^2] &\leq \frac{4}{\sqrt{2\pi}\,d}\left\{\exp[-d^2/8] + \sum_{\ell=2}^{\infty}\exp[-(2\ell-1)^2 d^2/8]\right\} \\ &= \frac{4}{\sqrt{2\pi}\,d}\exp[-d^2/8]\left\{1 + \sum_{\ell=2}^{\infty}\exp[-4\ell(\ell-1)d^2/8]\right\} \\ &\leq \frac{4}{\sqrt{2\pi}\,d}\exp[-d^2/8]\left\{\frac{1}{1-\exp(-d^2)}\right\} \\ &\leq \frac{1.6}{d}\exp\left[-\frac{d^2}{8}\right], \quad \text{for } d \geq 4. \qquad (29)\end{aligned}$$

□

We now define the rest of this new algorithm. We have defined the unconstrained signal U_1 at time 1 to be $\hat{m}_0 - m$ but have not specified the energy constraint to be used in amplifying U_1 to X_1. The analysis is simplified by defining X_1 in terms of a specified scaling factor between U_1 and X_1. The energy in X_1 is determined later by this scaling. In particular, let

$$X_1 = d_1 U_1, \quad \text{where } d_1 = \sqrt{8}\exp\left(\frac{d_0^2}{16}\right).$$

The peculiar expression for d_1 above looks less peculiar when expressed as $d_1^2/8 = \exp(d_0^2/8)$. When $Y_1 = X_1 + Z_1$ is received, we can visualize the situation from Fig. 2 again, where now d_0 is replaced by d_1. The signal set for X_1 is again a PAM set but it now has signal spacing d_1 and is centered on the signal corresponding to the transmitted source symbol m. The signals are no longer equally likely, but the analysis is simplified if an ML tentative decision $\hat{m}_1$ is again made. We see that $\hat{m}_1 = \hat{m}_0 - \hat{Y}_1$ where $\hat{Y}_1$ is the d_1-quantization of Y_1 (and where the receiver again allows $\hat{m}_1$ to be an arbitrary integer). We can now state the algorithm for each time i, $1 \leq i \leq n-1$.

$$d_i = \sqrt{8}\exp\left(\frac{d_{i-1}^2}{16}\right) \qquad (30)$$

$$X_i = d_i U_i \qquad (31)$$

$$\hat{m}_i = \hat{m}_{i-1} - \hat{Y}_i \qquad (32)$$

$$U_{i+1} = \hat{m}_i - m. \qquad (33)$$

where $\hat{Y}_i$ is the d_i-quantization of Y_i.

Lemma 4.2: For $d_0 \geq 4$, the algorithm of (31)–(33) satisfies the following for all alphabet sizes M and all message symbols m:

$$\frac{d_i^2}{8} = g_i\left(\frac{d_0^2}{8}\right) \geq g_i(2) \qquad (34)$$

$$\mathsf{E}[X_i^2] \leq \frac{12.8}{d_{i-1}} \qquad (35)$$

$$\sum_{i=1}^{\infty}\mathsf{E}[X_i^2] \leq 5 \qquad (36)$$

$$\Pr(\hat{m}_i \neq m) \leq 1/g_{i+1}(2) \qquad (37)$$

where $g_i(x) = \exp(\cdots(\exp(x))\cdots)$ with i exponentials.

Proof: From the definition of d_i in (30)

$$\frac{d_i^2}{8} = \exp\left(\frac{d_{i-1}^2}{8}\right) = \exp\left(\exp\left(\frac{d_{i-2}^2}{8}\right)\right) = \cdots = g_i\left(\frac{d_0^2}{8}\right).$$

This establishes the first part of (34) and the inequality follows since $d_0 \geq 4$ and $g_i(x)$ is increasing in x.

Next, since $X_i = d_i U_i$, we can use (34) and Lemma 4.1 to see that

$$\begin{aligned}\mathsf{E}\left[X_i^2\right] &= d_i^2\,\mathsf{E}\left[U_i^2\right] \\ &= \left(8\exp\left(\frac{d_{i-1}^2}{8}\right)\right)\left(\frac{1.6}{d_{i-1}}\exp\left(-\frac{d_{i-1}^2}{8}\right)\right) \\ &\leq \frac{12.8}{d_{i-1}}\end{aligned}$$

where we have canceled the exponential terms, establishing (35).

To establish (36), note that each d_i is increasing as a function of d_0, and thus each $\mathsf{E}[X_i^2]$ is upper-bounded by taking $d_0 \geq 4$ to be 4. Then $\mathsf{E}[X_1^2] = 3.2$, $\mathsf{E}[X_2^2] = 1.6648$, and the other terms can be bounded in a geometric series with a sum less than 0.12.

Finally, to establish (37), note that

$$\begin{aligned}\Pr(\hat{m}_i \neq m) = \Pr(|U_i|^2 \geq 1) \quad &\leq \mathsf{E}\left[U_{i+1}^2\right] \\ &\overset{(a)}{\leq} \frac{1.6}{d_i}\exp(-d_i^2/8) \overset{(b)}{\leq} \exp(-d_i^2/8) \\ &\overset{(c)}{=} 1/\exp(g_i(d_0^2/8)) \overset{(d)}{\leq} 1/g_{i+1}(2),\end{aligned}$$

where we have used Lemma 4.1 in (a), the fact that $d_i \geq 4$ in (b), and (34) in (c) and (d). □

We have now shown that, in this high-SNR regime, the error probability decreases with time i as an ith-order exponent. The constants involved, such as $d_0 \geq 4$, are somewhat *ad hoc*, and the details of the derivation are similarly *ad hoc*. What is happening, as stated before, is that by using PAM centered on the receiver's current tentative decision, one can achieve rapidly expanding signal point separation with small energy. This is the critical idea driving this algorithm, and in essence this idea was used earlier by[10] Zigangirov [15]

V. A Two-Phase Strategy

We now combine the Shalkwijk–Kailath (SK) scheme of Section III and the high-SNR scheme of Section IV into a two-phase strategy. The first phase, of block length n_1, uses the SK scheme. At time n_1-1, the equivalent received signal $Y_0 - \mathsf{E}[Z_0|\boldsymbol{Y}_1^{n_1-1}]$ (see (12)) is used in an ML decoder to detect the original PAM signal X_0 in the presence of additive Gaussian noise of variance $\sigma_{n_1}^2$.

Note that if we scale the equivalent received signal, $Y_0 - \mathsf{E}[Z_0|\boldsymbol{Y}_1^{n_1-1}]$ by a factor of $1/\sigma_{n_1}$ so as to have an equivalent unit variance additive noise, we see that the distance between adjacent signal points in the normalized PAM is $d_{n_1-1} = 2\gamma_{n_1}$ where γ_{n_1} is given in (13). If n_1 is selected to be large enough to satisfy $d_{n_1-1} \geq 4$, then this detection at time n_1-1 satisfies the criterion assumed at time 0 of the high-SNR algorithm of Section IV. In other words, the SK algorithm not only achieves the error probability calculated in Section III, but also, if the block length of the SK phase n_1 is chosen to be large enough, it creates the initial condition for the high-SNR algorithm. That is, it provides the receiver and the transmitter at time n_1-1 with the output of a high SNR PAM. Consequently, not only is the tentative ML decision at time n_1-1 correct with moderately high probability, but also the probability of the distant neighbors of the decoded messages vanishes rapidly.

The intuition behind this two-phase scheme is that the SK algorithm seems to be quite efficient when the signal points are so close (relative to the noise) that the discrete nature of the signal is not of great benefit. When the SK scheme is used enough times, however, the signal points become far apart relative to the noise, and the discrete nature of the signal becomes important. The increased effective distance between the signal points of the original PAM also makes the high-SNR scheme feasible. Thus, the two-phase strategy switches to the high-SNR scheme at this point and the high-SNR scheme drives the error probability to 0 as an n_2-order exponential.

We now turn to the detailed analysis of this two-phase scheme. Note that five units of energy must be reserved for phase 2 of the algorithm, so the power constraint S_1 for the first phase of the algorithm is $n_1 S_1 = nS - 5$. For any fixed rate $R < C(S)$, we will find that the remaining $n_2 = n - n_1$ time units are a linearly increasing function of n and yield an error probability upper-bounded by $1/g_{n_2+1}(2)$.

A. The Finite-Bandwidth Case

For the finite-bandwidth case, we assume an overall block length $n = n_1 + n_2$, an overall power constraint S, and an overall rate $R = (\ln M)/n$. The overall energy available for phase 1 is at least $nS-5$, so the average power in phase 1 is at least $(nS-5)/n_1$.

We observed that the distance d_{n_1-1} between adjacent signal points, assuming that signal and noise are normalized to unit noise variance, is twice the parameter γ_{n_1} given in (16). Rewriting (16) for the power constraint $(nS-5)/n_1$

$$\begin{aligned} d_{n_1} &\geq 2\sqrt{3}\left(1+\frac{nS-5}{n_1}-\frac{1}{n_1}\right)^{n_1/2}\exp(-nR) \\ &= 2\sqrt{3}\left(1+\frac{nS}{n_1}\right)^{n_1/2}\exp(-nR)\left(1-\frac{6}{nS+n_1}\right)^{n_1/2} \\ &\overset{(a)}{\geq} 2\sqrt{3}\left(1+\frac{nS}{n_1}\right)^{n_1/2}\exp(-nR)\left(1-\frac{1}{1+n_1/6}\right)^{n_1/2} \\ &\geq \frac{2\sqrt{3}}{e^3}\left(1+\frac{Sn}{n_1}\right)^{n_1/2}\exp(-nR) \end{aligned} \tag{38}$$

where to get (a) we assumed that $nS \geq 6$. We can also show that the multiplicative term, $(1-\frac{1}{1+n_1/6})^{n_1/2}$, is a decreasing function of n_1 satisfying

$$\left(1-\frac{1}{1+n_1/6}\right)^{n_1/2} \geq \lim_{n_1\to\infty}\left(1-\frac{1}{1+n_1/6}\right)^{n_1/2} = e^{-3}.$$

This establishes (38). In order to satisfy $d_{n_1} \geq 4$, it suffices for the right-hand side of (38) to be greater than or equal to 4. Letting $\nu = n_1/n$, this condition can be rewritten as

$$\exp\left[n\left(-R+\frac{\nu}{2}\ln\left(1+\frac{S}{\nu}\right)\right)\right] \geq \frac{2e^3}{\sqrt{3}}. \tag{39}$$

Define $\phi(\nu)$ by

$$\phi(\nu) = \frac{\nu}{2}\ln(1+S/\nu).$$

This is a concave increasing function for $0 < \nu \leq 1$ and can be interpreted as the capacity of the given channel if the number of available degrees of freedom is reduced from n to νn without changing the available energy per block, i.e., it can be interpreted as the capacity of a continuous-time channel whose bandwidth has been reduced by a factor of ν. We can then rewrite (39) as

$$\phi(\nu) \geq R + \frac{\beta}{n} \tag{40}$$

where $\beta = \ln(\frac{2e^3}{\sqrt{3}})$. This is interpreted in Fig. 3.

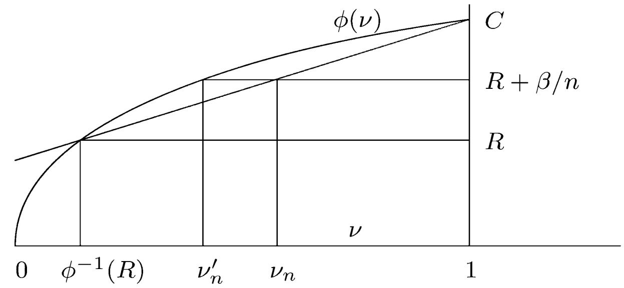

Fig. 3. The figure shows the function $\phi(\nu)$ and also the value of ν, denoted $\phi^{-1}(R)$, at which $\phi(\nu) = R$. It also shows ν_n', which satisfies $\phi(\nu_n') = R+\beta/n$, and gives the solution to (40) with equality. It turns out to be more convenient to satisfy (40) with inequality using ν_n, which by simple geometry satisfies $\nu_n = \phi^{-1}(R) + \frac{\beta(1-\phi^{-1}(R))}{n(C-R)}$.

The condition $d_{n_1} \geq 4$ is satisfied by choosing $n_1 = \lceil n\nu_n \rceil$ for ν_n defined in Fig. 3, i.e.,

$$n_1 = \left\lceil n\phi^{-1}(R) + \frac{\beta(1-\phi^{-1}(R))}{C-R}\right\rceil.$$

[10]However, unlike the scheme presented above, in Zigangirov's scheme the total amount of energy needed for transmission is increasing linearly with time.

Thus, the duration n_2 of phase 2 can be chosen to be

$$n_2 = \left\lfloor n[1-\phi^{-1}(R)] - \frac{\beta(1-\phi^{-1}(R))}{C-R} \right\rfloor. \tag{41}$$

This shows that n_2 increases linearly with n at rate $1-\phi^{-1}(R)$ for $n > \beta/(C-R)$. As a result of Lemma 4.2, the error probability is upper-bounded as

$$\Pr(\hat{m} \neq m) \leq 1/g_{n_2+1}(2). \tag{42}$$

Thus, the probability of error is bounded by an exponential order that increases at a rate $1-\phi^{-1}(R)$. We later derive a lower bound to error probability which has this same rate of increase for the exponential order of error probability.

B. The Broadband Case—Zero Error Probability

The broadband case is somewhat simpler since an unlimited number of degrees of freedom are available. For phase 1, we start with (24), modified by the fact that five units of energy must be reserved for phase 2.

$$\begin{aligned} d_{n_1} &\geq 2\sqrt{3}\exp\left[\frac{n_1}{2}\ln\left(1+\frac{\mathcal{P}T}{n_1}-\frac{6}{n_1}\right)-TR_\infty\right] \\ &\geq 2\sqrt{3}\exp\left[\frac{\mathcal{P}T}{2}-3-\frac{\mathcal{P}^2T^2}{4n_1}-TR_\infty\right] \end{aligned}$$

where, in order to get the inequality in the second step, we assumed that $\mathcal{P}T \geq 6$ and used the identity $\ln(1+x) \geq x - x^2/2$. As in the broadband SK analysis, we assume that n_1 is increasing quadratically with increasing T. Then $\frac{\mathcal{P}^2T^2}{4n_1}$ becomes just a constant. Specifically, if $n_1 \geq \frac{\mathcal{P}^2T^2}{4}$, we get

$$d_{n_1} \geq \frac{2\sqrt{3}}{e^4}\exp[T(C_\infty - R_\infty)].$$

It follows that $d_{n_1} \geq 4$ if

$$T \geq \frac{4+\ln 2 - 0.5\ln 3}{C_\infty - R_\infty}. \tag{43}$$

If (43) is satisfied, then phase 2 can be carried out for arbitrarily large n_2, with P_e satisfying (42). In principle, n_2 can be infinite, so P_e becomes 0 whenever T is large enough to satisfy (43).

One might object that the transmitter sequence is not well defined with $n_2 = \infty$, but in fact it is, since at most a finite number of transmitted symbols can be nonzero. One might also object that it is impossible to obtain an infinite number of ideal feedback signals in finite time. This objection is certainly valid, but the entire idea of ideal feedback with infinite bandwidth is unrealistic. Perhaps a more comfortable way to express this result is that 0 is the greatest lower bound to error probability when (43) is satisfied, i.e., any desired error probability, no matter how small, is achievable if the continuous-time block length T satisfies (43).

VI. A Lower Bound to Error Probability

The previous sections have derived upper bounds to the probability of decoding error for data transmission using particular block coding schemes with ideal feedback. These schemes are nonoptimal, with the nonoptimalities chosen both for analytical convenience and for algorithmic simplicity. It appears that the optimal strategy is quite complicated and probably not very interesting. For example, even with a block length $n = 1$, and a message set size $M = 4$, PAM with equispaced messages is neither optimal in the sense of minimizing average error probability over the message set (see [6, Exercise 6.3]) nor in the sense of minimizing the error probability of the worst message. Aside from this rather unimportant nonoptimality, the SK scheme is also nonoptimal in ignoring the discrete nature of the signal until the final decision. Finally, the improved algorithm of Section V is nonoptimal both in using ML rather than maximum *a posteriori* probability (MAP) for the tentative decisions and in not optimizing the choice of signal points as a function of the prior received signals.

The most important open question, in light of the extraordinarily rapid decrease of error probability with block length for the finite bandwidth case, is whether any strictly positive lower bound to error probability exists for fixed block length n. To demonstrate that there is such a positive lower bound we first derive a lower bound to error probability for the special case of a message set of size $M = 2$. Then, we generalize this to codes of arbitrary rate and show that for $R < C$, the lower bound decreases as a kth-order exponential where k increases with the block length n and has the form $k = an - b'$ where the coefficient a is the same as that in the upper bound in Section V. It is more convenient in this section to number the successive signals from 1 to n rather than 0 to $n-1$ as in previous sections.

A. A Lower Bound for $M = 2$

Although it is difficult to find and evaluate the entire optimal code, even for $M = 2$, it turns out to be easy to find the optimal encoding in the last step. Thus, for each $\boldsymbol{Y}_1^{n-1}$, we want to find the optimal choice of $X_n = f(U, \boldsymbol{Y}_1^{n-1})$ as a function of, first, the encoding functions $X_i = f(U, \boldsymbol{Y}_1^{i-1})$, $1 \leq i \leq n-1$, and, second, the allocation of energy, $\tilde{S} = \mathsf{E}[X_n^2|\boldsymbol{Y}_1^{n-1}]$ for that $\boldsymbol{Y}_1^{n-1}$. We will evaluate the error probability for such an optimal encoding at time n and then relate it to the error probability that would have resulted from decoding at time $n-1$. We will use this relation to develop a recursive lower bound to error probability at each time i in terms of that at time $i-1$.

For a given code function $X_i = f(U, \boldsymbol{Y}_1^{i-1})$ for $1 \leq i \leq n-1$, the conditional probability density[11] of $\boldsymbol{Y}_1^i$ given $U = 1$ or 2 is positive for all sample values for $\boldsymbol{Y}_1^i$; thus, the corresponding conditional probabilities of hypotheses $U = 1$ and $U = 2$ are positive, i.e.,

$$\Pr\left(U{=}m|\boldsymbol{Y}_1^i\right) > 0, \quad m \in \{1,2\}, \forall \boldsymbol{Y}_1^i \in \mathbb{R}^i.$$

In particular, for $m \in \{1,2\}$, define $\Phi_m = \Pr(U{=}m|\boldsymbol{Y}_1^{n-1})$ for some given $\boldsymbol{Y}_1^{n-1}$. Finding the error probability $\Psi = \Pr(\hat{U}(\boldsymbol{Y}_1^n) \neq U|\boldsymbol{Y}_1^{n-1})$ is an elementary binary detection problem for the given $\boldsymbol{Y}_1^{n-1}$. MAP detection, using the *a priori* probabilities Φ_1 and Φ_2, minimizes the resulting error probability.

[11] We do not use the value of this density, but for completeness, it can be seen to be $\prod_{j=1}^{i}\xi[Y_j - f(U, \boldsymbol{Y}_1^{j-1})]$, where $\xi(x)$ is the normal density $(2\pi)^{-1/2}\exp(-x^2/2)$.

For a given sample value of $\boldsymbol{Y}_1^{n-1}$, let b_1 and b_2 be the values of X_n for $U=1$ and 2, respectively. Let a be half the distance between b_1 and b_2, i.e., $2a = b_2 - b_1$. The error probability Ψ depends on b_1 and b_2 only through a. For a given $\tilde{S}$, we choose b_1 and b_2 to satisfy $\mathsf{E}[X_n|\boldsymbol{Y}_1^{n-1}] = 0$, thus maximizing a for the given $\tilde{S}$. The variance of X_n conditional on $\boldsymbol{Y}_1^{n-1}$ is given by

$$\mathrm{Var}\left(X_n|\boldsymbol{Y}_1^{n-1}\right) = \frac{1}{2}\sum_{i,j}\Phi_i\Phi_j(b_i-b_j)^2 = 4\Phi_1\Phi_2 a^2,$$

and since $\mathsf{E}[X_n|\boldsymbol{Y}_1^{n-1}] = 0$, this means that a is related to $\tilde{S}$ by $\tilde{S} = 4\Phi_1\Phi_2 a^2$.

Now let $\Phi = \min\{\Phi_1,\Phi_2\}$. Note that Φ is the probability of error for a hypothetical MAP decoder detecting U at time $n-1$ from $\boldsymbol{Y}_1^{n-1}$. The error probability Ψ for the MAP decoder at the end of time n is given by the classic result of binary MAP detection with *a priori* probabilities Φ and $1-\Phi$

$$\Psi = (1-\Phi)Q\left(a + \frac{\ln\eta}{2a}\right) + \Phi Q\left(a - \frac{\ln\eta}{2a}\right) \tag{44}$$

where $\eta = \frac{1-\Phi}{\Phi}$ and $Q(x) = \int_x^\infty (2\pi)^{-1/2}\exp(-z^2/2)\,dz$. This equation relates the error probability Ψ at the end of time n to the error probability Φ at the end of time $n-1$, both conditional on $\boldsymbol{Y}_1^{n-1}$. We are now going to view Ψ and Φ as functions of $\boldsymbol{Y}_1^{n-1}$, and thus as random variables. Similarly, $\tilde{S}\geq 0$ can be any nonnegative function of $\boldsymbol{Y}_1^{n-1}$, subject to a constraint S_n on its mean; so we can view $\tilde{S}$ as an arbitrary nonnegative random variable with mean S_n. For each $\boldsymbol{Y}_1^{n-1}$, $\tilde{S}$ and Φ determine the value of a; thus, a is also a nonnegative random variable.

We are now going to lower-bound the expected value of Ψ in such a way that the result is a function only of the expected value of Φ and the expected value S_n of $\tilde{S}$. Note that Ψ in (44) can be lower-bounded by ignoring the first term and replacing the second term with $\Phi Q(a)$. Thus

$$\begin{aligned}\Psi &\geq \Phi Q(a)\\ &= \Phi Q\left(\sqrt{\frac{\tilde{S}}{4\Phi(1-\Phi)}}\right)\\ &\geq \Phi Q\left(\sqrt{\frac{\tilde{S}}{2\Phi}}\right)\end{aligned} \tag{45}$$

where the last step uses the facts that $Q(x)$ is a decreasing function of x and that $1-\Phi > 1/2$.

$$\mathsf{E}[\Psi] \geq \mathsf{E}[\Phi]Q\left(\frac{1}{\mathsf{E}[\Phi]}\mathsf{E}\left[\Phi\sqrt{\frac{\tilde{S}}{2\Phi}}\right]\right) \tag{46}$$

$$= \mathsf{E}[\Phi]Q\left(\frac{1}{\sqrt{2}\mathsf{E}[\Phi]}\mathsf{E}\left[\sqrt{\Phi\tilde{S}}\right]\right)$$

$$\geq \mathsf{E}[\Phi]Q\left(\frac{1}{\sqrt{2}\mathsf{E}[\Phi]}\sqrt{\mathsf{E}[\Phi]\mathsf{E}[\tilde{S}]}\right) \tag{47}$$

$$= \mathsf{E}[\Phi]Q\left(\sqrt{\frac{S_n}{2\mathsf{E}[\Phi]}}\right). \tag{48}$$

In (46), we used Jensen's inequality, based on the facts that $Q(x)$ is a convex function for $x\geq 0$ and that $\Phi/\mathsf{E}[\Phi]$ is a probability distribution on $\boldsymbol{Y}_1^{n-1}$. In (47), we used the Schwarz inequality along with the fact that $Q(x)$ is decreasing for $x\geq 0$.

We now recognize that $\mathsf{E}[\Psi]$ is simply the overall error probability at the end of time n and $\mathsf{E}[\Phi]$ is the overall error probability (if a MAP decision were made) at the end of time $n-1$. Thus, we denote these quantities as p_n and p_{n-1} respectively

$$p_n \geq p_{n-1}Q\left(\sqrt{\frac{S_n}{2p_{n-1}}}\right). \tag{49}$$

Note that this lower bound is monotone increasing in p_{n-1}. Thus, we can further lower-bound p_n by lower-bounding p_{n-1}. We can lower-bound p_{n-1} (for a given p_{n-2} and S_{n-1}) in exactly the same way, so that $p_{n-1}\geq p_{n-2}Q(\sqrt{S_{n-1}/2p_{n-2}})$. These two bounds can be combined to implicitly bound p_n in terms of p_{n-2}, S_n, and S_{n-1}. In fact, the same technique can be used for each $i, 1\leq i\leq n$, getting

$$p_i \geq p_{i-1}Q\left(\sqrt{\frac{S_i}{2p_{i-1}}}\right). \tag{50}$$

This gives us a recursive lower bound on p_n for any given choice of $S_1,\ldots,S_n$ subject to the power constraint $\sum_i S_i \leq nS$.

We have been unable to find a clean way to optimize this over the choice of $S_1,\ldots,S_n$, so as a very crude lower bound on p_n, we upper-bound each S_i by nS. For convenience, multiply each side of (50) by $2/nS$

$$\frac{2p_i}{nS} \geq \frac{2p_{i-1}}{nS}Q\left(\sqrt{\frac{nS}{2p_{i-1}}}\right), \quad \text{for } 1\leq i\leq n. \tag{51}$$

At this point, we can see what is happening in this lower bound. As p_i approaches 0, $\frac{nS}{2p_i}\to\infty$. Also, $Q\left(\sqrt{\frac{nS}{2p_i}}\right)$ approaches 0 as $e^{-\frac{nS}{4p_i}}$. Now we will lower-bound the expression on the right-hand side of (51). We can check numerically[12] that for $x\geq 9$

$$\frac{1}{x}Q(\sqrt{x}) \geq \exp(-x). \tag{52}$$

Furthermore, $\frac{1}{x}Q(\sqrt{x})$ is decreasing in x for all $x>0$, and thus

$$\frac{1}{x}Q(\sqrt{x}) \geq \exp(-\max\{x,9\}), \quad \forall x>0.$$

Substituting this into (51) we get

$$\frac{2p_i}{nS} \geq \frac{1}{\exp\left(\max\left\{\frac{nS}{2p_{i-1}},9\right\}\right)}, \quad \text{for } 1\leq i\leq n.$$

Applying this recursively for $i=n$ down to $i=k+1$ for any $k\geq 0$ we get

$$\begin{aligned}\frac{2p_n}{nS} &\geq \frac{1}{\exp\left(\max\left\{\exp\left(\max\left\{\frac{nS}{2p_{n-2}},9\right\}\right),9\right\}\right)}\\ &\overset{(a)}{=} \frac{1}{\exp\left(\exp\left(\max\left\{\frac{nS}{2p_{n-2}},9\right\}\right)\right)}\\ &\geq \frac{1}{g_{n-k}\left[\max\left\{\frac{nS}{2p_k},9\right\}\right]}\end{aligned} \tag{53}$$

[12]That is, we can check numerically that (52) is satisfied for $x=9$ and verify that the right-hand side is decreasing faster than the left for $x>9$.

where (a) simply follows from the fact that $\exp(9) > 9$. This bound holds for $k = 0$, giving an overall lower bound on error probability in terms of p_0. In the usual case where the symbols are initially equiprobable, $p_0 = 1/2$ and

$$p_n \geq \frac{nS}{2g_n[\max(nS, 9)]}. \tag{54}$$

Note that this lower bound is an nth-order exponential. Although it is numerically much smaller than the upper bound in Section V, it has the same general form. The intuitive interpretation is also similar. In going from block length $n-1$ to n, with very small error probability at $n-1$, the symbol of large *a priori* probability is very close to 0 and the other symbol is approximately at $\sqrt{\tilde{S}/p_{n-1}}$. Thus, the error probability is decreased in one time unit by an exponential in p_{n-1}, leading to an nth-order exponential over n time units.

B. A Lower Bound for Arbitrary M

Next consider feedback codes of arbitrary rate $R < C$ with sufficiently large block length n and $M = e^{nR}$ codewords. We derive a lower bound on error probability by splitting n into an initial segment of length n_1 and a final segment of length $n_2 = n - n_1$. This segmentation is for bounding purposes only and does not restrict the feedback code. The error probability of a hypothetical MAP decoder at the end of the first segment, $P_e(n_1)$, can be lower-bounded by a conventional use of the Fano inequality. We will show how to use this error probability as the input of the lower bound for $M = 2$ case derived in the previous subsection, i.e., (53). There is still the question of allocating power between the two segments, and since we are deriving a lower bound, we simply assume that the entire available energy is available in the first segment, and can be reused in the second segment. We will find that the resulting lower bound has the same form as the upper bound in Section V.

Using energy Sn over the first segment corresponds to power Sn/n_1, and since feedback does not increase the channel capacity, the average directed mutual information over the first segment is at most $n_1C(Sn/n_1)$. Reusing the definitions $\nu = n_1/n$ and $\phi(\nu) = \frac{\nu}{2}\ln(1 + \frac{S}{\nu})$ from Section V,

$$n_1C(Sn/n_1) = n\phi(\nu).$$

The entropy of the source is $\ln M = nR$, and thus the conditional entropy of the source given $\boldsymbol{Y}_1^{n_1}$ satisfies

$$\begin{aligned} n[R - \phi(\nu)] &\leq H\left(U|\boldsymbol{Y}_1^{n_1}\right) \\ &\leq h(P_e(n_1)) + P_e(n_1)nR \\ &\leq \ln 2 + P_e(n_1)nR \end{aligned} \tag{55}$$

where we have used the Fano inequality and then bounded the binary entropy $h(p) = -p\ln p - (1-p)\ln(1-p)$ by $\ln 2$.

To use (55) as a lower bound on $P_e(n_1)$, it is necessary for $n_1 = n\nu$ to be small enough that $\phi(\nu)$ is substantially less than R, and to be specific we choose ν to satisfy

$$R - \phi(\nu) \geq \frac{1}{n}. \tag{56}$$

With this restriction, it can be seen from (55) that

$$P_e(n_1) \geq \frac{1 - \ln 2}{nR}. \tag{57}$$

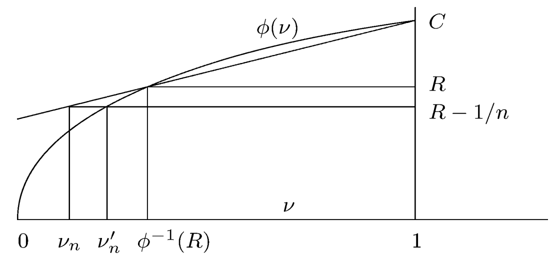

Fig. 4. The figure shows the value of ν, denoted $\phi^{-1}(R)$, at which $\phi(\nu) = R$. It also shows ν'_n, where $\phi(\nu'_n) = R - 1/n$. This gives the solution to (56) with equality, but $\nu_n = \phi^{-1}(R) - \frac{1-\phi^{-1}(R)}{n(C-R)}$ can be seen to be less than ν'_n and thus also satisfies (56).

Fig. 4 illustrates that the following choice of n_1 in (58) satisfies both (56) and (57). This uses the fact that $\phi(\nu)$ is a monotonically increasing concave function of ν

$$n_1 = \left\lfloor n\phi^{-1}(R) - \frac{1 - \phi^{-1}(R)}{C - R} \right\rfloor. \tag{58}$$

The corresponding choice for n_2 is

$$n_2 = \left\lceil n[1 - \phi^{-1}(R)] + \frac{1 - \phi^{-1}(R)}{C - R} \right\rceil. \tag{59}$$

Thus, with this choice of n_1, n_2, the error probability at the end of time n_1 satisfies (57).

The straightforward approach at this point would be to generalize the recursive relationship in (50) to arbitrary M. This recursive relationship could then be used, starting at time $i = n$ and using each successively smaller i until terminating the recursion at $i = n_1$ where (57) can be used. It is simpler, however, since we have already derived (50) for $M = 2$, to define a binary coding scheme from any given M-ary scheme in such a way that the binary results can be used to lower-bound the M-ary results. This technique is similar to one used earlier in [1].

Let $X_i = f(U, \boldsymbol{Y}_1^{i-1})$ for $1 \leq i \leq n$ be any given coding function for $U \in \mathcal{M} = \{1, \ldots, M\}$. That code is used to define a related binary code. In particular, for each received sequence $\boldsymbol{Y}_1^{n_1}$ over the first segment, we partition the message set $\mathcal{M}$ into two subsets, $\mathcal{M}_1(\boldsymbol{Y}_1^{n_1})$ and $\mathcal{M}_2(\boldsymbol{Y}_1^{n_1})$. The particular partition for each $\boldsymbol{Y}_1^{n_1}$ is defined later. This partitioning defines a binary random variable V as follows:

$$V = \begin{cases} 1, & U \in \mathcal{M}_1\left(\boldsymbol{Y}_1^{n_1}\right) \\ 2, & U \in \mathcal{M}_2\left(\boldsymbol{Y}_1^{n_1}\right). \end{cases}$$

At the end of the transmission, the receiver will use its decoder to decide $\hat{U}$. We define the decoder for V at time n, using the decoder of U as follows:

$$\hat{V} = \begin{cases} 1, & \hat{U} \in \mathcal{M}_1\left(\boldsymbol{Y}_1^{n_1}\right) \\ 2, & \hat{U} \in \mathcal{M}_2\left(\boldsymbol{Y}_1^{n_1}\right). \end{cases}$$

Note that with the above mentioned definitions, whenever the M-ary scheme decodes correctly, the related binary scheme does also, and thus the error probability $P_e(n)$ for the M-ary scheme must be greater than or equal to the error probability p_n of the related binary scheme.

The binary scheme, however, is one way (perhaps somewhat bizarre) of transmitting a binary symbol, and thus it satisfies the results[13] of Section VI-A. In particular, for the binary scheme, the error probability p_n at time n is lower-bounded by the error probability p_{n_1} at time n_1 by (53)

$$P_e(n) \geq p_n \geq \frac{nS}{2} \frac{1}{g_{n_2}\left[\max\left\{\frac{nS}{2p_{n_1}}, 9\right\}\right]}. \tag{60}$$

Our final task is to relate the error probability p_{n_1} at time n_1 for the binary scheme to the error probability $P_e(n_1)$ in (57) for the M-ary scheme. In order to do this, let $\Phi_m(\boldsymbol{Y}_1^{n_1})$ be the probability of message m conditional on the received first segment $\boldsymbol{Y}_1^{n_1}$. The MAP error probability for an M-ary decision at time n_1, conditional on $\boldsymbol{Y}_1^{n_1}$, is $1-\Phi_{\max}(\boldsymbol{Y}_1^{n_1})$ where

$$\Phi_{\max}(\boldsymbol{Y}_1^{n_1}) = \max\{\Phi_1(\boldsymbol{Y}_1^{n_1}), \ldots \Phi_M(\boldsymbol{Y}_1^{n_1})\}.$$

Thus, $P_e(n_1)$, given in (57), is the mean of $1-\Phi_{\max}(\boldsymbol{Y}_1^{n_1})$ over $\boldsymbol{Y}_1^{n_1}$.

Now p_{n_1} is the mean, over $\boldsymbol{Y}_1^{n_1}$, of the error probability of a hypothetical MAP decoder for V at time n_1 conditional on $\boldsymbol{Y}_1^{n_1}$, $p_{n_1}(\boldsymbol{Y}_1^{n_1})$. This is the smaller of the *a posteriori* probabilities of the subsets $\mathcal{M}_1$, $\mathcal{M}_2$ conditional on $\boldsymbol{Y}_1^{n_1}$, i.e.,

$$p_{n_1}\left(\boldsymbol{Y}_1^{n_1}\right) = \min\left\{\sum_{m\in\mathcal{M}_1(\boldsymbol{Y}_1^{n_1})} \Phi_m\left(\boldsymbol{Y}_1^{n_1}\right), \sum_{m\in\mathcal{M}_2(\boldsymbol{Y}_1^{n_1})} \Phi_m\left(\boldsymbol{Y}_1^{n_1}\right)\right\} \tag{61}$$

The following lemma shows that by an appropriate choice of partition for each $\boldsymbol{Y}_1^{n_1}$, this binary error probability is lower-bounded by $1/2$ the corresponding M-ary error probability.

Lemma 6.1: For any probability distribution $\Phi_1, \ldots, \Phi_M$ on a message set $\mathcal{M}$ with $M > 2$, let $\Phi_{\max} = \max\{\Phi_1, \ldots, \Phi_M\}$. Then there is a partition of $\mathcal{M}$ into two subsets, $\mathcal{M}_1$ and $\mathcal{M}_2$, such that

$$\sum_{m\in\mathcal{M}_1} \Phi_m \geq \frac{1-\Phi_{\max}}{2} \quad \text{and} \quad \sum_{m\in\mathcal{M}_2} \Phi_m \geq \frac{1-\Phi_{\max}}{2}. \tag{62}$$

Proof: Order the messages in order of decreasing Φ_m. Assign the messages one by one in this order to the sets $\mathcal{M}_1$ and $\mathcal{M}_2$. When assigning the kth most likely message, we calculate the total probability of the messages that have already been assigned to each set, and assign the kth message to the set which has the smaller probability mass. If the probability mass of the sets are the same, we choose one of the sets arbitrarily. With such a procedure, the difference in the probabilities of the sets, as they evolve, never exceeds $\Phi_{\max}$. After all messages have been assigned, let

$$\Phi_1' = \sum_{m\in\mathcal{M}_1} \Phi_m; \quad \Phi_2' = \sum_{m\in\mathcal{M}_2} \Phi_m.$$

We have seen that $|\Phi_1' - \Phi_2'| \leq \Phi_{\max}$. Since $\Phi_1' + \Phi_2' = 1$, (62) follows. □

Since the error probability for the binary scheme is now at least one half of that for the M-ary scheme for each $\boldsymbol{Y}_1^{n_1}$, we can take the mean over $\boldsymbol{Y}_1^{n_1}$, getting $p_{n_1} \geq P_e(n_1)/2$. Combining this with (60) and (57)

$$P_e(n) \geq \frac{nS}{2} \frac{1}{g_{n_2}\left[\max\left(\frac{n^2SR}{1-\ln 2}, 9\right)\right]} \tag{63}$$

where n_2 is given in (59). The exact terms in this expression are not particularly interesting because of the very weak bounds on energy at each channel use. What is interesting is that the order of exponent in both the upper bound of (42) and (41) and the lower bound here are increasing linearly[14] at the same rate $1-\phi^{-1}(R)$.

VII. Conclusion

The SK data transmission scheme can be viewed as ordinary PAM combined with the Elias scheme for noise reduction. The SK scheme can also be improved by incorporating the PAM structure into the transmission of the error in the receiver's estimate of the message, particularly during the latter stages. For the band-limited version, this leads to an error probability that decreases with an exponential order $an+b$ where $a = 1-\phi^{-1}(R)$ and b is a constant. In the broadband version, the error probability is zero for sufficiently large finite constraint durations T. A lower bound to error probability, valid for all $R < C$ was derived. This lower bound also decreases with an exponential order $an + b'(n)$ where again $a = 1-\phi^{-1}(R)$ and $b'(n)$ is essentially a constant.[15] It is interesting to observe that the strategy yielding the upper bound uses almost all the available energy in the first phase, using at most five units of energy in the second phase. The lower bound relaxed the energy constraint, allowing all the allowable energy to be used in the first phase and then to be used repeatedly in each time unit of the second phase. The fact that both bounds decrease with the same exponential order suggests that the energy available for the second phase is not of primary importance. An open theoretical question is the minimum overall energy under which the error probability for two codewords can be zero in the infinite bandwidth case.

References

[1] P. Berlin, B. Nakiboğlu, B. Rimoldi, and E. Telatar, "A simple converse of Burnashev's reliability function," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3074–3080, Jul. 2009.

[2] M. V. Burnashev, "Sequential discrimination of hypotheses with control of observations," *Math. USSR-Izv.*, vol. 15, no. 3, pp. 419–440, 1980.

[13] This is not quite as obvious as it sounds. The binary scheme here is not characterized by a coding function $f(V, \boldsymbol{Y}_1^{i-1})$ as in Section VI-A, but rather is a randomized binary scheme. That is, for a given $\boldsymbol{Y}_1^{n_1}$ and a given choice of V, the subsequent transmitted symbols X_i are functions not only of V and $\boldsymbol{Y}_1^{i-1}$, but also of a random choice of U conditional on V. The basic conclusion of (50) is then justified by averaging over both $\boldsymbol{Y}_1^{i-1}$ and the choice of U conditional on V.

[14] Note that the argument of g_{n_2} is proportional to n^2, so that this bound does not quite decrease with the exponential order n_2. It does, however, decrease with an exponential order $n_2 + \alpha(n)$, where $\alpha(n)$ increases with n much more slowly than, say, $\ln(\ln(n))$. Thus, $(n_2 + \alpha(n))/n$ is asymptotically proportional to $1-\phi^{-1}(R)$.

[15] $b'(n)$ is a sublinear function of n, i.e., $\lim_{n\to\infty} \frac{b'(n)}{n} = 0$.

[3] M. V. Burnashev, "Data transmission over a discrete channel with feedback and random transmission time," *Probl. Pered. Inform.*, vol. 12, no. 4, pp. 10–30, 1976.
[4] P. Elias, Channel Capacity Without Coding MIT ResearchLaboratory of Electronics, Cambaridge, MA, Quart. Progr. Rep., Oct. 15, 1956, Also published in *Lectures in Communication System Theory*, E. Baghdady, Ed., New York: McGraw-Hill, 1961..
[5] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[6] R. G. Gallager, *Principles of Digital Communication*. New York: Cambridge Univ. Press, 2008.
[7] Y.-H. Kim, A. Lapidoth, and T. Weissman, "The Gaussian channel with noisy feedback," in *Proc. IEEE Int. Symp. Information Theory (ISIT 2007)*, Nice, France, Jun. 2007, pp. 1416–1420.
[8] A. Kramer, "Improving communication reliability by use of an intermittent feedback channel," *IEEE Trans. Inf. Theory*, vol. IT-15, no. 1, pp. 52–60, Jan. 1969.
[9] B. Nakiboğlu and R. G. Gallager, "Error exponents for variable-length block codes with feedback and cost constraints," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 945–963, Mar. 2008.
[10] M. S. Pinsker, "The probability of error in block transmission in a memoryless Gaussian channel with feedback," *Probl. Pered. Inform.*, vol. 4, no. 4, pp. 1–14, 1968.
[11] A. Sahai, "Why do block length and delay behave differently if feedback is present?," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1860–1886, May 2008.
[12] J. Schalkwijk, "A coding scheme for additive noise channels with feedback—II: Band-limited signals," *IEEE Trans. Inf. Theory*, vol. IT-12, no. 2, pp. 183–189, Apr. 1966.
[13] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback—I: No bandwidth constraint," *IEEE Trans. Inf. Theory*, vol. IT-12, no. 2, pp. 172–182, Apr. 1966.
[14] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. Mathematical Statistics and Probability*, Berleley, CA, 1961, vol. 1, pp. 611–644, University of California Press.
[15] K. Sh. Zigangirov, "Upper bounds for the error probability for channels with feedback," *Probl. Pered. Inform.*, vol. 6, no. 2, pp. 87–92, 1970.
[16] L. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 4, pp. 623–629, Jul. 1984.
[17] L. Ozarow and S. Leung-Yan-Cheong, "An achievable region and outer bound for the Gaussian broadcast channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 4, pp. 667–671, Jul. 1984.
[18] G. Kramer, "Feedback strategies for white Gaussian interference networks," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1423–1438, Jun. 2002.
[19] S. I. Bross and M. A. Wigger, "On the relay channel with receiver transmitter feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 275–291, Jan. 2009.
[20] A. Sahai, S. C. Draper, and M. Gastpar, "Boosting reliability over awgn networks with average power constraints and noiseless feedback," in *Proc. IEEE Int. Symp. Information Theory (ISIT'2005)*, Adelaide, Australia, Sep. 2005, pp. 402–406.
[21] D. Gündüz, D. R. Brown, and H. V. Poor, "Secret communication with feedback," in *Proc. IEEE Symp. Information Theory and Its Applications (ISITA 2008)*, Auckland, New Zeland, Dec. 2008, pp. 1–6.
[22] Y.-H. Kim, "Feedback capacity of the first-order moving average Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3063–3079, Jul. 2006.
[23] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

Robert G. Gallager (S'58–M'61–F'68–LF'96) received the B.S.E.E. degree from the University of Pennsylvania, Philadelphia, in 1953 and the S.M. and Sc.D. degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 1957 and 1960, respectively.

From 1953 to 1956, he was at Bell Telephone Laboratories and then the U.S. Signal Corps. He joined the MIT faculty in 1960, became Fujitsu Professor in 1988, and is now Professor Emeritus. In the mid 1970s, his research focused on data networks and he coauthored the text, *Data Networks* (2nd ed., Englewood Cliffs, NJ: Prentice-Hall 1992) with D. Bertsekas. He wrote *Discrete Stochastic Processes* (Kluwer, 1996) and has just completed the textbook *Principles of Data Communication* (New York: Cambridge Univ. Press, 2008). His current interests are in information theory, wireless communication, all optical networks, data networks, and stochastic processes. He is proud of his many graduate students. He has consulted for a number of companies and has received five patents.

Prof. Gallager was President of the IEEE Information Theory Society in 1971, Chairman of the Advisory committee to the NSF Division on Networking and Communication Research and Infrastructure from 1989 to 1992, and has been on numerous academic visiting committees. His honors include University of Pennsylvania Moore School Gold Medal Award (1973), Guggenheim Fellowship (1978), election to the National Academy of Engineering (1979), IEEE Information Theory Society Shannon Award (1983), IEEE Medal of Honor (1990), election to the National Academy of Sciences (1992), Fellow of the American Academy of Arts and Sciences (1999), the Technion Harvey Prize (1999), the Eduard Rhein Prize for basic research (2002), the Marconi Fellowship Award (2003), and the Dijkstra Prize in Distributed Computing (2004). He won the MIT Graduate Student Council Teaching Award for 1993. His 1960 Sc.D. dissertation on "Low Density Parity Check Codes" won an IEEE Information Theory Society Golden Jubilee Paper Award in 1998 and is an active area of research today. "A Simple Derivation of the Coding Theorem and some Applications," won the IEEE Baker Prize (1966) and an IEEE Information Theory Society Golden Jubilee Paper Award (1998). His book, *Information Theory and Reliable Communication* (New York: Wiley 1968), placed Information Theory on a firm foundation. His joint papers in 1993 with Parekh, "A Generalized Processor Sharing Approach to Flow Control in ISN," won the 1993 William Bennett Prize Paper Award and the 1993 Infocomm Prize Paper Award.

Barış Nakiboğlu received the B.S. degrees in electrical and electronics engineering and physics from Middle East Technical University (METU), Ankara, Turkey, in 2002 and the M.S. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2005.

He is currently working toward the Ph.D. degree at the Research Laboratory of Electronics (RLE) at MIT.