Gaussian Channel With Noisy Feedback and Peak Energy Constraint

Yu Xiang, *Student Member, IEEE*, and Young-Han Kim, *Senior Member, IEEE*

***Abstract*—Optimal coding over the additive white Gaussian noise channel under the *peak* energy constraint is studied when there is noisy feedback over an orthogonal additive white Gaussian noise channel. As shown by Pinsker, under the peak energy constraint, the best error exponent for communicating an M-ary message, $M \geq 3$, with noise-free feedback is strictly larger than the one without feedback. This paper extends Pinsker's result and shows that if the noise power in the feedback link is sufficiently small, the best error exponent for communicating an M-ary message can be strictly larger than the one without feedback. The proof involves two feedback coding schemes. One is motivated by a two-stage noisy feedback coding scheme of Burnashev and Yamamoto for binary symmetric channels, while the other is a linear noisy feedback coding scheme that extends Pinsker's noise-free feedback coding scheme. When the feedback noise power α is sufficiently small, the linear coding scheme outperforms the two-stage (nonlinear) coding scheme, and is asymptotically optimal as α tends to zero. By contrast, when α is relatively larger, the two-stage coding scheme performs better.**

***Index Terms*—Error exponent, Gaussian channel, noisy feedback, peak energy constraint.**

I. Introduction and Main Results

We consider a communication problem for an additive white Gaussian noise (AWGN) *forward* channel with feedback over an orthogonal AWGN *backward* channel as depicted in Fig. 1. Suppose that the sender wishes to communicate a message $W \in [1:M] := \{1, 2, \ldots, M\}$ over the (forward) AWGN channel

$$Y_i = X_i + Z_i$$

where X_i, Y_i, and Z_i, respectively, denote the channel input, channel output, and additive Gaussian noise. The sender has a causal access to a noisy version $\tilde{Y}_i$ of Y_i over the feedback (backward) AWGN channel

$$\tilde{Y}_i = Y_i + \tilde{Z}_i$$

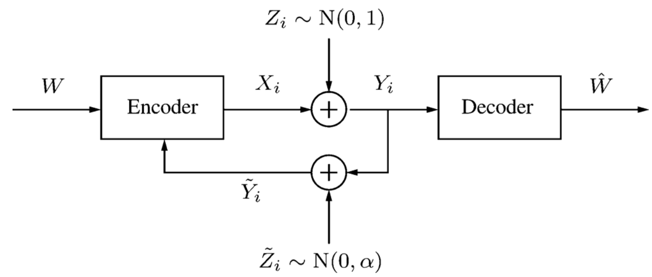


Fig. 1. Gaussian channel with noisy feedback.

where $\tilde{Z}_i$ is the Gaussian noise in the backward link. We assume that the forward noise process $\{Z_i\}_{i=1}^{\infty}$ and the backward noise process $\{\tilde{Z}_i\}_{i=1}^{\infty}$ are independent of each other, and white Gaussian $\mathrm{N}(0,1)$ and $\mathrm{N}(0,\alpha)$, respectively.

We define an (M,n) code with the encoding functions $x_i(w, \tilde{y}^{i-1})$, $i \in [1:n]$, and the decoding function $\hat{w}(y^n)$. We assume a *peak energy constraint*

$$\mathsf{P}\left\{\sum_{i=1}^{n} x_i^2(w, \tilde{Y}^{i-1}) \leq nP\right\} = 1 \quad \text{for all } w. \tag{1}$$

The probability of error of the code is defined as

$$\begin{aligned} P_e^{(n)} &= \mathsf{P}\{W \neq \hat{W}(Y^n)\} \\ &= \frac{1}{M}\sum_{w=1}^{M} \mathsf{P}\{W \neq \hat{W}(Y^n) | W = w\}, \end{aligned}$$

where W is distributed uniformly over $\{1, 2, \ldots, M\}$ and is independent of $(Z^n, \tilde{Z}^n)$.

As is well known, the capacity of the channel (the supremum of $(\log M)/n$ such that there exists a sequence of (M,n) codes with $\lim_{n\to\infty} P_e^{(n)} \to 0$) stays the same with or without feedback. Hence, our main focus is the reliability of communication, which is captured by the error exponent

$$\lim_{n\to\infty} -\frac{1}{n} \ln P_e^{(n)}$$

of the given code. The error exponent is sensitive to the presence of noise in the feedback link. Schalkwijk and Kailath showed in their celebrated work [1] that *noise-free* feedback can improve the error exponent dramatically under the *expected energy constraint*

$$\sum_{i=1}^{n} \mathsf{E}[x_i^2(w, \tilde{Y}^{i-1})] \leq nP \quad \text{for all } w \tag{2}$$

(in fact, $P_e^{(n)}$ decays much faster than exponentially in n). Kim *et al.* [2] studied the optimal error exponent under the expected

Manuscript received August 17, 2011; revised January 17, 2013; accepted March 09, 2013. Date of publication April 25, 2013; date of current version July 10, 2013. This work was supported in part by the National Science Foundation under Grant CCF-0729195. This paper was presented in part at the 2010 IEEE International Symposium on Information Theory.

The authors are with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093-0407 USA (e-mail: yxiang@ucsd.edu; yhk@ucsd.edu).

Communicated by E. Erkip, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2013.2254233

energy constraint and noisy feedback, and showed that the error exponent is inversely proportional to α for small α. Another important factor that affects the error exponent is the energy constraint on the channel inputs—the peak energy constraint in (1) versus the expected energy constraint in (2). Wyner [3] showed that the error probability of the Schalkwijk–Kailath coding scheme [1] degrades to an exponential form under the peak energy constraint. In fact, Shepp *et al.* [4] showed that for the binary-message case ($M = 2$), the best error exponent under the peak energy constraint is achieved by simple nonfeedback antipodal signaling, regardless of the presence of feedback. This negative result might lead to an impression that under the peak energy constraint, even noise-free feedback does not improve the reliability of communication. Pinsker [5] proved the contrary by showing that the best error exponent for sending an M-ary message does not depend on M and, hence can be strictly larger than the best error exponent without feedback for $M \geq 3$.

In this paper, we show that noisy feedback can improve the reliability of communication under the peak energy constraint, provided that the feedback noise power α is sufficiently small. Let

$$E_M(\alpha) := \limsup_{n\to\infty} -\frac{1}{n}\ln P_e^*(M,n)$$

where $P_e^*(M,n)$ denotes the best error probability over all (M,n) codes for the AWGN channel with the noisy feedback. Thus, $E_M(\infty)$ denotes the best error exponent for communicating an M-ary message over the AWGN channel *without* feedback. Shannon [6] showed that

$$E_M(\infty) = \frac{M}{4(M-1)}P. \tag{3}$$

This follows by first upper bounding the error exponent with the sphere packing bound and then achieving this upper bound by using a regular simplex code on the sphere of radius $\sqrt{nP}$, i.e., each codeword $x^n(w)$ satisfies $\sum_{i=1}^n x_i^2(w) = nP$ and is at the same Euclidean distance from every other codeword. In particular, for $M = 3$,

$$\begin{aligned} x^n(1) &= \sqrt{nP}\cdot(\quad 0, \qquad 1, 0, \ldots, 0)\\ x^n(2) &= \sqrt{nP}\cdot(-1/2, -\sqrt{3}/2, 0, \ldots, 0)\\ x^n(3) &= \sqrt{nP}\cdot(\ 1/2, -\sqrt{3}/2, 0, \ldots, 0), \end{aligned}$$

and

$$E_3(\infty) = \frac{3}{8}P.$$

At the other extreme, $E_M(0)$ denotes the best error exponent for communicating an M-ary message over the AWGN channel with *noise-free* feedback. Pinsker [5] showed that

$$E_M(0) \equiv \frac{P}{2}$$

for all M. In particular,

$$E_3(0) = \frac{P}{2}.$$

Clearly, $E_M(\alpha)$ is decreasing in α and

$$E_M(\infty) \leq E_M(\alpha) \leq E_M(0)$$

for every α and M.

Is $E_M(\alpha)$ strictly larger than $E_M(\infty)$ (i.e., is noisy feedback better than no feedback)? Does $E_M(\alpha)$ tend to $E_M(0)$ as $\alpha \to 0$ (i.e., does the performance degrade gracefully with small noise in the feedback link)? What is the optimal feedback coding scheme that achieves $E_M(\alpha)$? To answer these questions, we establish the following results.

Theorem 1: For $0 \leq s \leq 1$,

$$E_M(\alpha^*(s)) \geq \frac{P}{2}\left(1 - \frac{3(M-2)}{M(s^2-2s+4)+3(M-2)}\right)$$

where

$$\alpha^*(s) = \frac{3s^2}{4(s^2-2s+4)}.$$

By comparing the lower bound with (3) and identifying the critical point $\alpha = \alpha^*(1) = 1/4$, we obtain the following.

Corollary 1:

$$E_M(\alpha) > E_M(\infty) \quad \text{for } \alpha < \frac{1}{4}.$$

Thus, if the noise power in the feedback link is sufficiently small, then the noisy feedback improves the reliability of communication even under the peak energy constraint. The proof of Theorem 1 is motivated by recent results of Burnashev and Yamamoto in a series of papers [7], [8], where they considered a communication model with a forward BSC (p) and a backward BSC (αp), and showed that when α is sufficiently small, the best error exponent is strictly larger than the one without feedback.

The lower bound in Theorem 1 shows that

$$\liminf_{\alpha\to 0} E_M(\alpha) \geq 2PM/(7M-6),$$

which is strictly less than $E_M(0) = P/2$. To obtain a better asymptotic behavior for $\alpha \to 0$, we establish the following.

Theorem 2:

$$\begin{aligned} E_M(\alpha) &\geq \frac{P}{2}\frac{1}{1+\alpha+4(\lfloor M/2\rfloor)^2\alpha+4(\lfloor M/2\rfloor)\sqrt{\alpha(1+\alpha)}}\\ &\geq \frac{P}{2}\frac{1}{(\sqrt{\alpha}M+\sqrt{1+\alpha})^2}. \end{aligned}$$

This theorem leads to the following.

Corollary 2:

$$\lim_{\alpha\to 0} E_M(\alpha) = E_M(0).$$

Thus, the lower bound in Theorem 2 is tight for $\alpha \to 0$. The proof of Theorem 2 extends Pinsker's linear noise-free feedback coding scheme [5] to the noisy case.

Fig. 2 compares the two bounds for the $M = 3$ case. The linear noisy feedback coding scheme performs better when α

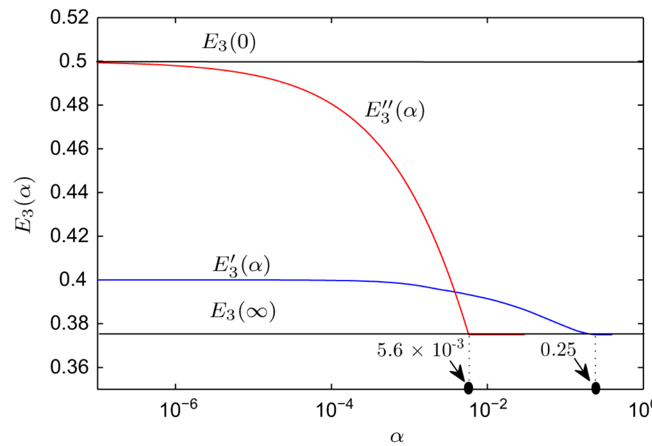


Fig. 2. Comparison of the two noisy feedback coding scheme for $M = 3$.

is sufficiently small, while the two-stage noisy feedback coding scheme performs better when α is relatively larger.

The rest of the paper is organized as follows. In Section II, we study a two-stage noisy feedback coding scheme motivated by recent results of Burnashev and Yamamoto and establish Theorem 1. In Section III, we extend Pinsker's noise-free linear feedback coding scheme to the noisy feedback case and establish Theorem 2. Section IV concludes the paper.

II. Two-Stage Noisy Feedback Coding Scheme

A. Background

It is instructive to first consider a two-stage noise-free feedback coding scheme for $M = 3$. This two-stage scheme has been studied by Schalkwijk and Barron [9] and Yamamoto and Itoh [10] for a general M.

Encoding. Fix some $\lambda \in (0, 1)$. For simplicity of notation, assume throughout that λn is an integer. To send message $w \in [1:3]$, during the transmission time interval $[1 : \lambda n]$ (namely, stage 1), the encoder uses the simplex signaling:

$$x^{\lambda n}(w) = \begin{cases} \sqrt{\lambda n P} \cdot (\quad 0, \qquad 1, 0, \dots, 0) & \text{for } w = 1 \\ \sqrt{\lambda n P} \cdot (-1/2, -\sqrt{3}/2, 0, \dots, 0) & \text{for } w = 2 \\ \sqrt{\lambda n P} \cdot (\ \ 1/2, -\sqrt{3}/2, 0, \dots, 0) & \text{for } w = 3. \end{cases} \tag{4}$$

Based on the feedback $y^{\lambda n}$, the encoder then chooses the two most probable message estimates $\hat{w}_1$ and $\hat{w}_2$, where

$$p(\hat{w}_1 | y^{\lambda n}) \ge p(\hat{w}_2 | y^{\lambda n}) \ge p(\hat{w}_3 | y^{\lambda n}) \tag{5}$$

and in case of a tie the one with the smaller index is chosen. Since the channel is Gaussian and W is uniform, (5) can be written as

$$\|x^{\lambda n}(\hat{w}_1) - y^{\lambda n}\| \le \|x^{\lambda n}(\hat{w}_2) - y^{\lambda n}\| \le \|x^{\lambda n}(\hat{w}_3) - y^{\lambda n}\|$$

where $\|\cdot\|$ denotes the Euclidean distance. During the transmission time interval $[\lambda n + 1 : n]$ (stage 2), the encoder uses antipodal signaling for w if $w \in \{\hat{w}_1, \hat{w}_2\}$ and transmits all-zero sequence, otherwise,

$$\begin{aligned} &x^n_{\lambda n+1}(w) \\ &= \begin{cases} \sqrt{(1-\lambda) n P} \cdot (1, 0, 0, \dots, 0) & \text{if } w = \min\{\hat{w}_1, \hat{w}_2\}, \\ \sqrt{(1-\lambda) n P} \cdot (-1, 0, 0, \dots, 0) & \text{if } w = \max\{\hat{w}_1, \hat{w}_2\}, \\ (0, 0, 0, \dots, 0) & \text{otherwise.} \end{cases} \end{aligned}$$

Decoding. At the end of stage 1, the decoder chooses the two most probable message estimates $\hat{w}_1$ and $\hat{w}_2$ based on $Y^{\lambda n}$ as the encoder does. At the end of stage 2, the decoder declares that $\hat{w}$ is sent if

$$\begin{aligned} \hat{w} &= \arg \min_{w \in \{\hat{w}_1, \hat{w}_2\}} \|x^n(w) - y^n\| \\ &= \arg \min_{w \in \{\hat{w}_1, \hat{w}_2\}} \Big(\|x^{\lambda n}(w) - y^{\lambda n}\|^2 \\ &\qquad\qquad + \|x^n_{\lambda n+1}(w) - y^n_{\lambda n+1}\|^2\Big)^{1/2}. \end{aligned}$$

Analysis of the probability of error. Let $\hat{W}_1$ and $\hat{W}_2$ denote the two most probable message estimates at the end of stage 1. The decoder makes an error if and only if one of the following events occurs:

$$\begin{aligned} \mathcal{E}_1 &= \{W \ne \hat{W}_1 \text{and } W \ne \hat{W}_2\} \\ \mathcal{E}_2 &= \{W \in \{\hat{W}_1, \hat{W}_2\} \text{ and } \hat{W} \ne W\}. \end{aligned}$$

Thus, the probability of error is

$$P_e^{(n)} = \mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\mathcal{E}_2).$$

By symmetry, we assume without loss of generality that $W = 1$ is sent. For brevity, we do not explicitly condition on the event $\{W = 1\}$ in probability expressions in the following, whenever it is clear from the context. Referring to Fig. 3, let

$$\begin{aligned} A_{23} = \{y^{\lambda n} &: \|x^{\lambda n}(1) - y^{\lambda n}\| \ge \|x^{\lambda n}(2) - y^{\lambda n}\| \\ &\text{and } \|x^{\lambda n}(1) - y^{\lambda n}\| \ge \|x^{\lambda n}(3) - y^{\lambda n}\|\}, \end{aligned}$$

we have

$$\begin{aligned} \mathsf{P}(\mathcal{E}_1) &= \mathsf{P}\{Y^{\lambda n} \in A_{23}\} \\ &\le Q(d_1) \\ &\overset{(a)}{\le} \frac{1}{2} \exp\left(-\frac{\lambda n P}{2}\right), \end{aligned}$$

where (a) follows since $Q(x) \le (1/2)\exp(-x^2/2)$ for $x \ge 0$ (see [11, Problem 2.26]).

On the other hand, $\mathsf{P}(\mathcal{E}_2)$ is determined by the distance between the simplex signaling in stage 1 and the distance between the antipodal signaling in stage 2 (see Fig. 4). In particular,

$$\|X^n(\hat{W}_1) - X^n(\hat{W}_2)\| = \sqrt{d_2^2 + d_3^2} = \sqrt{(4 - \lambda) n P}.$$

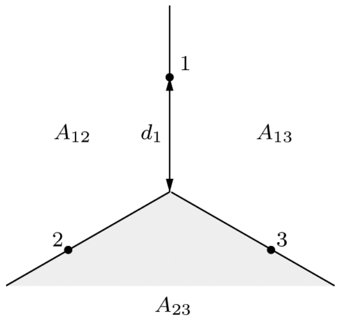


Fig. 3. The error event $\mathcal{E}_1$ when $W = 1$. Here $d_1 = \sqrt{\lambda n P}$ and 1, 2, and 3 denote $x^{\lambda n}(1)$, $x^{\lambda n}(2)$, and $x^{\lambda n}(3)$, respectively.

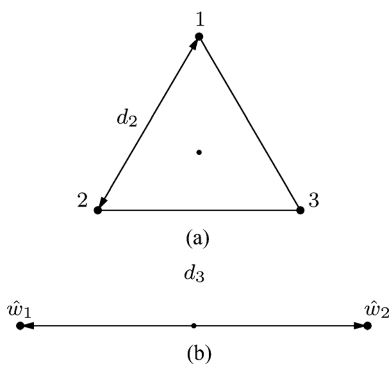


Fig. 4. The error event $\mathcal{E}_2$. Here $d_2 = \sqrt{3\lambda n P}$ and $d_3 = \sqrt{4(1-\lambda)nP}$.

Thus

$$\begin{aligned}\mathsf{P}(\mathcal{E}_2) &= Q\left(\frac{\|X^n(\hat{W}_1) - X^n(\hat{W}_2)\|}{2}\right)\\ &= Q\left(\sqrt{\left(1-\frac{\lambda}{4}\right)nP}\right)\\ &\le \frac{1}{2}\exp\left(-\frac{1}{2}\left(1-\frac{\lambda}{4}\right)nP\right).\end{aligned}$$

Therefore, the error exponent of the two-stage feedback coding scheme is lower bounded as

$$\begin{aligned}E_3'(0) &= \limsup_{n\to\infty} -\frac{1}{n}\ln P_e^{(n)}\\ &= \limsup_{n\to\infty} -\frac{1}{n}\max\{\ln \mathsf{P}(\mathcal{E}_1), \ln \mathsf{P}(\mathcal{E}_2)\}\\ &\ge \min\left\{\frac{\lambda P}{2}, \frac{P}{2}\left(1-\frac{\lambda}{4}\right)\right\}.\end{aligned}$$

Now, let $\lambda = 4/5$. Then, it can be readily verified that both terms in the minimum are the same and we have

$$E_3(0) \ge E_3'(0) \ge \frac{2P}{5}.$$

Remark 1: Since $E_3(0) = P/2$, this two-stage noise-free feedback coding scheme is strictly suboptimal.

Remark 2: We need only three transmissions: two for stage 1 and one for stage 2. Thus, λ actually divides only the total energy nP, not the block length n.

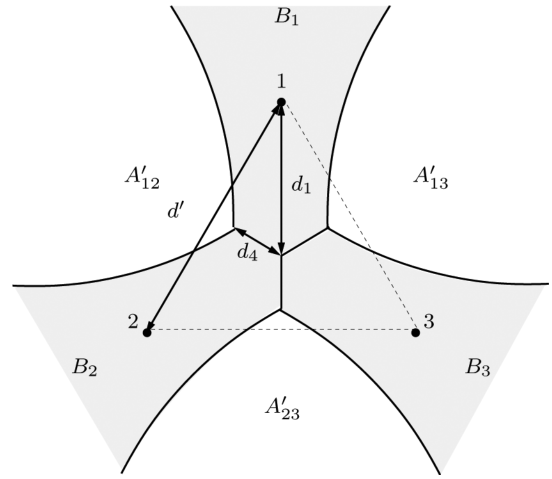


Fig. 5. Signal protection regions. The shaded areas B_w for $w = 1, 2, 3$ are the signal protection regions for $x^{\lambda n}(1)$, $x^{\lambda n}(2)$, and $x^{\lambda n}(3)$, respectively. Here $d_4 = sd_1/2 = (s/2)\sqrt{\lambda n P}$ for some parameter $s = s(t) \in [0, 1]$ to be optimized later.

B. Two-Stage Noisy Feedback Coding Scheme

Based on the two-stage noise-free feedback coding scheme in the previous section and a new idea of *signal protection* introduced by Burnashev and Yamamoto [7], [8], we present a two-stage noisy feedback coding scheme for $M = 3$. The coding scheme for an arbitrary M is given in the Appendix.

In the two-stage noise-free feedback coding scheme, the encoder and decoder agree on the same set of message estimates $\hat{w}_1$ and $\hat{w}_2$ at the end of stage 1. When there is noise in the feedback link, however, this coordination is not always possible. To solve this problem, we assign a signal protection region B_w, $w \in [1:3]$, to each signal $x^{\lambda n}(w)$ as depicted in Fig. 5. Let $x^{\lambda n}$ and $y^{\lambda n}$ denote the transmitted and received signals, respectively, and $\tilde{y}^{\lambda n}$ denote the feedback sequence at the encoder. Let $d' = \|x^{\lambda n}(1) - x^{\lambda n}(2)\| = \sqrt{3\lambda n P}$ and the signal protection region B_w for $x^{\lambda n}(w)$, $w \in [1:3]$ are defined as

$$\begin{aligned}B_w = \{y^{\lambda n}: &\|x^{\lambda n}(w) - y^{\lambda n}\| \le \|x^{\lambda n}(w') - y^{\lambda n}\|\\ &\text{for } w' \ne w\\ &\big|\|x^{\lambda n}(w') - y^{\lambda n}\| - \|x^{\lambda n}(w'') - y^{\lambda n}\|\big| \le td'\\ &\text{for } w', w'' \ne w\}\end{aligned} \tag{6}$$

which means that message w is the most probable and the other messages w' and w'' are of approximately equal posterior probabilities. Here $t \in [0, (\sqrt{3}-1)/2]$ is a fixed parameter which will be optimized later in the analysis.

Encoding. In stage 1, the encoder uses the same simplex signaling as in the noise-free feedback case [see (4)]. Then based on the noisy feedback $\tilde{y}^{\lambda n}$, the encoder chooses $\tilde{w}_1$ and $\tilde{w}_2$ such that

$$\|x^{\lambda n}(\hat{w}_1) - \tilde{y}^{\lambda n}\| \le \|x^{\lambda n}(\hat{w}_2) - \tilde{y}^{\lambda n}\| \le \|x^{\lambda n}(\hat{w}_3) - \tilde{y}^{\lambda n}\|.$$

In stage 2, the encoder uses antipodal signaling for w if $w \in \{\tilde{w}_1, \tilde{w}_2\}$ and transmits all-zero sequence otherwise.

Decoding. The decoder makes a decision immediately at the end of stage 1 if the received signal lies in one of the signal protection regions, i.e., $y^{\lambda n} \in B_w$ for $w \in [1:3]$. Otherwise, it chooses the two most probable message estimates $\hat{w}_1$ and $\hat{w}_2$ and wait for the transmission in stage 2. At the end of stage 2, the decoder declares that $\hat{w}$ is sent if

$$\begin{aligned}\hat{w} &= \arg \min_{w\in\{\hat{w}_1,\hat{w}_2\}} \|x^n(w) - y^n\| \\ &= \arg \min_{w\in\{\hat{w}_1,\hat{w}_2\}} \big(\|x^{\lambda n}(w) - y^{\lambda n}\|^2 \\ &\qquad\qquad + \|x^n_{\lambda n+1}(w) - y^n_{\lambda n+1}\|^2\big)^{1/2}.\end{aligned}$$

Remark 3: The signal protection region corresponds to the case in which the two least probable messages are of approximately equal posterior probabilities, i.e., $\|x^{\lambda n}(w) - y^{\lambda n}\| \ll \|x^{\lambda n}(w') - y^{\lambda n}\| \approx \|x^{\lambda n}(w'') - y^{\lambda n}\|$.

Analysis of the probability of error. Let $(\tilde{W}_1, \tilde{W}_2)$ and $(\hat{W}_1, \hat{W}_2)$ denote the pairs of the two most probable message estimates at the encoder and the decoder, respectively. As before, we assume that $W = 1$ is sent. Referring to Fig. 5, let

$$A'_{ww'} = A_{ww'}\backslash(\cup_{w''} B_{w''}), \quad w, w' \in [1:3]$$

where

$$\begin{aligned}A_{ww'} = \{y^{\lambda n}: \max\{\|y^{\lambda n} - x^{\lambda n}(w)\|, \|y^{\lambda n} - x^{\lambda n}(w')\|\} \\ \le \|y^{\lambda n} - x^{\lambda n}(w'')\|,\ w'' \ne w, w'\}.\end{aligned}$$

The decoder makes an error only if one or more of the following events occur.

1) Decoding error at the end of stage 1

$$\mathcal{E}_1 = \{Y^{\lambda n} \in B_2 \cup B_3 \cup A'_{23}\}.$$

2) Miscoordination due to the feedback noise

$$\begin{aligned}\tilde{\mathcal{E}}_{12} &= \{Y^{\lambda n} \in A'_{12}, \tilde{Y}^{\lambda n} \in A_{13} \cup A_{23}\} \\ \tilde{\mathcal{E}}_{13} &= \{Y^{\lambda n} \in A'_{13}, \tilde{Y}^{\lambda n} \in A_{12} \cup A_{23}\}.\end{aligned}$$

3) Decoding error at the end of stage 2

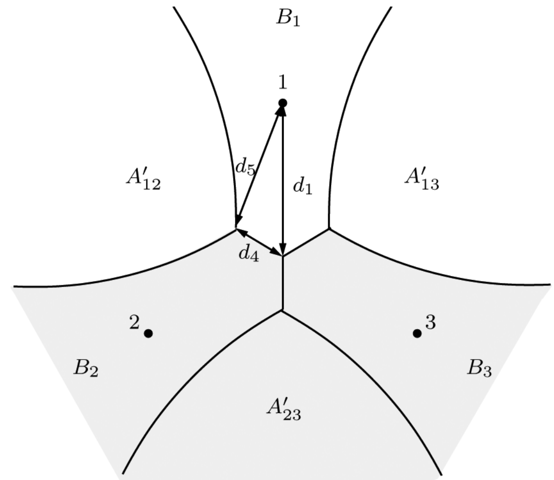
$$\mathcal{E}_2 = \{W \in \{\hat{W}_1, \hat{W}_2\} = \{\tilde{W}_1, \tilde{W}_2\} \text{and } \hat{W} \ne W\}.$$

Thus, the probability of error is upper bounded as

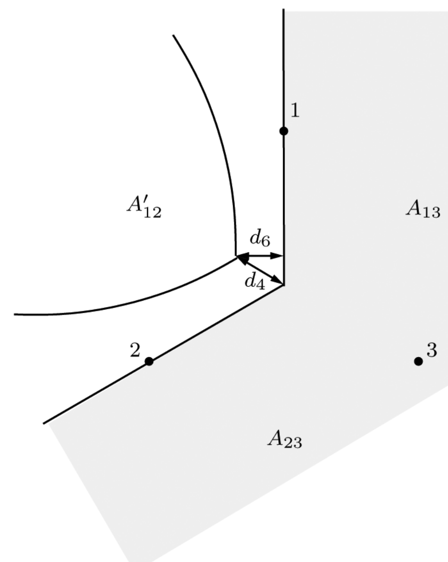
$$\begin{aligned}P_e^{(n)} &\le \mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\tilde{\mathcal{E}}_{12}) + \mathsf{P}(\tilde{\mathcal{E}}_{13}) + \mathsf{P}(\mathcal{E}_2) \\ &= \mathsf{P}(\mathcal{E}_1) + 2\mathsf{P}(\tilde{\mathcal{E}}_{12}) + \mathsf{P}(\mathcal{E}_2).\end{aligned}$$

To simplify the analysis, we introduce a new parameter $s \in [0, 1]$ such that $d_4 = sd_1/2 = (s/2)\sqrt{\lambda nP}$. It can be easily checked that $s \in [0, 1]$ corresponds to $t \in [0, (\sqrt{3} - 1)/2]$ and that this constraint guarantees that $d_5 = \min_{y^{\lambda n} \in A'_{23} \cup B_2 \cup B_3} \|x^{\lambda n}(1) - y^{\lambda n}\|$ [see Fig. 6(a)]. Hence, for the first term

$$\begin{aligned}\mathsf{P}(\mathcal{E}_1) &= \mathsf{P}\{Y^{\lambda n} \in A'_{23} \cup B_2 \cup B_3\} \\ &\le 2Q(d_5) \\ &\le \exp\left(-\frac{\lambda nP}{8}(s^2 - 2s + 4)\right).\end{aligned} \tag{7}$$



(a)



(b)

Fig. 6. (a) The error event $\mathcal{E}_1$ when $W = 1$. Since $0 \le s \le 1$, we have $d_5 = \sqrt{d_1^2 + d_4^2 - d_1 d_4} = \sqrt{(\lambda nP/4)(s^2 - 2s + 4)}$ (b) The error event $\tilde{\mathcal{E}}_{12}$ when $W = 1$ and $\{\tilde{W}_1, \tilde{W}_2\} = \{1, 3\}$. Here $d_6 = (\sqrt{3}/2)d_4 = s\sqrt{(3\lambda nP/16)}$.

The second term $\mathsf{P}(\tilde{\mathcal{E}}_{12})$ can be upper bounded [see Fig. 6(b)] as

$$\begin{aligned}\mathsf{P}(\tilde{\mathcal{E}}_{12}) &= \mathsf{P}\{Y^{\lambda n} \in A'_{12}, \tilde{Y}^{\lambda n} \in A_{13} \cup A_{23}\} \\ &\le \mathsf{P}\{\tilde{Y}^{\lambda n} \in A_{13} \cup A_{23} | Y^{\lambda n} \in A'_{12}\} \\ &\le 2Q\left(\frac{d_6}{\sqrt{\alpha}}\right) \\ &\le \exp\left(-\frac{3s^2\lambda nP}{32\alpha}\right).\end{aligned} \tag{8}$$

Finally, the third term $\mathsf{P}(\mathcal{E}_2)$ can be upper bounded in the exactly same manner as in the noise-free feedback case

$$\mathsf{P}(\mathcal{E}_2) \le \frac{1}{2}\exp\left(-\frac{1}{2}\left(1 - \frac{\lambda}{4}\right)nP\right).$$

Therefore, the error exponent of the two-stage noisy feedback coding scheme is lower bounded as

$$\begin{aligned} E_3'(\alpha) &= \limsup_{n\to\infty} -\frac{1}{n}\ln P_e^{(n)} \\ &\geq \limsup_{n\to\infty} -\frac{1}{n}\max\{\ln \mathsf{P}(\mathcal{E}_1), \ln \mathsf{P}(\tilde{\mathcal{E}}_{12}), \ln \mathsf{P}(\mathcal{E}_2)\} \\ &\geq \min\left\{\frac{\lambda P}{8}(s^2-2s+4),\ \frac{3\lambda s^2 P}{32\alpha},\ \frac{P}{2}\left(1-\frac{1}{4}\lambda\right)\right\}. \end{aligned}$$

Now let

$$\alpha = \alpha^*(s) = \frac{3s^2}{4(s^2-2s+4)}$$

and

$$\lambda = \lambda^*(s) = \frac{4}{s^2-2s+5}.$$

Then, it can be readily verified that all the three terms in the minimum are the same and we have

$$E_3'(\alpha^*(s)) \geq \frac{P}{2}\frac{s^2-2s+4}{s^2-2s+5} =: \phi(s). \tag{9}$$

Note that if $s < 1$,

$$\phi(s) > \frac{3}{8}P = E_3(\infty)$$

and $\alpha^*(s)$ is monotonically increasing over $s \in [0,1]$. Thus,

$$E_3(\alpha) > E_3(\infty) \quad \text{for } \alpha < \alpha^*(1) = \frac{1}{4}.$$

This completes the proof of Theorem 1 for the $M = 3$ case.

Remark 4: It can be easily checked that the lower bound in (9) is tight and characterizes the exact error exponent $E_3'(\alpha)$ of the two-stage noisy feedback coding scheme.

III. Linear Noisy Feedback Coding Scheme

A. Background

It is instructive to revisit (a slightly simplified version of) the linear noise-free feedback coding scheme by Pinsker [5], which shows that $E_M(0) \geq E_2(\infty) = P/2$ for all $M \geq 2$. This lower bound is tight since $E_2(0) = E_2(\infty)$[4] and $E_M(0)$ is nonincreasing in M.

Encoding. To send message $w \in [1 : M]$, the encoder transmits

$$X_1(w) = \begin{cases} \frac{L+1-w}{L}\sqrt{P} & \text{if} M = 2L+1, \\ \frac{L+1/2-w}{L}\sqrt{P} & \text{if } M = 2L. \end{cases} \tag{10}$$

Because of the feedback Y_1, the encoder can learn the noise $Z_1 = Y_1 - X_1$. Subsequently, it transmits

$$X_i = (1+\delta)Z_{i-1}, \quad i \in [2:\eta]$$

and $X_i = 0$ afterward, where $\delta > 0$ will be optimized later and the random time $\eta = \eta(w, Z^n)$ is the largest $k \leq \bar{n} = \sqrt{n}$ such that

$$\sum_{i=1}^{k} X_i^2 \leq nP.$$

Decoding. Upon receiving Y^n, the decoder estimates X_1 by

$$\hat{X}_1 = \sum_{i=1}^{\bar{n}} (-1)^{i-1}\frac{Y_i}{(1+\delta)^{i-1}}$$

and declares that $\hat{w}$ is sent if

$$\hat{w} = \arg\min_{w\in[1:M]} |X_1(w) - \hat{X}_1|.$$

Remark 5: It can be easily checked that each time $i \in [2:\eta]$, the encoder transmits the error

$$\sum_{j=1}^{i-1}(-1)^{j-1}\frac{Y_j}{(1+\delta)^{j-1}} - X_1 = (-1)^{i-2}\frac{Z_{i-1}}{(1+\delta)^{i-2}}$$

in the decoder's current estimate of the initial transmission (up to scaling). Thus, Pinsker's coding scheme is another instance of iterative refinement used in the Schalkwijk–Kailath coding scheme [1] for the Gaussian channel and the Horstein coding scheme [12] for the binary symmetric channel.

Analysis of the probability of error. For simplicity of notation, assume throughout that $\bar{n} = \sqrt{n}$ is an integer. We use ϵ_n to denote a generic sequence of nonnegative numbers that tends to zero as $n \to \infty$. When there are multiple such functions $\epsilon_n^{(1)}, \epsilon_n^{(2)}, \cdots, \epsilon_n^{(k)}$, we denote them all by ϵ_n with the understanding that $\epsilon_n = \max\{\epsilon_n^{(1)}, \epsilon_n^{(2)}, \cdots, \epsilon_n^{(k)}\}$. It is easy to see that decoding error occurs only if $|X_1(w) - \hat{X}_1| > \sqrt{P}/(2L)$. The probability of error is thus upper bounded as

$$P_e^{(n)} = \mathsf{P}\{W \neq \hat{W}\} \leq \mathsf{P}\left\{|X_1 - \hat{X}_1| > \frac{\sqrt{P}}{2L}\right\}.$$

The key idea in the analysis is to introduce a "virtual" transmission

$$X_i' = \begin{matrix} X_1 & \text{if} i = 1 \\ (1+\delta)Z_{i-1} & \text{if } i \in [2:\bar{n}] \\ 0 & \text{otherwise.} \end{matrix} \tag{11}$$

Let

$$Y_i' = X_i' + Z_i \tag{12}$$

and define the estimate $\hat{X}_1'$ of X_1' as

$$\hat{X}_1' = \sum_{i=1}^{\bar{n}}(-1)^{i-1}\frac{Y_i'}{(1+\delta)^{i-1}}. \tag{13}$$

Then, it can be easily shown that

$$\hat{X}_1' = X_1 + (-1)^{\bar{n}-1}\frac{Z_{\bar{n}}}{(1+\delta)^{\bar{n}-1}}.$$

Thus, we have

$$\begin{aligned}
&\mathsf{P}\left\{|X_1 - \hat{X}_1| > \frac{\sqrt{P}}{2L}\right\}\\
&\le \mathsf{P}\left\{|X_1 - \hat{X}_1'| + |\hat{X}_1' - \hat{X}_1| > \frac{\sqrt{P}}{2L}\right\}\\
&\le \mathsf{P}\left\{|X_1 - \hat{X}_1'| > \frac{\sqrt{P}}{2L}\right\} + \mathsf{P}\{|\hat{X}_1' - \hat{X}_1| > 0\}\\
&=: P_1 + P_2.
\end{aligned}$$

Now we upper bound the two terms. For the first term, we have

$$\begin{aligned}
P_1 &= \mathsf{P}\left\{\left|\frac{Z_{\bar{n}}}{(1+\delta)^{\bar{n}-1}}\right| > \frac{\sqrt{P}}{2L}\right\}\\
&= 2Q\left(\frac{\sqrt{P}(1+\delta)^{\bar{n}-1}}{2L}\right)\\
&\le \exp\left(-\frac{P(1+\delta)^{2(\bar{n}-1)}}{8L^2}\right).
\end{aligned}$$

For the second term, note that $X_i = X_i'$ for all $i \in [1:n]$ if and only if $\sum_{i=1}^{\bar{n}} X_i^2 \le nP$ (i.e., $\bar{n} = \eta$), and thus that $\hat{X}_1' \ne \hat{X}_1$ only if $\sum_{i=1}^{\bar{n}} X_i^2 > nP$. Therefore

$$\begin{aligned}
P_2 &\le \mathsf{P}\left\{\sum_{i=1}^{\bar{n}} X_i^2 > nP\right\}\\
&\overset{(a)}{\le} \mathsf{P}\left\{\sum_{i=2}^{\bar{n}}(1+\delta)^2 Z_{i-1}^2 > (n-1)P\right\}\\
&= \mathsf{P}\left\{\chi_{\bar{n}-1}^2 > \frac{(n-1)P}{(1+\delta)^2}\right\},
\end{aligned}$$

where (a) follows since $X_1^2 \le P$ [recall (10)] and $\chi_{\bar{n}-1}^2$ denotes a chi-square random variable with $\bar{n}-1$ degrees of freedom. By upper bounding the tail probability of the chi-square random variable [13] as

$$\mathsf{P}\{\chi_k^2 > x\} \le \exp\left(-\frac{x}{2} + \frac{k}{2}\log\frac{ex}{k}\right) \text{ for any } k \ge 1 \text{ and } x \ge k \tag{14}$$

we have

$$\begin{aligned}
P_2 &\le \mathsf{P}\left\{\chi_{\bar{n}-1}^2 > \frac{(n-1)P}{(1+\delta)^2}\right\}\\
&\le \exp\left(-\frac{1}{2}\frac{(n-1)P}{(1+\delta)^2} + \frac{\bar{n}-1}{2}\log\frac{e(n-1)P}{(\bar{n}-1)(1+\delta)^2}\right)\\
&\le \exp\left(-\frac{1}{2}\frac{(n-1)P}{(1+\delta)^2} + \frac{\bar{n}-1}{2}\log\frac{e(n-1)P}{(\bar{n}-1)}\right)\\
&\le \exp\left(-\frac{1}{2}\frac{nP}{(1+\delta)^2} + n\epsilon_n\right),
\end{aligned}$$

where ϵ_n tends to zero as $n \to \infty$. Therefore, the error exponent of the linear feedback coding scheme is lower bounded as

$$\begin{aligned}
E_M''(0) &\ge \limsup_{n\to\infty} -\frac{1}{n}\ln P_e^{(n)}\\
&= \limsup_{n\to\infty} -\frac{1}{n}\max\{\ln P_1, \ln P_2\}\\
&\ge \limsup_{n\to\infty} \min\left\{\frac{P(1+\delta)^{2(\bar{n}-1)}}{8nL^2},\ \frac{P}{2(1+\delta)^2}\right\}
\end{aligned}$$

for any $\delta > 0$. Now let

$$\delta = \delta(n) = \frac{\ln(4nL^2)}{2\bar{n}}$$

which tends to zero as $n \to \infty$. Then, the limits of both terms in the minimum are the same. Therefore,

$$E_M''(0) \ge \limsup_{n\to\infty}\frac{P}{2(1+\delta(n))^2} = \frac{P}{2}$$

which completes the proof of achievability.

B. Linear Noisy Feedback Coding Scheme

Now we formally describe and analyze a linear noisy feedback coding scheme based on Pinsker's noise-free feedback coding scheme.

Encoding. Fix some $\lambda \in (0,1)$. To send message $w \in [1:M]$, the encoder transmits

$$X_1(w) = \begin{cases} \frac{L+1-w}{L}\sqrt{\lambda n P} & \text{if} M = 2L+1,\\ \frac{L+1/2-w}{L}\sqrt{\lambda n P} & \text{if } M = 2L. \end{cases} \tag{15}$$

Because of the noisy feedback $\tilde{Y}_1$, the encoder can learn $Z_1 + \tilde{Z}_1 = \tilde{Y}_1 - X_1$. Subsequently, it transmits

$$X_i = (1+\delta)(Z_{i-1} + \tilde{Z}_{i-1}), \quad i \in [2:\eta]$$

where $\delta > 0$ will be optimized later and the random time $\eta = \eta(w, Z^n, \tilde{Z}^n)$ is the largest $k \le \bar{n} = \sqrt{n}$ such that

$$\sum_{i=1}^{k} X_i^2 \le nP.$$

Decoding. Upon receiving Y^n, the decoder estimates X_1 by

$$\hat{X}_1 = \sum_{i=1}^{\bar{n}}(-1)^{i-1}\frac{Y_i}{(1+\delta)^{i-1}}$$

and declares that $\hat{w}$ is sent if

$$\hat{w} = \arg\min_{w\in[1:M]}|X_1(w) - \hat{X}_1|.$$

Remark 6: The main difference between this noisy feedback coding scheme and Pinsker's noise-free feedback coding scheme in the previous section is that we let the power of the initial transmission grow linearly with the block length n [exploiting the peak energy constraint in (1)] and thus the initial transmission contains much more information about the message than in Pinsker's scheme. This makes the coding scheme more robust to combat the noise in the feedback link.

Analysis of the probability of error. As before, we assume that $\bar{n} = \sqrt{n}$ is an integer. Let

$$X_i' = \begin{cases} X_1 & \text{if } i = 1 \\ (1+\delta)(Z_{i-1} + \tilde{Z}_{i-1}) & \text{if } i \in [2:\bar{n}] \\ 0 & \text{otherwise,} \end{cases} \tag{16}$$

and let Y_i' and $\hat{X}_1'$ be defined as in (12) and (13). Then, it can be easily shown that

$$\hat{X}_1' = X_1 + (-1)^{\bar{n}-1}\frac{Z_{\bar{n}}}{(1+\delta)^{\bar{n}-1}} + \sum_{i=1}^{\bar{n}-1}(-1)^i\frac{\tilde{Z}_i}{(1+\delta)^{i-1}}.$$

Thus, we have

$$\begin{aligned} P_e^{(n)} &= \mathsf{P}\{W \neq \hat{W}\} \\ &\le \mathsf{P}\left\{|X_1 - \hat{X}_1| > \frac{\sqrt{\lambda nP}}{2L}\right\} \\ &\le \mathsf{P}\left\{|X_1 - \hat{X}_1'| + |\hat{X}_1' - \hat{X}_1| > \frac{\sqrt{\lambda nP}}{2L}\right\} \\ &\le \mathsf{P}\left\{|X_1 - \hat{X}_1'| > \frac{\sqrt{\lambda nP}}{2L}\right\} + \mathsf{P}\{|\hat{X}_1' - \hat{X}_1| > 0\} \\ &=: P_1 + P_2. \end{aligned}$$

Now we upper bound the two terms. For the first term, we have

$$\begin{aligned} P_1 &= \mathsf{P}\left\{\left|(-1)^{\bar{n}-1}\frac{Z_{\bar{n}}}{(1+\delta)^{\bar{n}-1}} + \sum_{i=1}^{\bar{n}-1}(-1)^i\frac{\tilde{Z}_i}{(1+\delta)^{i-1}}\right| > \frac{\sqrt{\lambda nP}}{2L}\right\} \\ &= 2Q\left(\frac{\sqrt{\lambda nP/N}}{2L}\right) \\ &\le \exp\left(-\frac{\lambda nP}{8L^2N}\right) \end{aligned}$$

where

$$\begin{aligned} N &= \sum_{i=1}^{\bar{n}-1}\frac{\alpha}{(1+\delta)^{2(i-1)}} + \frac{1}{(1+\delta)^{2(\bar{n}-1)}} \\ &= \frac{\alpha\left(1 - \frac{1}{(1+\delta)^{2(\bar{n}-2)}}\right)}{1 - \frac{1}{(1+\delta)^2}} + \frac{1}{(1+\delta)^{2(\bar{n}-1)}} \\ &\le \frac{\alpha(1+\delta)^2}{(1+\delta)^2 - 1} + \epsilon_n, \end{aligned}$$

where ϵ_n tends to zero as $n \to \infty$. Thus,

$$P_1 \le \exp\left(-\frac{\lambda nP}{8L^2}\left(\frac{\alpha(1+\delta)^2}{(1+\delta)^2-1} + \epsilon_n\right)^{-1}\right). \tag{17}$$

For the second term, we have

$$\begin{aligned} P_2 &\le \mathsf{P}\left\{\sum_{i=1}^{\bar{n}} X_i^2 > nP\right\} \\ &\overset{(a)}{\le} \mathsf{P}\left\{\sum_{i=2}^{\bar{n}}(1+\delta)^2(Z_{i-1} + \tilde{Z}_{i-1})^2 > (1-\lambda)nP\right\} \\ &= \mathsf{P}\left\{\chi^2_{\bar{n}-1} > \frac{(1-\lambda)nP}{(1+\delta)^2(1+\alpha)}\right\}, \end{aligned}$$

where (a) follows since $X_1 \le \lambda nP$ [recall (15)]. By (14), we have

$$\begin{aligned} P_2 &\le \mathsf{P}\left\{\chi^2_{\bar{n}-1} > \frac{(1-\lambda)nP}{(1+\delta)^2(1+\alpha)}\right\} \\ &\le \exp\left(-\frac{1}{2}\frac{(1-\lambda)nP}{(1+\delta)^2(1+\alpha)} + \frac{\bar{n}-1}{2}\log\frac{e(1-\lambda)nP}{(\bar{n}-1)(1+\delta)^2(1+\alpha)}\right) \\ &\le \exp\left(-\frac{1}{2}\frac{(1-\lambda)nP}{(1+\delta)^2(1+\alpha)} + n\epsilon_n\right), \end{aligned} \tag{18}$$

where ϵ_n tends to zero as $n \to \infty$. Therefore, the error exponent of the linear noisy feedback coding scheme is lower bounded as

$$\begin{aligned} E_M''(\alpha) &= \limsup_{n\to\infty} -\frac{1}{n}\ln P_e^{(n)} \\ &= \limsup_{n\to\infty} -\frac{1}{n}\max\{\ln P_1, \ln P_2\} \\ &\ge \min\left\{\frac{\lambda P}{8L^2\alpha}\frac{(1+\delta)^2-1}{(1+\delta)^2}, \frac{(1-\lambda)P}{2(1+\delta)^2(1+\alpha)}\right\}. \end{aligned}$$

Now let

$$\delta = \delta(\alpha) = \left(1 + \sqrt{\frac{4L^2\alpha}{1+\alpha}}\right)^{1/2} - 1$$

and

$$\lambda = \lambda(\alpha) = \left(1 + \sqrt{\frac{1+\alpha}{4L^2\alpha}}\right)^{-1}.$$

Then, it can be readily verified that both terms in the minimum are the same and we have

$$E_M''(\alpha) \ge \frac{P}{2}\frac{1}{1+\alpha+4(\lfloor M/2\rfloor)^2\alpha+4(\lfloor M/2\rfloor)\sqrt{\alpha(1+\alpha)}}$$

which completes the proof of Theorem 2.

IV. Discussion

When α is very small, the linear feedback coding scheme (which is optimal for noise-free feedback) outperforms the two-stage (nonlinear) feedback coding scheme. When α is relatively large, however, linear feedback coding scheme amplifies the feedback noise, while the two-stage scheme achieves a more robust performance via signal protection. While this dichotomy agrees with the usual engineering intuition, it would be aesthetically more pleasing if a single feedback coding scheme performs uniformly better over all ranges of α, and the search for such a coding scheme invites further investigation. We finally note that $\alpha^* = 1/4$ is the threshold for all M in the two-stage noisy feedback coding scheme (see the Appendix). In both schemes, the error exponents are strictly larger than those for the no feedback case only when α is sufficiently small. Thus, it is natural to ask whether the noisy feedback is useful for all α or there exists a fundamental threshold beyond which noisy feedback becomes useless.

Following Burnashev and Yamamoto's work [14] on noisy feedback communication over the binary symmetric channel at positive rates, we can extend our result to a positive rate, i.e., $M = e^{nR}$ with $R > 0$. Let $E(R;\alpha)$ denote the maximum error exponent, namely, the reliability function. Although the $E(R;\infty)$ is not known for all $R \in [0, C]$ (see, e.g., [15]), Shannon [6] showed that

$$E(0+;\infty) := \lim_{R\to 0} E(R;\infty) = \frac{P}{4}.$$

We can easily adapt the analysis of our two-stage noisy feedback coding scheme in the Appendix to show that

$$\lim_{\alpha\to 0} E(0+;\alpha) = \frac{2}{7}P > E(0+;\infty).$$

Moreover, it can be shown that

$$E(R;\alpha) > E(R;\infty) \quad \text{for } R < \frac{P}{24} \text{and } \alpha < \alpha(s)$$

where $s \in [0,1]$ is the root of $(s-1)^2 = 24R/P$. Thus, the best error exponent can be strictly larger than the one without feedback if the rate and the feedback noise power are sufficiently small.

Finally, we note that our discussion has been limited to the peak energy constraint (1). In some practical systems, however, it would be more relevant to consider peak power constraints

$$\mathsf{P}\{x_i^2(w, \tilde{Y}^{i-1}) \le P\} = 1 \quad \text{for all } w \text{ and } i$$

or

$$\mathsf{E}[x_i^2(w, \tilde{Y}^{i-1})] \le P \quad \text{for all } w \text{ and } i.$$

It remains to be seen whether noisy feedback still improves the reliability under these more stringent conditions.

Appendix
Proof of Theorem 1 for the General Case

Encoding. In stage 1, the encoder uses the simplex signaling for an M-ary message

$$x^{\lambda n}(w) = A\left(e_w - \frac{1}{M}\sum_{w=1}^{M} e_w\right) \quad \text{for } w \in [1:M]$$

where $A = \sqrt{M\lambda n P/(M-1)}$ and

$$e_w = (\underbrace{0,\cdots,0}_{w-1}, 1, 0, \cdots, 0).$$

Then based on the noisy feedback $\tilde{y}^{\lambda n}$, the encoder chooses the two most probable message estimates $\tilde{w}_1$ and $\tilde{w}_2$ among M candidates. In stage 2, the encoder uses antipodal signaling for w if $w \in \{\tilde{w}_1, \tilde{w}_2\}$ and transmits all-zero sequence otherwise.

Decoding. The signal protection region for the M-ary message is defined as in (6) (with w, w', $w'' \in [1:M]$). The decoder makes a decision immediately at the end of stage 1 if the received signal $y^{\lambda n}$ lies in one of the signal protection regions. Otherwise, it chooses the two most probable message estimates $\hat{w}_1$ and $\hat{w}_2$, and wait for the transmission in stage 2. At the end of stage 2, the decoder declares that $\hat{w}$ is sent if

$$\hat{w} = \arg\min_{w\in\{\hat{w}_1,\hat{w}_2\}} \left(\|x^{\lambda n}(w) - y^{\lambda n}\|^2 + \|x^n_{\lambda n+1}(w) - y^n_{\lambda n+1}\|^2\right)^{1/2}.$$

Analysis of the probability of error. Let $(\tilde{W}_1, \tilde{W}_2)$ and $(\hat{W}_1, \hat{W}_2)$ denote the pairs of the two most probable message estimates at the encoder and the decoder, respectively. The decoder makes an error only if one or more of the following events occur.

1) Decoding error at the end of stage 1

$$\mathcal{E}_1 = \{Y^{\lambda n} \in \cup_{w\neq 1} B_w \cup (\cup_{w,w'\neq 1} A'_{ww'})\}.$$

2) Miscoordination due to the feedback noise

$$\tilde{\mathcal{E}}_{1w} = \{Y^{\lambda n} \in A'_{1w} \text{ and } \tilde{Y}^{\lambda n} \in \cup_{\{w',w''\}\neq\{1,w\}} A_{w'w''}\}.$$

3) Decoding error at the end of stage 2

$$\mathcal{E}_2 = \{W \in \{\hat{W}_1, \hat{W}_2\} = \{\tilde{W}_1, \tilde{W}_2\} \text{ and } \hat{W} \neq W\}.$$

Thus, the probability of error is upper bounded as

$$P_e^{(n)} \le \mathsf{P}(\mathcal{E}_1) + M\mathsf{P}(\tilde{\mathcal{E}}_{1w}) + \mathsf{P}(\mathcal{E}_2).$$

As before, we assume that $W = 1$ was sent. For the first term, by the union of events bound,

$$\begin{aligned}\mathsf{P}(\mathcal{E}_1) &= \mathsf{P}\{Y^{\lambda n} \in \cup_{w\neq 1} B_w \cup (\cup_{w,w'\neq 1} A'_{ww'})\} \\ &\le M^2 \mathsf{P}\{Y^{\lambda n} \in B_2 \cup A'_{23}\}.\end{aligned}$$

For $\mathsf{P}(\tilde{\mathcal{E}}_{1w})$, again by the union of events bound,

$$\begin{aligned}\mathsf{P}(\tilde{\mathcal{E}}_{1w}) &= \mathsf{P}\{Y^{\lambda n} \in A'_{1w} \text{ and } \tilde{Y}^{\lambda n} \in \cup_{\{w',w''\}\neq\{1,w\}} A_{w'w''}\}\\ &\leq M^2 \mathsf{P}\{Y^{\lambda n} \in A'_{1w} \text{ and } \tilde{Y}^{\lambda n} \in A_{w'w''}\}.\end{aligned}$$

We use d'_j, $j \in [1:6]$, to denote the distances corresponding to d_j in the $M = 3$ case (see Fig. 6). It can be easily checked that $d'_j = d_j\sqrt{3(M-1)/(2M)}$. Thus, by replacing d_5 by d'_5 in (7) and d_6 by d'_6 in (8), we have

$$\begin{aligned}\mathsf{P}(\mathcal{E}_1) &\leq M^2 Q(d'_5)\\ &\leq \frac{M^2}{2}\exp\left(-\frac{M}{12(M-1)}\lambda n P(s^2-2s+4)\right)\end{aligned}$$

and

$$\mathsf{P}(\tilde{\mathcal{E}}_{12}) \leq M^2 Q\left(\frac{d'_6}{\sqrt{\alpha}}\right) \leq \frac{M^2}{2}\exp\left(-\frac{s^2 M}{16(M-1)\alpha}\lambda n P\right).$$

The third term $\mathsf{P}(\mathcal{E}_2)$ can be upper bounded in the same manner as for the $M = 3$ case,

$$\begin{aligned}\mathsf{P}(\mathcal{E}_2) &= Q\left(-\sqrt{\left(1-\frac{M-2}{2(M-1)}\lambda\right)nP}\right)\\ &\leq \frac{1}{2}\exp\left(-\frac{nP}{2}\left(1-\frac{M-2}{2(M-1)}\lambda\right)\right).\end{aligned}$$

Therefore,

$$\begin{aligned}&E'_M(\alpha)\\ &= \limsup_{n\to\infty} -\frac{1}{n}\ln P_e^{(n)}\\ &\geq \limsup_{n\to\infty} -\frac{1}{n}\max\{\ln \mathsf{P}(\mathcal{E}_1), \ln\big(M\mathsf{P}(\tilde{\mathcal{E}}_{12})\big), \ln \mathsf{P}(\mathcal{E}_2)\}\\ &\geq \min\left\{\frac{\lambda M P}{12(M-1)}(s^2-2s+4),\ \frac{s^2\lambda M P}{16(M-1)\alpha},\right.\\ &\qquad\qquad \left.\frac{P}{2}\left(1-\frac{M-2}{2(M-1)}\lambda\right)\right\}.\end{aligned}$$

Now let

$$\alpha = \alpha^*(s) = \frac{3s^2}{4(s^2-2s+4)}$$

and

$$\lambda = \lambda^*(s) = \left(\frac{M}{6(M-1)}(s^2-2s+4) + \frac{M-2}{2(M-1)}\right)^{-1}.$$

Then, it can be readily verified that all the three terms in the minimum are the same and we have

$$\begin{aligned}E'_M(\alpha^*(s)) &\geq \frac{P}{2}\left(1-\frac{3(M-2)}{M(s^2-2s+4)+3(M-2)}\right)\\ &=: \phi(s).\end{aligned}$$

Note that if $s < 1$,

$$\phi(s) > \frac{M}{4(M-1)}P = E_M(\infty)$$

and $\alpha^*(s)$ is monotonically increasing over $s \in [0,1]$. Thus,

$$E'_M(\alpha) > E_M(\infty) \quad \text{for } \alpha < \alpha^*(1) = \frac{1}{4}.$$

This completes the proof of Theorem 1 for the general case.

Remark 7: Note that $E'_M(\alpha)$ is decreasing in M, while $\alpha^*(s)$ is still independent of M.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for many valuable comments that helped to improve the presentation of this paper.

REFERENCES

[1] J. P. M. 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.

[2] Y.-H. Kim, A. Lapidoth, and T. Weissman, "The Gaussian channel with noisy feedback," in *Proc. IEEE Int. Symp. Inf. Theory*, Nice, France, Jun. 2007, pp. 1416–1420.

[3] A. D. Wyner, "On the Schalkwijk-Kailath coding scheme with a peak energy constraint," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 129–134, Jan. 1968.

[4] L. A. Shepp, J. K. Wolf, A. D. Wyner, and J. Ziv, "Binary communication over the Gaussian channel using feedback with a peak energy constraint," *IEEE Trans. Inf. Theory*, vol. IT-15, no. 4, pp. 476–478, Jul. 1969.

[5] M. S. Pinsker, "The probability of error in block transmission in a memoryless Gaussian channel with feedback," *Probl. Inf. Transmiss.*, vol. 4, no. 4, pp. 3–19, 1968.

[6] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, pp. 611–656, 1959.

[7] M. Burnashev and H. Yamamoto, "On BSC, noisy feedback and three messages," in *Proc. IEEE Int. Symp. Inf. Theory*, Toronto, Canada, Jul. 2008, pp. 886–889.

[8] M. Burnashev and H. Yamamoto, "On zero-rate error exponent for BSC with noisy feedback," *Probl. Inf. Transmiss.*, vol. 44, no. 3, pp. 33–49, 2008.

[9] J. P. M. Schalkwijk and M. Barron, "Sequential signalling under a peak power constraint," *IEEE Trans. Inf. Theory*, vol. IT-17, no. 3, pp. 278–282, May 1971.

[10] H. Yamamoto and K. Itoh, "Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 6, pp. 729–733, Nov. 1979.

[11] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. New York, NY, USA: Wiley, 1965.

[12] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Trans. Inf. Theory*, vol. IT-9, no. 3, pp. 136–143, Jul. 1963.

[13] T. Inglot and T. Ledwina, “Asymptotic optimality of new adaptive test in regression model,” *Ann. Inst. H. Poincaré Probab. Statist.*, vol. 42, no. 5, pp. 579–590, 2006.
[14] M. Burnashev and H. Yamamoto, “On reliability function of BSC with noisy feedback,” *Probl. Inf. Transmiss.*, vol. 46, no. 2, pp. 2–23, 2010.
[15] A. Ashikhmin, A. Barg, and S. Litsyn, “A new upper bound on the reliability function of the Gaussian channel,” *IEEE Trans. Inf. Theory*, vol. IT-46, no. 6, pp. 1945–1961, Sep. 2000.

Yu Xiang (S’10) received the B.S. degree in telecommunication engineering from Xidian University, Xi’an, China, in 2008 and the M.S. degree in electrical engineering from University of California, San Diego, CA, USA, in 2010. He is currently a Ph.D. candidate in the Department of Electrical and Computer Engineering, University of California, San Diego.

His research interests include information theory, statistics, and learning theory.

Young-Han Kim (S’99–M’06–SM’12) received the B.S. (Hons.) degree in electrical engineering from Seoul National University, Seoul, Korea, in 1996 and the M.S. degrees in electrical engineering and in statistics, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 2001, 2006, and 2006, respectively.

In July 2006, he joined the University of California, San Diego, where he is an Associate Professor of Electrical and Computer Engineering. His research interests include statistical signal processing and information theory, with applications in communication, control, computation, networking, data compression, and learning.

Dr. Kim is a recipient of the 2008 NSF Faculty Early Career Development (CAREER) Award, the 2009 US-Israel Binational Science Foundation Bergmann Memorial Award, and the 2012 IEEE Information Theory Paper Award. He is currently on the Editorial Board of the IEEE TRANSACTIONS ON INFORMATION THEORY, serving as an Associate Editor for Shannon Theory. He is also serving as a Distinguished Lecturer for the IEEE Information Theory Society.