On the AWGN Channel with Noisy Feedback and Peak Energy Constraint

Yu Xiang and Young-Han Kim
Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA 92093, USA
Email: {yxiang,yhk}@ucsd.edu

***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. Previously, Shepp, Wolf, Wyner, and Ziv, and Pinsker showed that under the *peak* energy constraint the best error exponent for transmission of *two* messages is achieved by 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 proved the contrary by showing that the best error exponent for sending M messages does not depend on M, and hence can be strictly larger than the best error exponent without feedback. This paper further extends this and shows that if the noise level in the feedback link is sufficiently small, then the best error exponent for transmission of *three* messages can be strictly larger than the one without feedback. This result is motivated by a series of recent papers of Burnashev and Yamamoto who considered a similar problem over binary symmetric channels.**

I. Introduction and Main Result

We consider a communication problem for an additive white Gaussian noise (AWGN) *forward* channel with feedback over an orthogonal additive white Gaussian noise *backward* channel as depicted in Fig. 1. One wishes to communicate a message

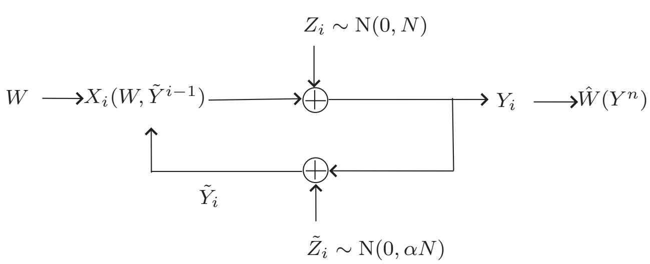

Fig. 1. Gaussian channel with noisy feedback.

index $W \in \{1, 2, \cdots, M\}$ over the (forward) additive white Gaussian noise channel

$$Y_i = X_i + Z_i, \tag{1}$$

where X_i, Y_i, and Z_i respectively denote the channel input, channel output, and additive Gaussian noise. Let further $\tilde{Y}_i$ denote a noisy version of Y_i over the feedback (backward) additive white Gaussian noise channel

$$\tilde{Y}_i = Y_i + \tilde{Z}_i, \tag{2}$$

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 respectively white Gaussian $\mathrm{N}(0, N)$ and $\mathrm{N}(0, \alpha N)$.

We define an (M, n) code with the encoding functions $x_i(w, \tilde{y}^{i-1})$, $i = 1, 2, \ldots, n$, and the decoding function $\hat{w}(y^n)$. Thus the encoder has the causal access to the noisy feedback $\tilde{Y}^n$. The probability of error $P_e^{(n)}$ of the code is defined by

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

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

There are two commonly used power constraints on the encoding functions:

1) Expected block power constraint (or expected energy constraint)

$$\sum_{i=1}^{n} \mathsf{E} X_i^2(w, \tilde{Y}^{i-1}) \leq nP \quad \text{for all } w. \tag{3}$$

2) Peak block power constraint (or peak energy constraint)

$$\mathsf{P}\{\sum_{i=1}^{n} X_i^2(w, \tilde{Y}^{i-1}) \leq nP\} = 1 \quad \text{for all } w. \tag{4}$$

The goal of this paper is to study the role of noisy feedback in communication. As is well known, the capacity stays the same with 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 [6] that *noise-free* feedback can improve the error exponent dramatically under the expected energy constraint. Kim, Lapidoth, and Weissman [4] studied the optimal error exponent under the *expected* energy constraint and noisy feedback and showed that the error exponent is inversely proportional to α for small α.

978-1-4244-7892-7/10/$26.00 ©2010 IEEE

Another important factor that determines the error exponent is the energy constraint on the channel inputs. Most notably, Wyner [7] pointed out that the error probability of the Schalkwijk–Kailath coding scheme [6] degrades to a single exponential form under the *peak* energy constraint. In fact, Shepp, Wolf, Wyner, and Ziv [3], and Pinsker [5] showed that for the two-message case ($M = 2$), the best error exponent is achieved by simple nonfeedback antipodal signaling. Thus, it might apparently seem that even noise-free feedback cannot improve the reliability of the communication under the peak energy constraint.

In this paper, we show that noisy feedback can improve the reliability of communication under the peak energy constraint, if the noise level α is small enough. More specifically, we consider the three-message case ($M = 3$) and show that the best error exponent is strictly larger than the one for the nonfeedback case.

Let

$$E_M := \lim_{n\to\infty} -\frac{1}{n} \ln P_e^*(M, n), \tag{5}$$

where $P_e^*(M, n)$ denotes the best error probability over all (M, n) codes for the AWGN channel without feedback ($\alpha = \infty$). In other words, E_M denotes the best error exponent for sending M messages over the AWGN channel without feedback. Shannon [8] showed that

$$E_M = \frac{M}{4(M-1)} \frac{P}{N}. \tag{6}$$

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 $\sqrt{nP}$, that is, each codeword $\mathbf{x}_w = x^n(w)$ satisfies $\sum_{i=1}^n x_i^2(w) = nP$ and is at the same Euclidean distance from other codewords. In particular, for $M = 3$

$$\begin{aligned} \mathbf{x}_1 &= \sqrt{nP} \cdot [1, 0, 0, \cdots, 0], \\ \mathbf{x}_2 &= \sqrt{nP} \cdot [-1/2, \sqrt{3}/2, 0, \cdots, 0], \\ \mathbf{x}_3 &= \sqrt{nP} \cdot [-1/2, -\sqrt{3}/2, 0, \cdots, 0], \end{aligned}$$

and

$$E_3 = \frac{3}{8} \frac{P}{N}. \tag{7}$$

Now denote by F_M the best error exponent for sending M messages over the AWGN channel with noise-free feedback ($\alpha = 0$). Pinsker [5] showed that

$$F_M \equiv \frac{P}{2N} \tag{8}$$

for all M. In particular,

$$F_3 = \frac{P}{2N}. \tag{9}$$

Generally, we denote by $F_M(\alpha)$ the best error exponent for sending M messages over the AWGN channel with the noisy feedback $\tilde{Z}_i \sim \mathrm{N}(0, \alpha N)$. Clearly, $F_M(\alpha)$ is decreasing in α and

$$E_M = F_M(\infty) \le F_M(\alpha) \le F_M(0) = F_M$$

for all α and all M.

We are ready to state our main result:

Theorem 1: If

$$\alpha \le \frac{25 - 2\sqrt{3}}{9(7 + 6\sqrt{3})^2} \approx 8 \times 10^{-3},$$

we have

$$F_3(\alpha) > E_3.$$

Thus, if the noise on 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 and resembles of recent results of Burnashev and Yamamoto in a series of papers [1], [2]. They considered a communication model with a BSC(p) forward channel and a BSC(αp) feedback channel, and showed that the best error exponent is strictly larger than that for the nonfeedback channel if α is below some positive level.

In the next section, we first describe the transmission strategy, which is essentially from Burnashev and Yamamoto. Then we provide an analysis of the probability of error of this strategy follows in Section III.

II. Transmission Strategy for Noisy Feedback

The transmission has two stages. During the transmission time interval $\{1, 2, \cdots, \lambda n\}$ (stage I) for some $\lambda \in (0, 1)$, we use simplex signaling of the three messages. Assume that at time λn, the posterior probability of one of the three messages is very small compared with the other two messages. Then, if the transmitter would be able to identify correctly and rule out this message from the noisy feedback, then the remaining transmission time interval $\{\lambda n + 1, \lambda n + 2, \cdots, n\}$ (stage II) can be used to distinguish between the other two (most probable) messages. Since the error exponent E_2 used to distinguish between the two messages is better than the similar exponent E_3 used for three messages, it should yield an overall better error exponent.

Trying to realize this idea, we immediately run into some pitfalls. With relatively high probability, the transmitter will fail to rank all three posterior probabilities $\mathsf{P}(w|\tilde{\mathbf{y}})$ correctly. In particular, with high probability, two least probable messages have approximately equal posterior probabilities, which can not be correctly distinguished from $\mathsf{P}(w|\tilde{\mathbf{y}})$ due to the feedback noise. Therefore, in our transmission strategy, when switching from stage I to stage II, we should not rely on the ranking of the message posterior probabilities alone.

Now we formally describe the transmission strategy. We introduce a parameter $0 \le t \le 1$. In the stage I of length λn, we use simplex signaling. And let d denote the (Euclidean) distance between each pair of codewords of their stage I, i.e., $d = \sqrt{3\lambda nP}$. Let $\mathbf{x}$ be the transmitted codeword (of length λn), $\mathbf{y}$ be the received sequence at the decoder, and $\tilde{\mathbf{y}}$ be the feedback sequence at the encoder.

Arrange the distance $d(\mathbf{x}_w, \mathbf{y})$, $w = 1, 2, 3$, in the increasing order, denoting

$$\min_w d(\mathbf{x}_w, \mathbf{y}) = d_1 \le d_2 \le d_3 = \max_w d(\mathbf{x}_w, \mathbf{y}),$$

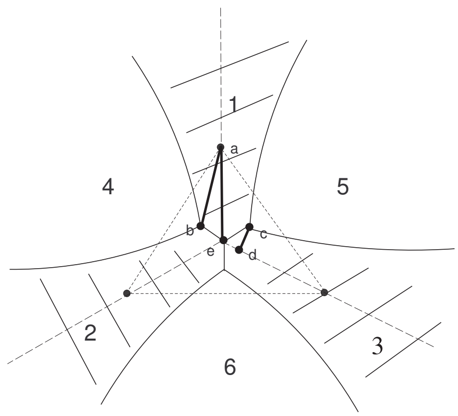

Fig. 2. Decision region.

(in case of equality, we use any order). We consider two cases.

Case I. If $d_3 \leq d_2 + td$, which means that the two least probable massages are of approximately equal posterior probabilities, then the receiver performs decoding immediately after stage I (in favor of the codeword closest to $\mathbf{y}$). Although the transmitter may continue its work, the receiver has already made its decision.

This case corresponds to the three shades areas $(1, 2, 3)$ in Fig. 2. Intuitively, since noisy feedback may "drag" the true signal far away enough so that the transmitter and the receiver will lose coordination, the shaded areas are introduced to "protect" the true signals.

Case II. If $d_3 > d_2 + td$, then after stage I the receiver considers the two most probable messages for the next stage. Now the transmitter selects its own candidates of the two most probable messages by observing the noisy feedback. Assuming that the receiver agrees on the same two messages (the order between them does not matter), the transmitter uses the antipodal signaling of length $(1-\lambda)n$ to help the receiver to distinguish between the two messages. After stage II, the receiver makes a decision using distances over *both* stages.

Note that to perform in coordination with the receiver, in case II it is important that the transmitter can correctly identify the two most probable messages for the receiver. Of course, an error in this choice is possible, but this probability must be sufficiently small (which will be denoted by P_3).

III. Analysis of the probability of error

Without loss of generality, assume that $w = 1$. Let P_1 denotes the decoding error probability and P_2 denote the decoding error probabilities in case I (after stage I) and case II (after stage II), respectively, for the *noise-free* feedback channel. Let P_3 to denote the probability of the union of the following two events:

1) The true codeword $\mathbf{x}_1$ is one of the two most probable codewords for the receiver, but is not so for the transmitter. For example, $\mathsf{P}(1|\mathbf{y}) > \mathsf{P}(2|\mathbf{y}) > \mathsf{P}(3|\mathbf{y})$, but $\mathsf{P}(2|\tilde{\mathbf{y}}) > \mathsf{P}(3|\tilde{\mathbf{y}}) > \mathsf{P}(1|\tilde{\mathbf{y}})$.
2) The true codeword $\mathbf{x}_1$ is one of the two most probable codewords, but the other candidate codeword for the receiver is different from the one at the transmitter. For example, $\mathsf{P}(1|\mathbf{y}) > \mathsf{P}(2|\mathbf{y}) > \mathsf{P}(3|\mathbf{y})$, but $\mathsf{P}(1|\tilde{\mathbf{y}}) > \mathsf{P}(3|\tilde{\mathbf{y}}) > \mathsf{P}(2|\tilde{\mathbf{y}})$.

Then, it can be checked that the total decoding error probability $P_e^{(n)}$ is upper bounded as

$$P_e^{(n)} \leq P_1 + P_2 + P_3. \tag{10}$$

To further upper bound each term in the right-hand side of (10), we need the following:

Lemma 1: If $t \leq 1 - \sqrt{3}/2 \approx 0.37$, then

$$t^k < \left(\frac{1}{2}\right)^{k-2} t^2 < \left(\frac{1}{2}\right)^{k-1} t,$$

where $k > 2$ is a positive integer.

Lemma 1 follows immediately from simple algebra and the given condition on t. This constraint has a simple geometric interpretation. As t decreases from 1 to 0, the unshaded areas in Figure 2 become closer to the triangular formed by the three message points. And P_1 starts to decrease until the "head" of the unshaded areas hits the triangular. The condition $t \leq 0.37$ corresponds to this critical point.

Henceforth, we assume that $t \leq 0.37$.

Since the three probabilities in (10) are determined by the minimum distance between some specific points and sets, the evaluation of them reduces to computing several distances and lower bounding them by using Lemma 1.

First, P_1 can be decomposed into two parts. Let P_{11} denotes the probability that the receiver will decode the received signal as message 2 or 3 immediately after stage I, i.e., the received signal lies in shaded area 2 or 3 in Fig. 2. P_{11} is determined by the minimum distance d' (between $\mathbf{x}_1$ and the set

$$\{\mathbf{y} : \|\mathbf{x}_2 - \mathbf{y}\| \leq \|\mathbf{x}_1 - \mathbf{y}\| \text{ and } \|\mathbf{x}_3 - \mathbf{y}\| \leq \|\mathbf{x}_1 - y\| + td\}.$$

which is the distance between a and b in Fig. 2. Let P_{12} denotes the probability that the received signal lies in the region that messages 2 and 3 are the most two probable messages for the receiver, i.e., the received signal lies in the triangular shaped area formed by area 6 and the lower halves of area 2 and 3. Then P_{12} is determined by the distance between points a and e in Fig. 2. Thus we have

$$P_1 \leq 2P_{11} + P_{12}.$$

However, since P_{11} dominates P_{12}, the main contribution to P_1 is given by P_{11}. By simple geometry, it can be readily checked that

$$\begin{aligned}(d')^2 = {} & \frac{d^2}{(3-4t^2)^2} \\ & \cdot (4t^6 - 4\sqrt{3}t^5 - t^4 + 6\sqrt{3}t^3 - 5t^2 - 2\sqrt{3}t + 3) \\ > {} & d^2\left(\frac{1}{3} - \frac{7+6\sqrt{3}}{24}t\right),\end{aligned}$$

where the inequality follows from Lemma 1. Thus, we have

$$P_1 < \Phi\left(-\frac{d_1(t,d)}{\sqrt{N}}\right)e^{o(n)},$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}\,dy$ and

$$d_1(t,d) := d\sqrt{\frac{8-(7+6\sqrt{3})t}{24}}.$$

Second, P_2 depends on the total of the distance between the simplex signaling (of radius $\sqrt{\lambda P n}$) in stage I and the distance between the antipodal signaling (of radius $\sqrt{(1-\lambda)Pn}$) in stage II. By simple algebra, we have

$$P_2 \le \Phi\left(-\sqrt{\left(1-\frac{1}{4}\lambda\right)n\frac{P}{N}}\right)e^{o(n)}.$$

Third, P_3 is determined by the minimum distance d'' between

$$\begin{aligned} &S_1 = \\ &\{\mathbf{y} : \|\mathbf{x}_1-\mathbf{y}\| \le \|\mathbf{x}_3-\mathbf{y}\| \le \|\mathbf{x}_3-\mathbf{y}\| + td \le \|\mathbf{x}_2-\mathbf{y}\|, \\ &\quad \|\mathbf{x}_3-\mathbf{y}\| \le \|\mathbf{x}_1-\mathbf{y}\| \le \|\mathbf{x}_1-\mathbf{y}\| + td \le \|\mathbf{x}_2-\mathbf{y}\|\} \end{aligned}$$

and

$$S_2 = \{\tilde{\mathbf{y}} : \|\mathbf{x}_2-\tilde{\mathbf{y}}\| \le \|\mathbf{x}_3-\tilde{\mathbf{y}}\| \text{ or } \|\mathbf{x}_2-\tilde{\mathbf{y}}\| \le \|\mathbf{x}_1-\tilde{\mathbf{y}}\|\},$$

which is the distance between c and d in Fig. 2, where S_1 corresponds to the event that the codeword $\mathbf{x}_2$ is not one of the two most probable codewords for the receiver, and S_2 corresponds to the event that the codeword $\mathbf{x}_2$ is one of the two most probable codewords for the transmitter. Again it can be readily checked that

$$\begin{aligned} (d'')^2 &= \frac{d^2}{4(3-4t^2)^2}(12t^6 - 4\sqrt{3}t^5 - 23t^4 + 4\sqrt{3}t^3 + 12t^2) \\ &> t^2 d^2 \frac{25-2\sqrt{3}}{144}. \end{aligned}$$

Thus

$$P_3 < \Phi\left(-\frac{d_2(t,d)}{\sqrt{\alpha N}}\right)e^{o(n)},$$

where

$$d_2(t,d) := \frac{td}{12}\sqrt{25-2\sqrt{3}}.$$

Combining the three terms and substituting $d = \sqrt{3\lambda n P}$, d_1, and d_2, we have

$$\begin{aligned} &\tilde{F}_3(\alpha) \\ &\ge \lim_{n\to\infty} -\frac{1}{n}\ln P_e^{(n)} \\ &\ge \lim_{n\to\infty} -\frac{1}{n}\max\{\ln P_1, \ln P_2, \ln P_3\} \\ &\ge \min\left\{\frac{8-(7+6\sqrt{3})t}{16}\lambda,\ \frac{1}{2}-\frac{\lambda}{8},\ \frac{(25-2\sqrt{3})t^2}{96\alpha}\lambda\right\}\frac{P}{N}. \end{aligned}$$

By choosing t and λ such that

$$\alpha = \alpha(t) := \frac{(25-2\sqrt{3})t^2}{6(8-(7+6\sqrt{3})t)}$$

and

$$\lambda = \lambda(t) := \frac{8}{10-(7+6\sqrt{3})t},$$

we can make the three terms in the minimum the same. Therefore,

$$F_3(\alpha(t)) \ge \frac{1}{2}\left(1-\frac{2}{10-(7+6\sqrt{3})t}\right)\frac{P}{N}.$$

Note that $\alpha(t)$ is monotonically increasing if

$$0 < t \le \frac{8}{7+6\sqrt{3}} \approx 0.46,$$

which is satisfied under our standing assumption $0 < t \le 0.37$. Hence,

$$F_3(\alpha) \ge \frac{1}{2}\left(1-\frac{2}{10-(7+6\sqrt{3})t(\alpha)}\right)\frac{P}{N}, \qquad (11)$$

where $t(\alpha)$ is the unique t such that $\alpha = \alpha(t)$.

Finally, we can compare (11) and $E_3 = (3/8)(P/N)$ to conclude that

$$F_3(\alpha) > E_3,$$

if

$$\alpha \le \frac{25-2\sqrt{3}}{9(7+6\sqrt{3})^2} \approx 8\times 10^{-3}.$$

This completes the proof of Theorem 1.

Note that

$$F_3 = \frac{1}{2}\frac{P}{N} > \tilde{F}_3(0) = \frac{2}{5}\frac{P}{N} > E_3 = \frac{3}{8}\frac{P}{N}.$$

Hence, our coding strategy is strictly suboptimal for noise-free feedback.

It remains an open problem to find a coding strategy that performs gracefully over all values of α and is (asymptotically) optimal for the noise-free case.

References

[1] M. V. Burnashev and H. Yamamoto, "On BSC, noisy feedback and three messages," *Proc. IEEE International Symposium on Information Theory, Toronto, Canada*, pp. 886 - 889, July 2008.

[2] M. V. Burnashev and H. Yamamoto, "On zero-rate error exponent for BSC with noisy feedback," *Problems of Information Transmission*, vol. 44, no. 3, pp. 33 - 49, 2008.

[3] 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. on Inform. Theory*, vol. 15, pp. 476C478, July 1969.

[4] Y.-H. Kim, A. Lapidoth and T. Weissman, "The Gaussian channel with noisy feedback," *Proc. IEEE International Symposium on Information Theory, Nice, France*, pp. 1416 - 1420, June 2007.

[5] M.S. Pinsker, "The probability of error in block transmission in a memoryless gaussian channel with feedback," *Problems of Information Transmission*, vol. 4, no. 4, pp. 3 - 19, 1968.

[6] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback," *IEEE Trans. Inform. Theory*, vol. 12, no. 3, pp. 172 - 182, 1966.

[7] A. D. Wyner. On the Schalkwiik-Kailath coding scheme with a peak ene;gy constraint," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 129 - 134, Jan. 1968.

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