

# On the Reliability of Gaussian Channels with Noisy Feedback

Young-Han Kim, Amos Lapidoth, and Tsachy Weissman

**Abstract**— We derive bounds on the reliability of the additive white Gaussian noise channel  $Y = X + Z$  with output fed back to the transmitter over independent additive white Gaussian noise channel  $\tilde{Y} = Y + \tilde{Z}$ . These bounds appear to be new even for transmission at zero rate. As a corollary, it is shown that linear feedback coding schemes of finite-dimensional constellation, such as the celebrated Schalkwijk–Kailath coding scheme, not only fail to achieve the capacity, but in fact cannot achieve any positive rate. Our approach is applicable to the derivation of upper bounds on the error exponents in various other scenarios involving channels with feedback.

## I. INTRODUCTION

That perfect feedback can dramatically improve the reliability of a memoryless channel is pointed out by Shannon in [18]. For the additive white Gaussian noise channel, this fact was reconfirmed by Schalkwijk–Kailath [17] in the strong sense that perfect feedback allows for schemes under which the probability of error diminishes double-exponentially fast in block length at any rate below capacity. This double exponential decay has been further extended to the diminishing probability of any number of nested exponential levels by Pinsker [14], Kramer [8], and Zigangirov [19].

Much less explored and understood is how noise in the feedback link affects the achievable reliability (cf. [6], [15], [11] for some recent progress). In this paper we restrict attention, for concreteness, to the additive white Gaussian channel  $Y = X + Z$  with additive white Gaussian noise corrupted feedback  $\tilde{Y} = Y + \tilde{Z}$ , and derive bounds on the reliability function.

The paper is organized as follows. After describing the problem setup in Section II, we review the case of perfect feedback in Section III. In subsequent two sections, we derive upper bounds on the reliability using two different methods. The first method, based on a change-of-measure argument, leads to a family of upper bounds on the reliability in Section IV. Section V presents another upper bound on the same exponent via a ‘genie-aided’ argument. For lower bounds, we develop a simple binary messaging scheme in Section VI and then show that true zero rate can be achieved by a concatenated code based on the Schalkwijk–Kailath feedback coding scheme with error exponent blowing up as the feedback noise variance tends to zero. For positive rate, however, we show that any linear encoding scheme not only

fails to achieve capacity, but fails to reliably communicate at any positive rate. We conclude in Section VIII with a description of the way our bounds extend to the non-Gaussian and to the discrete setting.

Throughout  $\log$  will denote the natural logarithm, and capacity and rate will be in nats per channel use.

## II. CHANNEL MODEL

We wish to communicate a message index  $W \in \{1, \dots, e^{nR}\}$  over the additive white Gaussian noise channel

$$Y_i = X_i + Z_i, \quad (1)$$

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

$$\tilde{Y}_i = Y_i + \tilde{Z}_i, \quad (2)$$

where  $\tilde{Z}_i$  is the Gaussian noise in the backward link. We define a  $(e^{nR}, n)$  code with the encoding function of the form

$$X_i = X_i(W, \tilde{Y}^{i-1}) \quad (3)$$

under the expected average power constraint

$$\frac{1}{n} \sum_{i=1}^n EX_i^2(W; \tilde{Y}^{i-1}) \leq P,$$

and the decoding function  $\hat{W}(Y^n) \in \{1, \dots, e^{nR}\}$ .

Thus, the encoder has the causal access to the noisy feedback. Equivalently, we can consider the encoding function to be of the form

$$X_i(W, S^{i-1}), \quad (4)$$

where

$$S_i = Z_i + \tilde{Z}_i, \quad (5)$$

since, given  $X^i$ , there is a one-to-one transformation from  $\tilde{Y}^i$  to  $S^i$ . The forward and backward noise processes,  $\{Z_i\}$  and  $\{\tilde{Z}_i\}$ , are independent of each other, and independent and identically distributed over time, with respective variances

$$\begin{aligned} Z_i &\sim \mathcal{N}(0, 1) \\ S_i &\sim \mathcal{N}(0, \varepsilon^2). \end{aligned} \quad (6)$$

The probability of error  $P_e^{(n)}$  is defined by

$$\begin{aligned} P_e^{(n)} &= \Pr(W \neq \hat{W}(Y^n)) \\ &= \frac{1}{e^{nR}} \sum_{w=1}^{e^{nR}} \Pr(W \neq \hat{W}(Y^n) | W = w) \end{aligned}$$

with  $W$  and  $(Z^n, S^n)$  independent of each other.

Young-Han Kim is with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093, USA yhk@ucsd.edu

Tsachy Weissman is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA tsachy@stanford.edu

Amos Lapidoth is with the Department of Electrical Engineering, Swiss Federal Institute of Technology (ETH), Zurich, CH-8092, Switzerland lapidoth@isi.ee.ethz.ch

Finally, let  $E_{FB}(R) = E_{FB}(R; P, \varepsilon^2)$  denote the reliability function associated with this communication problem at rate  $R$ , power constraint  $P$ , and the feedback noise variance  $\varepsilon^2$ . As usual, the reliability function is defined as the rate of decay for error probability of the optimal sequence of  $(e^{nR}, n)$  codes, i.e.,

$$E_{FB}(R) = \limsup_{n \rightarrow \infty} -\frac{1}{n} \log P_{e, \text{opt}}^{(n)}(R).$$

We will use the notation  $E(R; P)$  to denote the reliability of the additive white Gaussian noise channel without feedback under signal-to-noise ratio  $P$ .

### III. PERFECT FEEDBACK

Here we give a brief review of the case when the feedback noise  $\hat{Z}$  is constant, or equivalently,  $\varepsilon^2 = 0$ . Our analysis is purely information theoretic and thus very simple, compared to the original analysis of the Schalkwijk [16] and the later one by Butman [2], [3].

Let  $\theta$  be one of  $e^{nR}$  equally spaced real numbers on some interval, say,  $[-1, 1]$ , with distance  $2\Delta$  between nearest neighbors.

Initially, the transmitter sends

$$X_0 = \theta.$$

With noiseless feedback, the channel input  $X_i$  can be an arbitrary function of the previous output values  $Y_0^{i-1} = (Y_0, \dots, Y_{i-1})$ . Hence, the transmitter can subsequently send

$$X_1 = \alpha_1(Y_0 - X_0) = \alpha_1 Z_0$$

and

$$X_i = \alpha_i(Z_0 - \hat{Z}_0(Y_1^{i-1})), \quad i = 2, 3, \dots,$$

where  $\hat{Z}_0(Y_1^{i-1})$  is the minimum mean square error (MMSE) estimate of  $Z_0$  given  $(Y_1, \dots, Y_{i-1})$ , and  $\{\alpha_i\}_{i=1}^{\infty}$  are chosen to satisfy the power constraint  $EX_i^2 = P$  for each  $i$ .

From the orthogonality property of the MMSE estimate and the joint Gaussianity, it is easy to see that  $(X_i, Z_i, Y_i)$  are independent of  $Y^{i-1}$ . Furthermore, from our choice of the scaling factor  $\alpha_i$ ,  $\{Y_i\}_{i=1}^{\infty}$  is i.i.d.  $\sim \mathcal{N}(0, P+1)$ .

Now we look at the mutual information  $I(Z_0; Y_1^n)$ . On one hand, we have

$$\begin{aligned} I(Z_0; Y_1^n) &= h(Y_1^n) - h(Y_1^n | Z_0) \\ &= \sum_{i=1}^n h(Y_i | Y_1^{i-1}) - h(Y_i | Z_0, Y_1^{i-1}) \\ &= \sum_{i=1}^n h(Y_i) - h(X_i + Z_i | Z_0, Y_1^{i-1}) \\ &= \sum_{i=1}^n h(Y_i) - h(Z_i | Z_0, Y_1^{i-1}) \\ &= \sum_{i=1}^n h(Y_i) - h(Z_i) \\ &= \frac{n}{2} \log(1+P) \\ &= nC. \end{aligned} \tag{7}$$

On the other hand, we have

$$I(Z_0; Y_1^n) = h(Z_0) - h(Z_0 | Y_1^n) = \frac{1}{2} \log \frac{1}{\text{var}(Z_0 | Y_1^n)},$$

which combined with (7) implies that

$$\text{var}(Z_0 | Y_1^n) = e^{-2nC}.$$

Finally, upon receiving  $(Y_0, \dots, Y_n)$ , the receiver forms the maximum likelihood estimate  $\hat{\theta}_n$  of  $\theta$  as

$$\begin{aligned} \hat{\theta}_n &= Y_0 - \hat{Z}_0(Y_1^n) \\ &= \theta + Z_0 - \hat{Z}_0(Y_1^n) \\ &\sim \mathcal{N}(\theta, e^{-2nC}) \end{aligned}$$

and performs the nearest neighbor decoding.

It is easy to see that the decoding error happens only if  $\hat{\theta}_n$  is closer to the nearest neighbors of the true  $\theta$ , that is, if  $\mathcal{N}(0, e^{-2nC}) > \Delta$ . Since

$$\Delta = c_0 e^{-nR}$$

with  $c_0$  being a fixed constant depending only on the message constellation, the probability of error is given by

$$P_e^{(n)} = 2 \left( 1 - Q \left( c_0 e^{n(C-R)} \right) \right)$$

where

$$Q(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$$

is the standard Gaussian cumulative density function. Therefore, if  $R < C$ , we have

$$P_e^{(n)} \leq 2 \exp\left(-\frac{c_0^2 e^{2n(C-R)}}{2}\right),$$

that is, the probability of error decays doubly exponentially fast in block size  $n$ . In particular, this implies

$$E_{FB}(R; P, \varepsilon^2 = 0) = \infty$$

for all  $R < C$ .

### IV. UPPER BOUNDS VIA CHANGE OF MEASURE

In this section, we derive upper bounds on the reliability  $E_{FB}(R; P, \varepsilon^2)$  via a change-of-measure method. The idea is to change the joint law of the noises in the forward and the backward links into one under which the noisy feedback is useless. An upper bound on the error exponent of interest is then given by the error exponent under the latter law (which is a classical channel coding error exponent) plus an additional ‘penalty’ term stemming from the change of measure.

Let  $(Z, S)$  be a generic pair of random variables distributed as the pair  $(Z_i, S_i)$  of (5) and (6), and we let  $(Z', S)$  be a pair of *independent* Gaussians with  $Z' \sim \mathcal{N}(0, \sigma^2)$  and  $S' \sim \mathcal{N}(0, 1 + \varepsilon^2)$ . Let further  $f$  and  $f'$  denote the respective densities of  $(Z, S)$  and  $(Z', S')$ . Finally, let  $\Lambda_{\varepsilon^2, \sigma^2}^*$  denote the Fenchel-Legendre transform (see, e.g., [5]) of the random variable

$$\log \frac{f'(Z', S')}{f(Z', S')}.$$

We now present the main result of this section. The proof involves changing the law of the noise components in the original channel from  $(Z_i, S_i) \stackrel{d}{=} (S, Z)$  to  $(S_i, Z_i) \stackrel{d}{=} (Z', S')$ . The feedback in the latter case is useless, so the associated error exponent is  $E(R; P/\sigma^2)$ .

**Theorem 1.** For every  $\sigma^2 > 0$ ,

$$E_{\text{FB}}(P, \varepsilon^2, R) \leq \gamma + E_{\text{NoFB}}(P/\sigma^2, R) \quad (8)$$

where  $\gamma$  uniquely solves

$$\Lambda_{\varepsilon^2, \sigma^2}^*(\gamma) = E_{\text{NoFB}}(P/\sigma^2, R)$$

in the region

$$\begin{aligned} \gamma &\geq D(f' \| f) \\ &= \frac{1}{2} \left[ \frac{\sigma^2 - \varepsilon^2 + \sigma^2 \varepsilon^2 + 1}{\varepsilon^2} - \log[\sigma^2 \varepsilon^2 (1 + \varepsilon^2)] \right]. \end{aligned}$$

Note that this theorem gives a family of bounds, indexed by  $\sigma^2$ , which should be viewed as a parameter to be optimized over for a given value of  $R$ .

The Fenchel-Legendre transform  $\Lambda_{\varepsilon^2, \sigma^2}^*$  can be explicitly obtained. Towards this end note that

$$f(z, s) = \frac{\varepsilon}{2\pi} \exp \left\{ -\frac{(1 + \varepsilon^2)z^2 - 2zs + s^2}{2\varepsilon^2} \right\} \quad (9)$$

and

$$f'(z, s) = \frac{1}{2\pi\sqrt{\sigma^2(1 + \varepsilon^2)}} \exp \left\{ -\frac{1}{2} \left[ \frac{z^2}{\sigma^2} + \frac{s^2}{1 + \varepsilon^2} \right] \right\} \quad (10)$$

so that

$$\begin{aligned} \log \frac{f'(Z', S')}{f(Z', S')} &= -\frac{1}{2} \log[\sigma^2 \varepsilon^2 (1 + \varepsilon^2)] \\ &\quad + \frac{1}{2} \left[ \left( \frac{Z'}{\sigma} \right)^2 \frac{\sigma^2 - \varepsilon^2 + \sigma^2 \varepsilon^2}{\varepsilon^2} - \frac{2\sigma\sqrt{1 + \varepsilon^2}}{\varepsilon^2} \right. \\ &\quad \left. \frac{Z'}{\sigma} \frac{S'}{\sqrt{1 + \varepsilon^2}} + \left( \frac{S'}{\sqrt{1 + \varepsilon^2}} \right)^2 \frac{1}{\varepsilon^2} \right]. \end{aligned} \quad (11)$$

Since both  $\frac{Z'}{\sigma}$  and  $\frac{S'}{\sqrt{1 + \varepsilon^2}}$  are standard Gaussian random variables, it is evident from (11) that

$$\Lambda_{\varepsilon^2, \sigma^2}^*(\gamma) = \tilde{\Lambda}_{\varepsilon^2, \sigma^2}^*(2\gamma + \log[\sigma^2 \varepsilon^2 (1 + \varepsilon^2)]), \quad (12)$$

where  $\tilde{\Lambda}_{\varepsilon^2, \sigma^2}^*$  is the Fenchel-Legendre transform of

$$U^2 \frac{\sigma^2 + \sigma^2 \varepsilon^2 - \varepsilon^2}{\varepsilon^2} - \frac{2\sigma\sqrt{1 + \varepsilon^2}}{\varepsilon^2} UV + V^2 \frac{1}{\varepsilon^2}, \quad (13)$$

with independent standard Gaussian random variables  $U$  and  $V$ . Fortunately,  $\tilde{\Lambda}_{\varepsilon^2, \sigma^2}^*$  can be derived in closed form

$$\begin{aligned} \tilde{\Lambda}_{\varepsilon^2, \sigma^2}^*(\alpha) &= \frac{1}{4} (-2 - \alpha(1 + \sigma^2 + \varepsilon^2(\sigma^2 - 1))) \\ &\quad + \sqrt{4 + \alpha^2(\varepsilon^4(\sigma^2 - 1)^2 + (\sigma^2 + 1)^2 + 2\varepsilon^2(1 + \sigma^4))} + \frac{1}{2} \times \end{aligned}$$

$$\log \frac{-2 + \sqrt{4 + \alpha^2(\varepsilon^4(1 - \sigma^2)^2 + (1 + \sigma^2)^2 + 2\varepsilon^2(1 + \sigma^4))}}{\alpha^2 \varepsilon^2} \quad (14)$$

which, taken with (12), gives the explicit form of  $\Lambda_{\varepsilon^2, \sigma^2}^*(\gamma)$ .

While we have obtained  $\Lambda_{\varepsilon^2, \sigma^2}^*(\gamma)$  explicitly, solving for the  $\gamma$  that satisfies  $\Lambda_{\varepsilon^2, \sigma^2}^*(\gamma) = E_{\text{NoFB}}(P/\sigma^2, R)$ , namely the inverse function of  $\Lambda_{\varepsilon^2, \sigma^2}^*$ , is elusive. It is therefore useful to express the bound of Theorem 1 in parametric form. Towards this end, let  $R(s, e)$  denote the ‘‘rate-reliability’’ function of the AWGN channel without feedback under signal-to-noise ratio  $P$  and error exponent  $E$ , i.e.,

$$\begin{aligned} R(E; P) &= \\ &\begin{cases} R \text{ such that } E(R; P) = E, & \text{for } 0 \leq E < E(0; P), \\ 0, & \text{for } E \geq E(0; P). \end{cases} \end{aligned} \quad (15)$$

Note, in particular, that

$$R(0; P) = \frac{1}{2} \log(1 + P). \quad (16)$$

Theorem 1 can be stated equivalently as follows.

**Theorem 2.** For every  $\sigma^2 > 0$ ,  $E_{\text{FB}}(R; P, \varepsilon^2)$  is upper bounded by the following curve, which is given in parametric form by

$$[R(\Lambda_{\varepsilon^2, \sigma^2}^*(\gamma); P/\sigma^2), \Lambda_{\varepsilon^2, \sigma^2}^*(\gamma) + \gamma] \quad (17)$$

where  $\gamma$  varies in the interval

$$\left[ \frac{1}{2} \left[ \frac{\sigma^2 - \varepsilon^2 + \sigma^2 \varepsilon^2 + 1}{\varepsilon^2} - \log[\sigma^2 \varepsilon^2 (1 + \varepsilon^2)] \right], \gamma_{\max, \sigma^2} \right], \quad (18)$$

and  $\gamma_{\max, \sigma^2}$  is the value of  $\gamma$  for which  $\Lambda_{\varepsilon^2, \sigma^2}^*(\gamma) = E(0; P/\sigma^2)$ .

Although the functions  $E(\cdot; P)$  and  $R(\cdot; P)$  are not explicitly known, bounds on these functions can be combined with the theorems above to obtain concrete bounds, as illustrated in the following corollaries.

**Corollary 1.** For every  $\sigma^2 > 0$ , the following curve, in parametric form, is an upper bound to the curve of  $E_{\text{FB}}(\cdot; P, \varepsilon^2)$

$$\left( \frac{1}{2} \left[ 1 - \frac{2\Lambda_{\varepsilon^2, \sigma^2}^*(\gamma)}{(P/\sigma^2)} \right] \log(1 + (P/\sigma^2)), \Lambda_{\varepsilon^2, \sigma^2}^*(\gamma) + \gamma \right) \quad (19)$$

where  $\gamma$  varies in the range given in (18) and  $\gamma'_{\max, \sigma^2}$  is the value of  $\gamma$  for which  $\Lambda_{\varepsilon^2, \sigma^2}^*(\gamma) = \frac{P}{2\sigma^2}$ .

*Proof.* The corollary follows by further bounding the bound in Theorem 2 using the sphere packing bound [1]:

$$E(R; P) \leq E_{\text{SP}}(R; P) \leq \frac{P}{2} \left[ 1 - \frac{R}{\frac{1}{2} \log(1 + P)} \right], \quad (20)$$

which also implies

$$R(E; P) \leq \frac{1}{2} \log(1 + P) \left[ 1 - \frac{2E}{P} \right]. \quad (21) \quad \square$$

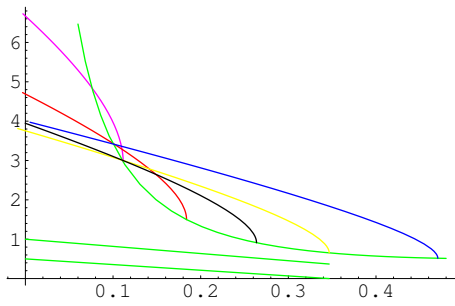


Fig. 1. Upper bounds on  $E_{FB}(1, 1, R)$ . The purple, red, black, yellow, and blue curves are the parametric bounds of Corollary 1 for the respective values of  $\sigma$ : 2, 1.5, 1.2, 1, and 0.8. The green curve is the bound of Corollary 3. The green line (higher of the two) is the straight line bound on the sphere packing bound (inequality (20)), applied to the bound of Proposition 1. The lower green line, for comparison, is the straight line bound on the sphere packing bound of the same channel with no feedback. This line intersects the  $x$ -axis at the capacity of this channel,  $\frac{1}{2} \log 2 \approx 0.347$  nats. Thus, the region where the bounds can be useful is  $0 \leq R \leq \frac{1}{2} \log 2$ .

Figure 1 shows the curves of Corollary 1, for the case of  $\varepsilon^2 = 1$ , and different values of  $\sigma^2$ . It is seen that there is no one value which dominates for all values of  $R$ , although there are values that are dominated for all values of  $R$  at the relevant range  $[0, C]$  (blue curve, corresponding to  $\sigma = 0.8$ ).

Another immediate consequence of Theorem 1 and (20) is the following result.

**Corollary 2.**

$$E(0; P, \varepsilon^2) \leq \frac{P}{2\sigma^2} + \gamma, \quad (22)$$

for any  $\sigma^2 > 0$  and  $\gamma$  satisfying

$$\Lambda_{\varepsilon^2, \sigma^2}^*(\gamma) \geq \frac{P}{2\sigma^2} \quad (23)$$

and

$$\gamma \geq \frac{\sigma^2 + \sigma^2 \varepsilon^2 - \varepsilon^2 + 1}{2\varepsilon^2} - \frac{1}{2} \log[\sigma^2 \varepsilon^2 (1 + \varepsilon^2)].$$

For another working point, note that by choosing  $\sigma^2 = P/(e^{2R} - 1)$  (smallest value of  $\sigma^2$  for which  $E(R; P/\sigma^2) = 0$ ) in Theorem 1, we obtain:

**Corollary 3.** For any  $R > 0$ ,

$$E(R; P, \varepsilon^2) \leq \frac{\frac{P}{e^{2R}-1}(1 + \varepsilon^2) - \varepsilon^2 + 1}{2\varepsilon^2} - \frac{1}{2} \log \frac{P\varepsilon^2(1 + \varepsilon^2)}{e^{2R} - 1}. \quad (24)$$

The bound of Corollary 3, for the case  $\varepsilon^2 = 1$ , is plotted in Figure 1 (green curve). Note that, as it should, this curve passes through the endpoints of the curves of Corollary 1. We conclude this section with the proof of Theorem 1.

*Proof of Theorem 1.* Fix a particular coding scheme, for the setting of (1)–(6), operating at average power upper bounded by  $P$ . Let  $P_{e|m}$  denote its probability of error when

transmitting the message  $m$ . In other words, for block length  $n$ ,

$$P_{e|W=w} = \Pr((Z^n, S^n) \in A_w), \quad (25)$$

$A_w$  denoting the error set

$$A_w = \{(z^n, s^n) : \hat{W}(y^n) \neq w\}, \quad (26)$$

where  $\hat{W}(y^n)$  in (26) denotes the decoder estimate under the fixed coding scheme and the realized values  $(z^n, s^n)$  when encoding for the message  $m$ .

Consider now the following scenario of communications with useless feedback:

$$Y'_i = X'_i + Z'_i, \quad X'_i = X'_i(w, S'^{i-1}), \quad (27)$$

where  $\{Z'_i\}$  and  $\{S'_i\}$  are independent white noises with

$$Z'_i \sim \mathcal{N}(0, \sigma^2), \quad S'_i \sim \mathcal{N}(0, 1 + \varepsilon^2). \quad (28)$$

Let  $P'_{e|w}$  denote the probability of error of the coding scheme associated with  $P_{e|w}$ , when operating in the useless feedback setting. Thus, denoting

$$B_\gamma = \left\{ (z^n, s^n) : \frac{1}{n} \log \frac{f'(z^n, s^n)}{f(z^n, s^n)} \leq \gamma \right\}, \quad (29)$$

we have

$$\begin{aligned} P'_{e|w} &= \Pr((Z'^n, S'^n) \in A_w) \\ &= \int_{A_w} f'(z^n, s^n) dz^n ds^n \\ &= \int_{A_w \cap B_\gamma} f'(z^n, s^n) dz^n ds^n \\ &\quad + \int_{A_w \cap B_\gamma^c} f'(z^n, s^n) dz^n ds^n \\ &\leq e^{n\gamma} \int_{A_w \cap B_\gamma} f(z^n, s^n) dz^n ds^n \\ &\quad + \int_{A_w \cap B_\gamma^c} f'(z^n, s^n) dz^n ds^n \\ &\leq e^{n\gamma} \int_{A_w} f(z^n, s^n) dz^n ds^n \\ &\quad + \int_{B_\gamma^c} f'(z^n, s^n) dz^n ds^n \\ &= e^{n\gamma} \Pr((Z^n, S^n) \in A_w) + \Pr((Z'^n, S'^n) \in B_\gamma^c) \\ &= e^{n\gamma} P_{e|w} + \Pr((Z'^n, S'^n) \in B_\gamma^c). \end{aligned} \quad (30)$$

Averaging the two sides of (30) over  $w$  gives

$$P'_e \leq e^{n\gamma} P_e + \Pr((Z'^n, S'^n) \in B_\gamma^c) \quad (31)$$

or, equivalently,

$$e^{-n\gamma} P'_e - e^{-n\gamma} \Pr((Z'^n, S'^n) \in B_\gamma^c) \leq P_e. \quad (32)$$

Assuming  $\gamma$  was chosen such that  $\Pr((Z'^n, S'^n) \in B_\gamma^c) < P'_e$ , (32) is, in turn, equivalent to

$$-\frac{1}{n} \log P_e \leq \gamma - \frac{1}{n} \log [P'_e - \Pr((Z'^n, S'^n) \in B_\gamma^c)] \quad (33)$$

implying

$$-\frac{1}{n} \log P_e \leq \gamma - \frac{1}{n} \log [P_{\min}^m(R; P, \sigma^2) - \Pr((Z^m, S^m) \in B_\gamma^c)], \quad (34)$$

where  $P_{\min}^m(R; P, \sigma^2)$  denotes the minimum probability of error achievable with block-length  $l$  in the useless feedback setting of (27) and (28).<sup>1</sup> Since, by definition of  $E(R; P)$ ,

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log P_{\min}^m(R; P, \sigma^2) = E(R; P/\sigma^2), \quad (35)$$

it follows that

$$E_{FB}(R; P, \varepsilon^2) \leq \gamma + E(R; P/\sigma^2) \quad (36)$$

for  $\gamma$  sufficiently large that

$$\liminf_{n \rightarrow \infty} -\frac{1}{n} \log \Pr((Z^m, S^m) \in B_\gamma^c) > E(R; P/\sigma^2). \quad (37)$$

The  $\liminf$  in (8), however, is in fact a limit we can explicitly characterize. Indeed, by the definition of  $B_\gamma$  in (29),

$$\Pr((Z^m, S^m) \in B_\gamma^c) \quad (38)$$

$$= \Pr\left(\frac{1}{n} \log \frac{f'(Z^m, S^m)}{f(Z^m, S^m)} > \gamma\right) \quad (39)$$

$$= \Pr\left(\frac{1}{n} \sum_{i=1}^n \log \frac{f'(Z'_i, S'_i)}{f(Z'_i, S'_i)} > \gamma\right) \quad (40)$$

thus, for  $\gamma \geq E\left[\log \frac{f'(Z', S')}{f(Z', S')}\right] = D(f' \| f)$ , Cramér's theorem (cf. [5, Th. 2.2.3]) implies

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \Pr((Z^m, S^m) \in B_\gamma^c) = \Lambda^*(\gamma), \quad (41)$$

where  $\Lambda^*$  is the Fenchel-Legendre transform of the random variable  $\log \frac{f'(Z', S')}{f(Z', S')}$ . It follows by substitution of (41) in (37) that (8) holds for all  $\gamma \geq D(f' \| f)$  satisfying  $\Lambda^*(\gamma) > E(R; P/\sigma^2)$  and, consequently, by the continuity and strict monotonicity of  $\Lambda^*(\gamma)$ , for the value of  $\gamma$  satisfying  $\Lambda^*(\gamma) = E(R; P/\sigma^2)$ .  $\square$

### V. UPPER BOUND VIA GENIE

Consider a genie-aided scheme where encoding is allowed to depend on the  $S_i$  sequence non-causally, i.e., to be of the form  $X_i = X_i(W, S^n)$  instead of  $X_i(W, S^{i-1})$ . Assume further that the decoder is also given access to  $S^n$  in addition to  $Y^n$ , i.e.,  $\hat{W} = \hat{W}(Y^n, S^n)$ . By conditioning on  $S^n$  we then see that the capacity and error exponent for this setting is exactly that for the standard additive white Gaussian noise channel with no feedback and noise variance equals to  $\text{Var}(Z_i|S_i) = \text{Var}(Z_i|Z_i + \tilde{Z}_i) = \frac{\varepsilon^2}{\varepsilon^2 + 1}$ . Of course, the capacity and error exponent for this problem upper bound those of our problem, since here encoder and decoder are supplied with more information. Therefore we have the following:

<sup>1</sup>Note that we have used here the fact that  $S^n \stackrel{d}{=} S'^n$ , which implies that the power used by the scheme in the useless feedback setting is identical to that used in the original setting.

### Proposition 1.

$$E_{FB}(R; P, \varepsilon^2) \leq E\left(R; \frac{P(\varepsilon^2 + 1)}{\varepsilon^2}\right). \quad (42)$$

Simple as the argument leading to it may be, the bound of Proposition 1 is, in many cases, tighter than those of the previous section (see Figure 1 for a comparison in the case  $\varepsilon^2 = 1$ ). Furthermore, the bound allows us to conclude that the noisy feedback (at least insofar as the fundamental limits go) can be no more useful than having the power increase by  $P/\varepsilon^2$  in the absence of feedback. Furthermore, when combined with the sphere packing bound on  $E_{\text{NoFB}}$ , Proposition 1 gives

$$E_{FB}(R; P, \varepsilon^2) \leq \frac{P \varepsilon^2 + 1}{2 \varepsilon^2} \quad (43)$$

implying that  $E_{FB}(R; P, \varepsilon^2)$  increases with small  $\varepsilon$  essentially no faster than  $\frac{P}{2\varepsilon^2}$ . The following section shows that this bound is very tight at  $R = 0$ .

Finally, we note that the bound of Proposition 1 can potentially be tightened by denying the decoder of the genie-aided scheme access to  $S^n$ . The non-feedback reliability on the right-hand side of (42) can thus be replaced by the error exponent of the corresponding dirty paper problem [4], [10]. Unfortunately, it is as yet unknown whether the latter is strictly smaller than the standard nonfeedback reliability.

### VI. ZERO RATE AND SMALL FEEDBACK NOISE

In this section we consider the asymptotic regime of zero rate (i.e.,  $R \rightarrow 0$ ) and feedback noise of very small variance (i.e.,  $\varepsilon^2 \rightarrow 0$ ). Specializing inequality (43) to this regime yields

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon^2 \cdot E_{FB}(R; P, \varepsilon^2) \leq \frac{P}{2} \quad (44)$$

for all  $R$ .

We shall first show that this upper bound is achievable in the two-message setting. Denoting the two-message reliability by  $E_{FB, \text{binary}}(P, \varepsilon^2)$ , we shall thus show that

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon^2 \cdot E_{FB, \text{binary}}(P, \varepsilon^2) \geq \frac{P}{2}. \quad (45)$$

Towards this end, we consider the following scheme for communicating one bit, '0' or '1'. Suppose we wish to communicate '1'. Then we take the channel input as

$$X_i = a - S_{i-1} = a - Z_{i-1} - \tilde{Z}_{i-1} \quad (46)$$

for some positive  $a$  whose value is to be determined. The channel output will then be

$$Y_i = X_i + Z_i = a - Z_{i-1} + Z_i - \tilde{Z}_{i-1}. \quad (47)$$

If '0' is to be communicated, then we use  $-a$  instead of  $a$ . The decoder declares that '1' was sent if  $\sum_{i=1}^n Y_i > 0$ ; otherwise, it decides on '0'. Thus, conditioned on '1' being sent,

$$\begin{aligned} \sum_{i=1}^n Y_i &= na + Z_n - \sum_{i=1}^{n-1} \tilde{Z}_i \\ &\sim \mathcal{N}(na, 1 + (n-1)\varepsilon^2), \end{aligned} \quad (48)$$

which implies

$$\begin{aligned} P_e^{(n)} &= P\left(\frac{\sum_{i=1}^n Y_i - na}{\sqrt{1 + (n-1)\varepsilon^2}} \leq \frac{-na}{\sqrt{1 + (n-1)\varepsilon^2}}\right) \\ &= Q\left(\frac{na}{\sqrt{1 + (n-1)\varepsilon^2}}\right) \\ &\doteq \exp\left(-n\frac{a^2}{2\varepsilon^2}\right), \end{aligned} \quad (49)$$

while the average power is

$$E[(X_i)^2] = a^2 + 1 + \varepsilon^2. \quad (50)$$

Thus we have the following lower bound on the reliability.

**Proposition 2.** For all  $a > 0, \varepsilon > 0$

$$E_{FB, \text{binary}}(a^2 + 1 + \varepsilon^2, \varepsilon^2) \geq \frac{a^2}{2\varepsilon^2} \quad (51)$$

and (by operating only fraction  $\alpha$  of the time) for any  $0 < \alpha < 1$ ,

$$E_{FB, \text{binary}}(\alpha(a^2 + 1 + \varepsilon^2), \varepsilon^2) \geq \alpha\frac{a^2}{2\varepsilon^2}. \quad (52)$$

We can now use this proposition to prove (45). Proposition 2 implies that for  $P > 1$ ,

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon^2 \cdot E_{FB, \text{binary}}(P, \varepsilon^2) \geq \frac{P-1}{2} \quad (53)$$

and, consequently, for any  $0 \leq \alpha \leq 1$  and  $Q > 1$ ,

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon^2 \cdot E_{FB, \text{binary}}(\alpha Q, \varepsilon^2) \geq \alpha\frac{Q-1}{2}. \quad (54)$$

It follows that for any  $P > 0$  and  $0 < \alpha < \min\{1, P\}$ , by taking  $Q = P/\alpha$  in (54),

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon^2 \cdot E_{FB, \text{binary}}(P, \varepsilon^2) \geq \frac{P-\alpha}{2}, \quad (55)$$

implying (45) when taking  $\alpha \rightarrow 0$ .

Now we consider the true zero rate (i.e., any sub-exponential number of messages). The basic idea is to use  $k$  iterations of Schalkwijk–Kailath coding as inner code to transform the noisy feedback channel into a standard additive white Gaussian noise channel without feedback under new signal-to-noise ratio  $P_k$ . Then, by employing the optimal code for this new channel, we can achieve the error exponent

$$\frac{E(kR; P_k)}{k}.$$

In other words, we are using the concatenated coding with the  $k$  feedback iterations as the inner code and the optimal Gaussian nonfeedback code as the outer code. The details follow.

Let

$$\alpha = \sqrt{\frac{1+P+\varepsilon^2}{1+\varepsilon^2}}$$

and

$$\beta = \frac{P}{1+P+\varepsilon^2}.$$

If the transmitter uses the Schalkwijk–Kailath described in Section III with the noisy output  $\tilde{Y}$  in place of the real output  $Y$ , we have

$$\begin{aligned} X_i &= \alpha^{i-1}(X_1 - \hat{X}_1(\tilde{Y}^{i-1})) \\ &= \alpha(X_{i-1} - \beta\tilde{Y}_{i-1}) \end{aligned}$$

where  $\hat{X}_i(\tilde{Y}^{i-1}) = E(X_i|\tilde{Y}^{i-1})$ . From our analysis of the Schalkwijk–Kailath coding scheme, it is straightward to check that

$$\text{var}(X_1 - \hat{X}_1(\tilde{Y}^n)) = P\left(\frac{1+\varepsilon^2}{1+P+\varepsilon^2}\right)^n.$$

It is also easy to check that

$$\hat{X}_1(\tilde{Y}^{i-1}) = \alpha\beta \sum_{j=1}^{i-1} \alpha^{-j} \tilde{Y}_j.$$

Now let  $\hat{X}_1(Y^{i-1})$  denote the true receiver's estimate of  $X_1$ . Then,

$$\begin{aligned} &\text{var}\left(\hat{X}_1(Y^{i-1}) - \hat{X}_1(\tilde{Y}^{i-1})\right) \\ &\leq \text{var}\left(\alpha\beta \sum_{j=1}^{i-1} \alpha^{-j} (Y_j - \tilde{Y}_j)\right) \\ &= \text{var}\left(\alpha\beta \sum_{j=1}^{i-1} \alpha^{-j} \tilde{Z}_j\right) \\ &= (\alpha\beta)^2 \sum_{j=1}^{i-1} \alpha^{-2j} \varepsilon^2 \\ &\leq \frac{P\varepsilon^2}{1+P+\varepsilon^2}. \end{aligned}$$

Therefore,

$$\begin{aligned} &\text{var}(X_1 - \hat{X}_1(Y^n)) \\ &\leq \text{var}(X_1 - \hat{X}_1(Y^n) + \hat{X}_1(Y^n) - \hat{X}_1(Y^n)) \\ &\leq \text{var}(X_1 - \hat{X}_1(Y^n)) + \text{var}(\hat{X}_1(Y^n) - \hat{X}_1(Y^n)) \\ &\quad + 2\text{var}(X_1 - \hat{X}_1(Y^n))^{1/2} \text{var}(\hat{X}_1(Y^n) - \hat{X}_1(Y^n))^{1/2} \\ &\leq P\left(\frac{1+\varepsilon^2}{1+P+\varepsilon^2}\right)^n + \frac{P\varepsilon^2}{1+\varepsilon^2+P} \\ &\quad + 2P\sqrt{\left(\frac{1+\varepsilon^2}{1+P+\varepsilon^2}\right)^n \cdot \frac{\varepsilon^2}{1+\varepsilon^2+P}} \\ &=: V_n(P, \varepsilon^2). \end{aligned}$$

Now take

$$n^* = n^*(\varepsilon^2) = \left\lceil -\frac{\log \varepsilon^2}{\log(1+P+\varepsilon^2)} \right\rceil - 1.$$

Then, it is easy to check that

$$V_{n^*}(P, \varepsilon^2) \leq \frac{4P\varepsilon^2}{1+P+\varepsilon^2}$$

so that the  $n$  iterations of the Schalkwijk–Kailath coding scheme gives a nonfeedback Gaussian channel with effective signal-to-noise ratio

$$P_{n^*}(P, \varepsilon^2) = \frac{P}{V_{n^*}(P, \varepsilon^2)} - 1 \geq \frac{1 + \varepsilon^2 + P}{4P\varepsilon^2} - 1.$$

By concatenating the optimal code for standard Gaussian nonfeedback channels as the outer code, we thus achieve

$$E_{FB}(R; P, \varepsilon^2) \geq \frac{E(n^*R; P_{n^*})}{n^*}.$$

Finally, taking  $R \rightarrow 0$ , we have

$$E_{FB}(0; P, \varepsilon^2) \geq \frac{E(0; P_{n^*})}{n^*} = \Omega\left(\frac{1}{\varepsilon^2 \log(1/\varepsilon^2)}\right).$$

### VII. FRAGILITY OF LINEAR FEEDBACK CODING

In this section, we show that applying the standard Schalkwijk–Kailath coding scheme or its variants directly does not achieve any positive rate, not to mention the capacity of the channel. More precisely, we prove the following result:

**Proposition 3.** *Consider any sequence of  $(M_n, n)$  codes, with average power bounded by  $P$ , and the following structure: the message index  $W \in \{1, \dots, e^{nR}\}$  is first mapped to a finite dimensional constellation point as*

$$\theta(W) \in \mathbb{R}^k$$

for some finite  $k$  independent of  $n$ , and then encoded linearly for each time index  $i$  as

$$X_i(W, \tilde{Y}^{i-1}) = L_i(\theta(W), \tilde{Y}^{i-1})$$

for some affine map  $L_i$ . Assume further that the Then,

$$P_e^{(n)} \rightarrow 0$$

implies that

$$\frac{\log(M_n)}{n} \rightarrow 0.$$

This result is rather surprising, especially because the linear feedback coding scheme such as the Kailath–Schalkwijk coding scheme its variants have been very successful in many communication scenarios under perfect feedback, such as the Gaussian nonwhite feedback channel [7], Gaussian multiple access channel [13], [9], writing-on-dirty paper with feedback [12], to name a few.

*Proof.* Proof by contradiction. Suppose there exists a sequence of linear feedback coding schemes that achieve  $R > 0$ . For simplicity, we assume that  $k = 1$ . Then according to the structure of the linear coding scheme, we have

$$\theta = \theta(W) \in \mathbb{R},$$

which, combined with the positive achievable rate, implies that there exists  $\alpha > 0$  such that

$$\frac{1}{n}I(\theta; Y^n) \geq \alpha$$

for all  $n$ .

Now from the restriction of linear feedback coding and the additive nature of the channel, we can represent the channel as

$$Y_i = \alpha_i \theta + \xi_i, \quad i = 1, 2, \dots, n$$

for a Gaussian random vector  $(\xi_1, \dots, \xi_n)$ . Since a Gaussian input maximizes the mutual information over Gaussian channels, we have

$$\frac{1}{n}I(\tilde{\theta}; Y^n) \geq \frac{1}{n}I(\theta; Y^n) \geq \alpha$$

where  $\tilde{\theta} \sim N(E(\theta(W)), \text{var}(\theta(W)))$ .

Now from joint Gaussianity of  $(\tilde{\theta}, Y^n)$ , it is easy to check that there exists a linear function  $L(Y^n)$  such that

$$I(\tilde{\theta}; L(Y^n)) = I(\tilde{\theta}; Y^n) > \alpha \quad (56)$$

and

$$L(Y^n) = \tilde{\theta} + \xi$$

where  $\xi \sim \mathcal{N}(0, E\xi^2)$  is independent of  $\tilde{\theta}$ . Furthermore, from (56), we can easily lower bound  $E\xi^2$  as

$$E\xi^2 \leq \frac{E\theta^2}{e^{2n\alpha} - 1}.$$

But suppose we use the uniform message constellation for the channel  $\theta \rightarrow L(Y^n)$ , as in the original Schalkwijk–Kailath coding scheme. Then, we can achieve

$$P_e^{(n)} \leq \exp\left(-\frac{1}{2}e^{2n(\alpha-R)}\right)$$

for any rate  $R < \alpha$ . In particular,

$$E_{FB}(R; P, \varepsilon^2) = \infty$$

for  $R < \alpha$ , which contradicts our upper bounds stating  $E_{FB}(R; P, \varepsilon^2) < \infty$  for all  $R$ .  $\square$

### VIII. CONCLUDING REMARKS

We have obtained upper bounds on the reliability of Gaussian channels with noisy feedback, which show that the reliability  $E_{FB}(R; P, \varepsilon^2)$  under power constraint  $P$  and noisy feedback with noise variance  $\varepsilon^2$  is finite for every positive  $\varepsilon^2$  and in fact

$$E_{FB}(R; P, \varepsilon^2) = E(R; O(1/\varepsilon^2))$$

in low  $\varepsilon^2$  asymptotics. As a matching lower bound, we presented a feedback coding scheme that achieves

$$E_{FB}(R = 0; P, \varepsilon^2) \geq E(R = 0; O(1/\varepsilon^{2-\delta}))$$

for all  $\delta > 0$ . However, the problem is still open which feedback coding scheme can achieve the error exponent that blows up as  $\varepsilon^2 \rightarrow 0$  at strictly positive rate, if this is indeed possible. Linear feedback coding schemes, which are well-known to be very effective under perfect feedback (and in many other scenarios), fail to achieve this goal in the strong sense that they cannot achieve a positive rate of reliable communication.

Although we have restricted our attention to the Gaussian channel with noisy feedback, our approach and techniques

are applicable more generally. One straightforward generalization is to the case of non-Gaussian channels. It is readily verified that the proof of Theorem 1 carries over to the case where the noise components are generally distributed. More specifically, if instead of (6) we have  $Z \sim f(z)$  and  $\tilde{Z} \sim f(s)$  that are independent (and, as before,  $S = Z + \tilde{Z}$ ), we can take  $(Z', S')$  to be a pair of independent variables where  $Z'$  can be arbitrarily distributed  $\sim f'(z)$  and  $S' \stackrel{d}{=} S$ . Then, we have

$$E_{FB}(R; P, f') \leq \gamma + E(R; P, f') \quad (57)$$

for any  $\gamma \geq D(f' \| f)$  for which  $\Lambda_{\varepsilon^2, \sigma^2}^*(\gamma) \geq E(R; P, f')$ , where  $\Lambda_{\varepsilon^2, \sigma^2}^*$  is the Fenchel–Legendre transform of the random variable  $\log \frac{f'(Z', S')}{f(Z', S')}$ . Note that the restriction  $S' \stackrel{d}{=} S$  guarantees that the input constraint  $P$  used by the scheme in the useless feedback setting is the same as in the original setting. A similar bound can be obtained for finite-alphabet channels with modulo-additive noise. For this case, in the absence of power (or cost) constraints, the restriction  $S' \stackrel{d}{=} S$  is no longer required. The approach behind the bound of Section V can also be extended to the non-Gaussian and discrete cases.

As another direction of extensions, we can allow encoding of the output signal  $Y$  over the backward channel  $\tilde{Y} = \tilde{X} + \tilde{Z}$  under the feedback function  $\tilde{X}_i = \tilde{X}_i(Y^i)$  with power constraint  $\tilde{P}$ . Thus, we use the Gaussian two-way channel for one-way information flow. Moreover, instead of block coding, we may use the variable-length coding. With these two additional power, we can show that the reliability is lower-bounded as

$$E_{FB}(R; P, \tilde{P}, \varepsilon^2) \geq E(R; O(1/\varepsilon^2)).$$

Unfortunately, our upper bounding techniques are not applicable to the encoded feedback. Also for the variable-length coding with noisy feedback, the problem formulation involves some subtlety, because the transmitter and the receiver cannot agree upon the stopping time of communication, due to the uncertainty stemming from noise in the feedback link. These and other extensions will be detailed elsewhere.

#### ACKNOWLEDGMENT

Discussions with Amir Dembo are gratefully acknowledged.

#### REFERENCES

- [1] A. E. Ashikhmin, A. Barg, and S. N. Litsyn, "New Upper Bound on the Reliability Function of the Gaussian Channel," *IEEE Trans. Inform. Theory*, vol. IT-46, Number 6, pp. 1945–1961, September 2000.
- [2] S. Butman, "A general formulation of linear feedback communication systems with solutions," *IEEE Trans. Inf. Theory*, vol. IT-15, no. 3, pp. 392–400, May 1969.
- [3] —, "Linear feedback rate bounds for regressive channels," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 3, pp. 363–366, May 1976.
- [4] M. H. M. Costa, "Writing on Dirty Paper," *IEEE Trans. Inform. Theory*, vol. IT-29 pp. 439–441, May 1983.
- [5] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. Springer, New York, 1997.
- [6] S. C. Draper and A. Sahai, "Noisy feedback improves communication reliability", In *Int. Symp. Inf. Th.*, Seattle, Washington, July 2006
- [7] Y.-H. Kim, "Feedback capacity of stationary Gaussian channels," submitted to *IEEE Trans. Info. Theory*, Feb. 2006. available at [arxiv.org/cs.IT/060291](http://arxiv.org/cs.IT/060291)
- [8] A. J. Kramer, "Improving communication reliability by use of an intermittent feedback channel," *IEEE Trans. Inf. Theory*, vol. IT-15, pp. 52–60, Jan. 1969.
- [9] G. Kramer and M. Gastpar, "Dependence balance and the Gaussian multiaccess channel with feedback," in *IEEE Information Theory Workshop*, Punta del Este, Uruguay, March 2006, pp. 198–202.
- [10] T. Liu, P. Moulin, and R. Koetter, "On error exponents of modulo lattice additive noise channels," *IEEE Trans. Inf. Theory*, vol. IT-52, no. 2, pp. 454–471, Feb 2006.
- [11] N.C. Martins and T. Weissman, "Coding Schemes for Additive White Noise Channels with Feedback Corrupted by Quantization or Bounded Noise," available at the arxiv.
- [12] N. Merhav and T. Weissman, "Coding for the feedback Gel'fand-Pinsker channel and the feedforward Wyner-Ziv source," *Int. Symp. Inf. Th.*, Adelaide, Australia, September 2005.
- [13] L. H. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 4, pp. 623–629, 1984.
- [14] M. S. Pinsker, "The probability of error in block transmission in a memoryless Gaussian channel with feedback," *Problemy Peredači Informacii*, vol. 4, no. 4, pp. 3–19, 1968.
- [15] A. Sahai and T. Şimşek, "On the variable-delay reliability function of discrete memoryless channels with access to noisy feedback", In *IEEE Inf. Th. Workshop*, San Antonia, Texas, 2004.
- [16] J. P. M. Schalkwijk, "A Coding Scheme for Additive Noise Channels with Feedback-II: Band-limited Signals," *IEEE Trans. Inform. Theory*, vol. 12 pp. 183–189, Apr. 1966.
- [17] J. P. M. Schalkwijk and T. Kailath, "A Coding Scheme for Additive Noise Channels with Feedback-I: No Bandwidth Constraint," *IEEE Trans. Inform. Theory*, vol. 12 pp. 183–189, Apr. 1966.
- [18] C. E. Shannon, "The Zero-Error Capacity of a Noisy Channel," *IRE Trans. Inform. Theory*, vol. IT-12, pp. 8–19, Sept. 1956.
- [19] K. Sh. Zigangirov, "Upper Bounds for the Error Probability for Channels with Feedback," *Prob. Inf. Trans.*, Volume 6, Number 2, April-June, 1970, pp. 159-163.