

# Bounds on the Error Exponent of the AWGN Channel with AWGN-Corrupted Feedback

Young-Han Kim

Electrical Engineering

Stanford University

Stanford 94305-9510, USA

yhk@stanford.edu

Amos Lapidoth

ETH-Zentrum, ETF E107

Swiss Federal Institute of Technology (ETH)

Zurich, CH-8092, Switzerland

lapidoth@isi.ee.ethz.ch

Tsachy Weissman

Electrical Engineering

Stanford University

Stanford 94305-9510, USA

tsachy@stanford.edu

**Abstract**— We derive bounds on the error exponent of the AWGN channel with AWGN-corrupted feedback. The bounds appear to be new even for transmission at zero 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

The fact that noise-free feedback can dramatically improve the reliability of a memoryless channel was pointed out by Shannon in [7]. For the AWGN channel this fact was shown in [6] to be true in the strong sense that noise-free feedback allows for schemes under which the probability of error diminishes not only super-exponentially (i.e., infinite error exponent) but, in fact, double-exponentially with block length (at any rate below capacity). Zigangirov in [8] subsequently showed that, in fact, the error can be made to diminish at a rate of any number of exponential levels.

Much less explored and understood is the way in which noise in the feedback link affects the achievable reliability (cf. [4], [5] for recent exceptions). In this extended abstract we restrict attention, for concreteness, to the AWGN channel with AWGN-corrupted feedback, and derive bounds on its error exponent. After describing the channel in Section II, we present a family of upper bounds on its error exponent in Section III. The bounds are obtained via a change of measure argument. 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. Section IV presents another upper bound on the same exponent via a ‘genie-aided’ scheme. In Section V we discuss the zero-rate and small feedback noise regime. We close in Section VI with a description of the way our bounds extend to the non-Gaussian and to the discrete settings.

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

## II. THE CHANNEL

Except in Section VI, we assume an AWGN channel

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

where  $X_i, Y_i$  are, respectively, the channel input and output, and  $N_i$  is the noise in the forward link. Let further  $Z_i$  denote a noisy version of  $Y_i$ ,

$$Z_i = Y_i + W_i, \quad (2)$$

where  $W_i$  is the noise in the backward link. Encoding here is of the form

$$X_i = X_i(m, Z^{i-1}), \quad (3)$$

where  $m \in \{1, \dots, 2^{nR}\}$  is the message. In other words, the encoder has noisy feedback. Note that, equivalently, we can consider the encoding to be of the form

$$X_i(m, V^{i-1}), \quad (4)$$

where

$$V_i = N_i + W_i, \quad (5)$$

since, given  $X^i$ , there is a one-to-one transformation from  $Z^i$  to  $V^i$ . The forward and backward noise processes,  $\{N_i\}$  and  $\{W_i\}$ , are independent and white, with variances

$$N_i \sim \mathcal{N}(0, 1), \quad W_i \sim \mathcal{N}(0, \varepsilon^2). \quad (6)$$

Let  $E_{\text{NoisyFB}}(P, \varepsilon^2, R)$  denote the error exponent associated with this setting (where the first argument stands for the power constraint).

## III. UPPER BOUNDS ON $E_{\text{NoisyFB}}(P, \varepsilon^2, R)$ VIA CHANGE OF MEASURE

Let  $E_{\text{NoFB}}(s, R)$  denote the regular error exponent of the AWGN channel (in the absence of feedback) at SNR level  $s$ . To state the main result of this section, we let  $(N, V)$  be a generic pair of random variables distributed as the pair  $(N_i, V_i)$  of (5) and (6), and we let  $(N', V')$  be a pair of independent Gaussians with  $N' \sim \mathcal{N}(0, \sigma^2)$  and  $V' \sim \mathcal{N}(0, 1 + \varepsilon^2)$ . Let further  $f_{N, V}$  and  $f_{N', V'}$  denote the respective densities of  $(N, V)$  and  $(N', V')$ . Finally, let  $\Lambda_{\varepsilon^2, \sigma^2}^*$  denote the Fenchel-Legendre transform (cf., e.g., [3]) of the random variable  $\log \frac{f_{N', V'}(N', V')}{f_{N, V}(N, V)}$ . We can now present the main result of this section. The proof involves changing the law of the noise components in the original channel from  $(N_i, V_i) \stackrel{d}{=} (N, V)$  to  $(N_i, V_i) \stackrel{d}{=} (N', V')$ . The feedback in the latter case is useless, so the associated error exponent is  $E_{\text{NoFB}}$ .

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

$$E_{\text{NoisyFB}}(P, \varepsilon^2, R) \leq \gamma + E_{\text{NoFB}}(P/\sigma^2, R) \quad (7)$$

for the (unique)  $\gamma$  solving

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

in the region  $\gamma \geq D(f_{N', V'} \| f_{N, V})$  where

$$D(f_{N', V'} \| f_{N, V}) = \frac{1}{2} \left[ \frac{\sigma^2 - \varepsilon^2 + \sigma^2 \varepsilon^2 + 1}{\varepsilon^2} - \log[\sigma^2 \varepsilon^2 (1 + \varepsilon^2)] \right]. \quad (8)$$

Note that the 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

$$f_{N, V}(n, v) = \frac{\varepsilon}{2\pi} \exp \left\{ -\frac{(1 + \varepsilon^2)n^2 - 2nv + v^2}{2\varepsilon^2} \right\}, \quad (9)$$

and

$$f_{N', V'}(n', v') = \frac{1}{2\pi \sqrt{\sigma^2(1 + \varepsilon^2)}} \exp \left\{ -\frac{1}{2} \left[ \frac{n'^2}{\sigma^2} + \frac{v'^2}{1 + \varepsilon^2} \right] \right\} \quad (10)$$

so that

$$\log \frac{f_{N', V'}(N', V')}{f_{N, V}(N', V')} = -\frac{1}{2} \log[\sigma^2 \varepsilon^2 (1 + \varepsilon^2)] + \frac{1}{2} \left[ \left( \frac{N'}{\sigma} \right)^2 \frac{\sigma^2 - \varepsilon^2 + \sigma^2 \varepsilon^2}{\varepsilon^2} - \frac{2\sigma \sqrt{1 + \varepsilon^2} N'}{\varepsilon^2} \frac{V'}{\sigma \sqrt{1 + \varepsilon^2}} + \left( \frac{V'}{\sqrt{1 + \varepsilon^2}} \right)^2 \frac{1}{\varepsilon^2} \right]. \quad (11)$$

Since both  $\frac{N'}{\sigma}$  and  $\frac{V'}{\sqrt{1 + \varepsilon^2}}$  are standard Normal 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

$$A^2 \frac{\sigma^2 + \sigma^2 \varepsilon^2 - \varepsilon^2}{\varepsilon^2} - \frac{2\sigma \sqrt{1 + \varepsilon^2}}{\varepsilon^2} AB + B^2 \frac{1}{\varepsilon^2}, \quad (13)$$

$A$  and  $B$  being independent standard normal random variables. Fortunately  $\tilde{\Lambda}_{\varepsilon^2, \sigma^2}^*$  can be derived in closed form

$$\tilde{\Lambda}_{\varepsilon^2, \sigma^2}^*(\alpha) = \frac{1}{4} \left( -2 - \alpha(1 + \sigma^2 + \varepsilon^2(\sigma^2 - 1)) + \sqrt{4 + \alpha^2(\varepsilon^4(\sigma^2 - 1)^2 + (\sigma^2 + 1)^2 + 2\varepsilon^2(1 + \sigma^4))} \right) + \frac{1}{2} \times \log \frac{-2 + \sqrt{4 + \alpha^2(\varepsilon^4(1 - \sigma^2)^2 + (1 + \sigma^2)^2 + 2\varepsilon^2(1 + \sigma^4))}}{\alpha^2 \varepsilon^2} \quad (14)$$

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  $R_{\text{NoFB}}(s, e)$  denote the ‘‘rate-reliability’’ function of the AWGN channel in the absence of feedback, i.e.

$$R_{\text{NoFB}}(s, e) = \begin{cases} R \text{ solving } E_{\text{NoFB}}(s, R) = e & \text{for } 0 \leq e < E_{\text{NoFB}}(s, 0) \\ 0 & \text{for } e \geq E_{\text{NoFB}}(s, 0) \end{cases}. \quad (15)$$

Note, in particular, that

$$R_{\text{NoFB}}(s, 0) = \frac{1}{2} \log(1 + s). \quad (16)$$

Theorem 1 can be stated equivalently as follows.

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

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

where  $\gamma$  is varied in the range

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

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

Though the functions  $E_{\text{NoFB}}$  and  $R_{\text{NoFB}}$  are unfortunately unknown, 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  $E_{\text{NoisyFB}}(P, \varepsilon^2, \cdot)$

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

where  $\gamma$  is varied in the range given in (18) and  $\gamma'_{\text{max}, \sigma^2}$  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

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

(cf., e.g., [1]) which implies also

$$R_{\text{NoFB}}(s, e) \leq \frac{1}{2} \log(1 + s) \left[ 1 - \frac{2e}{s} \right]. \quad Q.E.D. \quad (21)$$

Figure 1 shows the curves of Corollary 1, for the case of  $\varepsilon^2 = 1$ , and several values of  $\sigma$ . 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 *Corollary 2:*

$$E_{\text{NoisyFB}}(P, \varepsilon^2, 0) \leq \frac{P}{2\sigma^2} + \gamma, \quad (22)$$

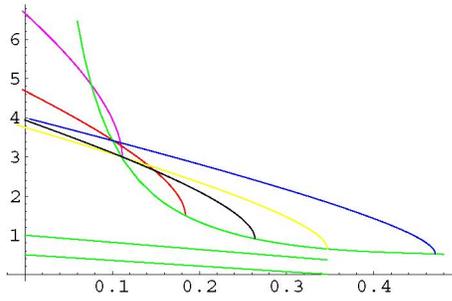


Fig. 1. Upper bounds on  $E_{\text{NoisyFB}}(1, 1, R)$ . The purple, red, black, yellow, and blue curves are the parametric bounds of Corollary 1 for the respective values of  $\sigma$ : 2, 1.5, 1.2, 1, and 0.8. The green curve is the bound of Corollary 3. The green line (higher of the two) is the straight 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  $0 \leq R \leq \frac{1}{2} \log 2$ .

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

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

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

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

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

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

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  $l$ ,

$$P_{e|m} = \Pr((N^l, V^l) \in A_m), \quad (25)$$

$A_m$  denoting the error set

$$A_m = \{(n^l, v^l) : \hat{m} \neq m\}, \quad (26)$$

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

Consider now the following scenario of communications with useless feedback:

$$Y_i' = X_i' + N_i', \quad X_i' = X_i'(m, V^{i-1}), \quad (27)$$

where  $\{N_i'\}$  and  $\{V_i'\}$  are independent white noises with

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

Let  $P'_{e|m}$  denote the probability of error of the coding scheme associated with  $P_{e|m}$ , when operating in the useless feedback setting. Thus, letting  $f_{N^l, V^l}$  and  $f_{N^l, V^l}$  denote the respective densities of  $(N^l, V^l)$  and  $(N^l, V^l)$ , and denoting

$$B_\gamma = \left\{ (n^l, v^l) : \frac{1}{l} \log \frac{f_{N^l, V^l}(n^l, v^l)}{f_{N^l, V^l}(n^l, v^l)} \leq \gamma \right\}, \quad (29)$$

we have

$$\begin{aligned} P'_{e|m} &= \Pr((N^l, V^l) \in A_m) \\ &= \int_{A_m} f_{N^l, V^l}(n^l, v^l) dn^l dv^l \\ &= \int_{A_m \cap B_\gamma} f_{N^l, V^l}(n^l, v^l) dn^l dv^l \\ &\quad + \int_{A_m \cap B_\gamma^c} f_{N^l, V^l}(n^l, v^l) dn^l dv^l \\ &\leq e^{l\gamma} \int_{A_m \cap B_\gamma} f_{N^l, V^l}(n^l, v^l) dn^l dv^l \\ &\quad + \int_{A_m \cap B_\gamma^c} f_{N^l, V^l}(n^l, v^l) dn^l dv^l \\ &\leq e^{l\gamma} \int_{A_m} f_{N^l, V^l}(n^l, v^l) dn^l dv^l \\ &\quad + \int_{B_\gamma^c} f_{N^l, V^l}(n^l, v^l) dn^l dv^l \\ &= e^{l\gamma} \Pr((N^l, V^l) \in A_m) + \Pr((N^l, V^l) \in B_\gamma^c) \\ &= e^{l\gamma} P_{e|m} + \Pr((N^l, V^l) \in B_\gamma^c). \end{aligned} \quad (30)$$

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

$$P'_e \leq e^{l\gamma} P_e + \Pr((N^l, V^l) \in B_\gamma^c) \quad (31)$$

or, equivalently,

$$e^{-l\gamma} P'_e - e^{-l\gamma} \Pr((N^l, V^l) \in B_\gamma^c) \leq P_e. \quad (32)$$

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

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

implying

$$\begin{aligned} -\frac{1}{l} \log P_e &\leq \gamma - \\ &\frac{1}{l} \log [P_{\min}^l(P, \sigma^2, R) - \Pr((N^l, V^l) \in B_\gamma^c)], \end{aligned} \quad (34)$$

where  $P_{\min}^l(P, \sigma^2, R)$  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  $E_{\text{NoFB}}$ ,

$$\lim_{l \rightarrow \infty} -\frac{1}{l} \log P_{\min}^l(P, \sigma^2, R) = E_{\text{NoFB}}(P/\sigma^2, R) \quad (35)$$

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

it follows that

$$E_{\text{NoisyFB}}(P, \varepsilon^2, R) \leq \gamma + E_{\text{NoFB}}(P/\sigma^2, R) \quad (36)$$

for  $\gamma$  sufficiently large that

$$\liminf_{l \rightarrow \infty} -\frac{1}{l} \log \Pr((N^l, V^l) \in B_\gamma^c) > E_{\text{NoFB}}(P/\sigma^2, R). \quad (37)$$

The lim inf in (7), however, is in fact a limit we can explicitly characterize. Indeed, by the definition of  $B_\gamma$  in (29),

$$\Pr((N^l, V^l) \in B_\gamma^c) \quad (38)$$

$$= \Pr\left(\frac{1}{l} \log \frac{f_{N^l, V^l}(N^l, V^l)}{f_{N^l, V^l}(N^l, V^l)} > \gamma\right) \quad (39)$$

$$= \Pr\left(\frac{1}{l} \sum_{i=1}^l \log \frac{f_{N^l, V^l}(N_i^l, V_i^l)}{f_{N, V}(N_i^l, V_i^l)} > \gamma\right) \quad (40)$$

thus, for  $\gamma \geq E\left[\log \frac{f_{N^l, V^l}(N^l, V^l)}{f_{N, V}(N^l, V^l)}\right] = D(f_{N^l, V^l} \| f_{N, V})$ , Cramér's theorem (cf. [3, Th. 2.2.3]) implies

$$\lim_{l \rightarrow \infty} -\frac{1}{l} \log \Pr((N^l, V^l) \in B_\gamma^c) = \Lambda^*(\gamma), \quad (41)$$

where  $\Lambda^*$  is the Fenchel-Legendre transform of the random variable  $\log \frac{f_{N^l, V^l}(N^l, V^l)}{f_{N, V}(N^l, V^l)}$ . It follows by substitution of (41) in (37) that (7) holds for all  $\gamma \geq D(f_{N^l, V^l} \| f_{N, V})$  satisfying  $\Lambda^*(\gamma) > E_{\text{NoFB}}(P/\sigma^2, R)$  and, consequently, by the continuity and strict monotonicity of  $\Lambda^*(\gamma)$ , for the value of  $\gamma$  satisfying  $\Lambda^*(\gamma) = E_{\text{NoFB}}(P/\sigma^2, R)$ . Q.E.D.

#### IV. UPPER BOUND ON $E_{\text{NOISYFB}}$ VIA GENIE

Consider a genie-aided scheme where encoding is allowed to depend on the  $V_i$  sequence non-causally, i.e., to be of the form  $X_i = X_i(m, V^n)$  instead of  $X_i(m, V^{i-1})$ . Assume further that the decoder is given access, in addition to  $Y^n$ , also to  $V^n$ , i.e.,  $\hat{m} = \hat{m}(Y^n, V^n)$ . By conditioning on  $V^n$  we then see that the capacity and error exponent for this setting is exactly that for the standard AWGN channel with no feedback and noise variance equals to  $\text{Var}(N_i|V_i) = \text{Var}(N_i|N_i+W_i) = \frac{\varepsilon^2}{\varepsilon^2+1}$ . 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. Thus:

*Proposition 1:*

$$E_{\text{NoisyFB}}(P, \varepsilon^2, R) \leq E_{\text{NoFB}}\left(P \frac{\varepsilon^2 + 1}{\varepsilon^2}, R\right). \quad (42)$$

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 a power increase of  $P/\varepsilon^2$  in the absence of feedback. Furthermore, when combined with the sphere packing bound on  $E_{\text{NoFB}}$ , Proposition 1 gives

$$E_{\text{NoisyFB}}(P, \varepsilon^2, 0) \leq \frac{P \varepsilon^2 + 1}{2 \varepsilon^2} \quad (43)$$

implying that  $E_{\text{NoisyFB}}(P, \varepsilon^2, 0)$  increases with small  $\varepsilon$  essentially no faster than  $\frac{P}{2\varepsilon^2}$ . The section that follows shows

that this bound is in fact tight. Finally, we note that the bound of Proposition 1 can potentially be tightened by denying the decoder of the genie-aided scheme access to  $V^n$ .  $E_{\text{NoFB}}$  on the right-hand side of (7) can thus be replaced by the error exponent of the corresponding dirty paper problem [2]. Unfortunately, it is as yet unknown whether the latter is strictly smaller than  $E_{\text{NoFB}}$ .

#### V. ZERO RATE AND SMALL FEEDBACK NOISE

In this section we briefly address the asymptotic regime of zero rate and feedback noise of very small variance. Specializing Inequality (43) to this regime yields

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon^2 \cdot E_{\text{NoisyFB}}(P, \varepsilon^2, 0) \leq \frac{P}{2}. \quad (44)$$

We shall next show that the RHS of the above is achievable in the two-message setting. Denoting the two-message error exponent by  $E_{\text{binary, NoisyFB}}(P, \varepsilon^2)$  we shall thus show that

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

To this end, we consider the following scheme for communicating one (unbiased) bit: Assuming that the value say '1' is to be communicated, the channel input is

$$X_i = a - V_{i-1} = a - N_{i-1} - W_{i-1}. \quad (46)$$

The channel output will then be

$$Y_i = X_i + N_i = a - N_{i-1} + N_i - W_{i-1}. \quad (47)$$

If '0' is to be communicated then  $a$  will be replaced by  $-a$ . The decoder computes  $S_n = \sum_{i=1}^n Y_i$  and decides that a '1' was sent if  $S_n > 0$ , otherwise it decides on '0'. Thus, conditioned on '1' being sent,

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

which implies

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

while the average power is

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

We thus obtain

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

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

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

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

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

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

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

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

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

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

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

## VI. EXTENSIONS

In order to give a reasonably detailed and self-contained description, we have chosen to restrict the exposition to the case of the AWGN channel with AWGN-corrupted feedback. Our approach and techniques, however, 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  $N_i \sim f_N$  and  $W_i \sim f_W$  that are independent (and, as before,  $V_i = N_i + W_i$ ), we can take  $(N', V')$  to be a pair of independent variables where  $N'$  can be arbitrarily distributed and  $V' \stackrel{d}{=} V_i$ . Then we would have

$$E_{\text{NoisyFB}}(P, f_N, f_W, R) \leq \gamma + E_{\text{NoFB}}(P, f_{N'}, R) \quad (56)$$

for any  $\gamma \geq D(f_{N',V'} \| f_{N,V})$  for which  $\Lambda_{\varepsilon^2, \sigma^2}^*(\gamma) \geq E_{\text{NoFB}}(P, f_{N'}, R)$ , where  $\Lambda_{\varepsilon^2, \sigma^2}^*$  is the Fenchel-Legendre transform of the random variable  $\log \frac{f_{N',V'}(N',V')}{f_{N,V}(N',V')}$ . Here  $E_{\text{NoisyFB}}(P, f_N, f_W, R)$  denotes the error exponent at power level  $P$  of the channel of interest (namely where  $N_i \sim f_N$  and  $W_i \sim f_W$ ) and  $E_{\text{NoFB}}(P, f_{N'}, R)$  of the channel with no feedback and memoryless additive noise components distributed as  $f_{N'}$ . Note that the restriction to  $V'$  such that  $V' \stackrel{d}{=} V_i$  is to guarantee that the power 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 (where densities would be replaced by PMFs). For this case, in the absence of power (or cost) constraints, the restriction  $V' \stackrel{d}{=} V_i$  is no longer required. The approach behind the bound of Section IV can also be extended to the non-Gaussian and discrete cases.

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] M. H. M. Costa, "Writing on Dirty Paper," *IEEE Trans. Inform. Theory*, vol. IT-29 pp. 439–441, May 1983.
- [3] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. Springer, New York, 1997.
- [4] S. C. Draper and A. Sahai, "Noisy feedback improves communication reliability", In *Int. Symp. Inf. Th.*, Seattle, Washington, July 2006
- [5] 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 Antonio, Texas, 2004.
- [6] 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.
- [7] C. E. Shannon, "The Zero-Error Capacity of a Noisy Channel," *IRE Trans. Inform. Theory*, vol. IT-12, pp. 8–19, Sept. 1956.
- [8] 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.