

Coding for the Gaussian Channel with Intermittent Feedback

Christoph Bunte, Amos Lapidoth, and Lars Palzer
ETH Zurich

Email: {bunte, lapidoth}@isi.ee.ethz.ch, palzerl@student.ethz.ch

Abstract—Optimal error probabilities for transmission over the average-power-limited Gaussian channel with intermittent feedback are studied. For the two-message case, the asymptotic decay of the probability of error in the blocklength is double-exponential and is fully characterized. For positive rates a critical rate is identified below which a double-exponential decay is possible and above which it is not.

I. INTRODUCTION

The asymptotic behavior of the best achievable error probability over the memoryless Gaussian channel has received much attention over the years both in the absence and in the presence of feedback. Much of the work on feedback has focused on ideal feedback (as in [1] and references therein), but there has recently been growing interest in imperfect feedback, e.g., [4]–[8]. Most relevant to the present work is the model introduced in [2] where the feedback is “intermittent,” i.e., where each channel output is fed back with some given probability ρ and the receiver is cognizant of which outputs were fed back. While [2] dealt with the multiple-access channel, here we focus on the single-user channel and on the achievable probabilities of error. As in [2], the results in this paper are derived under the assumption that the receiver knows which channel outputs are fed back to the transmitter. (A partial result for the case when the receiver does not know which symbols are fed back is presented in Section V.)

We show that if only one of two equally-likely messages is to be transmitted over the channel, then the optimal error probability decays with the blocklength double-exponentially with second-order exponent $-\log(1-\rho)$; see Theorem III.1 ahead. When transmitting at a positive rate R , we show that a double-exponential decay of the probability of error is possible if $R < \rho C$, where C denotes channel capacity. If the rate exceeds ρC , then the optimal probability of error decays at most exponentially in the blocklength; see Theorem IV.1 ahead.

Our model is more pessimistic than the one in [4], and indeed our converse for the positive-rate case is based on the converse in [4]. In this paper, most of the effort goes into the achievability proofs.

II. NOTATION AND PRELIMINARIES

We denote by $\mathcal{Q}(\xi)$ the probability that a standard Gaussian exceeds the value ξ . The following bounds will be useful.

$$\frac{1}{\sqrt{2\pi\xi^2}}e^{-\xi^2/2}\left(1 - \frac{1}{\xi^2}\right) < \mathcal{Q}(\xi) \leq \frac{1}{2}e^{-\xi^2/2}, \quad \xi \geq 0. \quad (1)$$

We use the notation $x^n = (x_1, \dots, x_n)$ for n -tuples. The complement of an event E is denoted by E^c . All logarithms are natural logarithms.

III. THE TWO-MESSAGE CASE

Consider the problem of transmitting one of two equiprobable messages in n uses of a channel whose output at time k is

$$Y_k = x_k + Z_k, \quad (2)$$

where x_k is the input of the channel at time k , and where $\{Z_k\}_{k=1}^\infty$ are IID Gaussian of zero mean and unit variance. We assume that the transmitter receives intermittent feedback in the sense that Y_k is revealed causally to the transmitter if, and only if, $V_k = 1$, where $\{V_k\}_{k=1}^\infty$ are independent of the noise $\{Z_k\}_{k=1}^\infty$ and of the message and are IID Bernoulli with $\Pr(V_k = 1) = 1 - \Pr(V_k = 0) = \rho$ for some $\rho \in (0, 1)$. The input of the channel at time k is thus a function of the message, of V^{k-1} , and of those Y_i with $1 \leq i \leq k-1$ for which $V_i = 1$. The receiver knows which symbols were fed back to the transmitter, i.e., it knows V^n .

To be more precise, let (Ω, \mathcal{F}, P) be the underlying probability space and let \mathcal{F}_k be the σ -field of all events $A \in \mathcal{F}$ such that for all $v^k \in \{0, 1\}^k$ we have $\{V^k = v^k\} \cap A \in \sigma(V^k, Y_{i(1)}, \dots, Y_{i(\ell)})$ where $i(1), \dots, i(\ell)$ are those indices $i \in \{1, \dots, k\}$ for which $v_i = 1$. A two-message blocklength- n code for this channel comprises two random n -tuples $X^n(m)$, $m \in \{0, 1\}$, such that each $X_k(m)$ is \mathcal{F}_{k-1} -measurable, and a (Borel-measurable) decoder mapping

$$\varphi: \mathbb{R}^n \times \{0, 1\}^n \rightarrow \{0, 1\}. \quad (3)$$

The probability of error is

$$\Pr(\text{error}) = \Pr(\varphi(Y^n, V^n) \neq M), \quad (4)$$

where conditional on $M = m$ the input to the channel is $X^n(m)$, and $\Pr(M = 0) = \Pr(M = 1) = 1/2$. We permit only codes that satisfy the power constraint

$$\mathbb{E}\left[\sum_{k=1}^n X_k(m)^2\right] \leq nP, \quad m \in \{0, 1\}, \quad (5)$$

where P is a given positive number. We denote by $P_e^{(n)}$ the least probability of error of any two-message blocklength- n code satisfying (5).

Theorem III.1. For any $P > 0$ and any $\rho \in (0, 1)$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log(-\log P_e^{(n)}) = -\log(1 - \rho). \quad (6)$$

Proof. We begin with the converse part. Let N denote the event that the transmitter receives no feedback at all. Then $\Pr(N) = (1 - \rho)^n$. Suppose for the moment that the transmitter knows in advance which output symbols will be fed back, i.e., V^n is revealed noncausally. This can only decrease the minimum probability of error. Moreover,

$$\Pr(\text{error}) \geq \Pr(\text{error}|N)(1 - \rho)^n, \quad (7)$$

and

$$\mathbb{E} \left[\sum_{k=1}^n X_k(m)^2 \right] \geq \mathbb{E} \left[\sum_{k=1}^n X_k(m)^2 \middle| N \right] (1 - \rho)^n. \quad (8)$$

To satisfy (5), we must therefore have

$$\mathbb{E} \left[\sum_{k=1}^n X_k(m)^2 \middle| N \right] \leq (1 - \rho)^{-n} n P. \quad (9)$$

Subject to the constraint (9), the conditional probability of error $\Pr(\text{error}|N)$ is minimized by using binary antipodal signaling whenever N occurs:

$$X_1(m) = \begin{cases} \sqrt{(1 - \rho)^{-n} n P} & \text{if } m = 0, \\ -\sqrt{(1 - \rho)^{-n} n P} & \text{if } m = 1, \end{cases} \quad (10)$$

and $X_2(m) = \dots = X_n(m) = 0$; and to remain silent otherwise: $X_1(m) = \dots = X_n(m) = 0$. Thus,

$$\Pr(\text{error}|N) \geq \mathcal{Q}(\sqrt{(1 - \rho)^{-n} n P}). \quad (11)$$

Using the lower bound in (1), it follows from (7) and (11) that

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log(-\log P_e^{(n)}) \leq -\log(1 - \rho), \quad (12)$$

which is the desired converse result.

To prove the direct part we exhibit a sequence of codes that achieves a second-order error exponent of $-\log(1 - \rho)$. We describe the construction of $X^n(m)$ when $m = 0$; the description for $m = 1$ is similar with obvious reversals of signs and inequalities. To simplify notation, we henceforth write X_k instead of $X_k(0)$. In the rest of this proof, all probabilities and expectations are implicitly conditioned on $M = 0$.

Fix the blocklength n and let T_1 and T_2 denote the times of occurrence of the first and second feedback symbols, i.e.,

$$T_1 = \min\{k \geq 1 : V_k = 1\}, \quad T_2 = \min\{k > T_1 : V_k = 1\}.$$

Note that, by convention, the minimum of an empty set is infinity: $\min \emptyset = +\infty$.

It will be convenient to describe the coding scheme separately for the case where $T_2 \leq n - 1$, the case where $T_2 > n - 1$ and $T_1 < n - 1$, and the case where $T_1 \geq n - 1$. We begin with the case where $T_2 \leq n - 1$.

- 1st phase: $X_k = A_1, 1 \leq k \leq T_1$.

- 2nd phase:

$$X_k = \begin{cases} 0 & \text{if } Y_{T_1} \geq 0, \\ A_2 & \text{if } Y_{T_1} < 0, \end{cases} \quad T_1 + 1 \leq k \leq T_2, \quad (13)$$

and $X_k = 0$ for $T_2 + 1 \leq k \leq n - 1$ if $T_2 < n - 1$.

- 3rd phase:

$$X_n = \begin{cases} A_3 & \text{if } Y_{T_2} < -A_2/2 \text{ or } (Y_{T_1} < 0 \text{ and } \\ & Y_{T_2} < A_2/2), \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

Consider now the case where $T_1 < n - 1$ and $T_2 > n - 1$.

- 1st phase: $X_k = A_1, 1 \leq k \leq T_1$.
- 2nd phase:

$$X_k = \begin{cases} 0 & \text{if } Y_{T_1} \geq 0, \\ A_2 & \text{if } Y_{T_1} < 0, \end{cases} \quad T_1 + 1 \leq k \leq n - 1. \quad (15)$$

- 3rd phase: $X_n = \tilde{A}_3$.

Finally, consider the case $T_1 \geq n - 1$.

- 1st phase:

$$X_k = A_1, \quad 1 \leq k \leq n - 1. \quad (16)$$

- 2nd phase: $X_n = \tilde{A}_3$.

It is readily verified that the different cases are consistent and that X^n satisfies the required measurability condition (i.e., that each X_k depends causally on the feedback symbols).

The amplitudes are chosen as follows.

$$A_1 = \sqrt{\alpha_1 \rho n P}, \quad A_2 = \sqrt{2\alpha_2 \rho n P} e^{\frac{\alpha_1 \rho n P}{4}}, \quad (17)$$

$$A_3 = \sqrt{2\alpha_3 n P} e^{\frac{\alpha_2 \rho n P}{8}} e^{\frac{\alpha_1 \rho n P}{2}}, \quad \tilde{A}_3 = e^{\frac{n}{2}(\log \frac{1}{1-\rho} - \delta)}, \quad (18)$$

where the constants $\alpha_1, \alpha_2, \alpha_3$ may be any positive numbers satisfying

$$\alpha_1 + \alpha_2 + \alpha_3 < 1, \quad (19)$$

and where δ is arbitrary in the interval $(0, -\log(1 - \rho))$.

The decoding rule is

$$\varphi(Y^n, V^n) = \begin{cases} \begin{cases} \text{if } (T_2 > n - 1, Y_n \geq 0) \text{ or} \\ (T_2 \leq n - 1, Y_n \geq A_3/2) \text{ or} \\ (T_2 \leq n - 1, Y_n \geq -A_3/2, \\ Y_{T_2} \geq A_2/2) \text{ or } (T_2 \leq n - 1, \\ Y_n \geq -A_3/2, Y_{T_2} \geq -A_2/2, \\ Y_{T_1} \geq 0), \end{cases} \\ 0 \\ 1 \end{cases} \text{ otherwise.} \quad (20)$$

Next, we analyze the probability of error. Let D denote the event that $T_2 \leq n - 1$. Conditioning on the complement of D ,

$$\begin{aligned} \Pr(\text{error}|D^c) &= \Pr(Y_n < 0|D^c) \\ &= \mathcal{Q}(\tilde{A}_3). \end{aligned} \quad (21)$$

Conditioning on D ,

$$\begin{aligned} \Pr(\text{error}|D) &= \Pr(Y_n < -A_3/2|D) \\ &+ \Pr(Y_{T_2} < -A_2/2, Y_n \in [-A_3/2, A_3/2]|D) \\ &+ \Pr(Y_{T_1} < 0, Y_{T_2} \in [-A_2/2, A_2/2], \\ &Y_n \in [-A_3/2, A_3/2]|D). \end{aligned} \quad (22)$$

For the first probability on the RHS of (22),

$$\Pr(Y_n < -A_3/2|D) \leq \mathcal{Q}(A_3/2). \quad (23)$$

For the second probability on the RHS of (22),

$$\begin{aligned} & \Pr(Y_{T_2} < -A_2/2, Y_n \in [-A_3/2, A_3/2]|D) \\ & \leq \Pr(Y_n < A_3/2|Y_{T_2} < -A_2/2, D) \\ & = \mathcal{Q}(A_3/2). \end{aligned} \quad (24)$$

For the third probability on the RHS of (22),

$$\begin{aligned} & \Pr(Y_{T_1} < 0, Y_{T_2} \in [-A_2/2, A_2/2], Y_n \in [-A_3/2, A_3/2]|D) \\ & \leq \Pr(Y_n < A_3/2|Y_{T_1} < 0, Y_{T_2} \in [-A_2/2, A_2/2], D) \\ & = \mathcal{Q}(A_3/2). \end{aligned} \quad (25)$$

Thus, the overall probability of error can be bounded as

$$\begin{aligned} \Pr(\text{error}) &= \Pr(\text{error}|D^c) \Pr(D^c) + \Pr(\text{error}|D) \Pr(D) \\ &\leq \mathcal{Q}(\tilde{A}_3) + 3\mathcal{Q}(A_3/2). \end{aligned} \quad (26)$$

Recalling (18) and (1), it follows from (26) that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log(-\log \Pr(\text{error})) \geq -\log(1 - \rho) - \delta. \quad (27)$$

Letting $\delta \rightarrow 0$ gives the desired lower bound.

It remains to verify that the power constraint is satisfied. Conditioning on D and computing the power in the three phases separately, we have

$$\begin{aligned} \mathbb{E}\left[\sum_{k=1}^{T_1} X_k^2 \middle| D\right] \Pr(D) &= A_1^2 \mathbb{E}[T_1|D] \Pr(D) \\ &\leq A_1^2 \mathbb{E}[T_1] \\ &= \alpha_1 nP, \end{aligned} \quad (28)$$

and

$$\begin{aligned} & \mathbb{E}\left[\sum_{k=T_1+1}^{T_2} X_k^2 \middle| D\right] \Pr(D) \\ &= A_2^2 \mathbb{E}[T_2 - T_1|D, Y_{T_1} < 0] \Pr(Y_{T_1} < 0|D) \Pr(D) \\ &= A_2^2 \mathcal{Q}(A_1) \mathbb{E}[T_2 - T_1|D] \Pr(D) \\ &\leq A_2^2 \mathcal{Q}(A_1) \mathbb{E}[T_2 - T_1] \\ &\leq \frac{A_2^2}{2\rho} e^{-A_1^2/2} \\ &= \alpha_2 nP, \end{aligned} \quad (29)$$

and

$$\begin{aligned} & \mathbb{E}[X_n^2|D] \\ &= A_3^2 (\Pr(Y_{T_2} < -A_2/2|Y_{T_1} \geq 0, D) \Pr(Y_{T_1} \geq 0|D) \\ & \quad + \Pr(Y_{T_2} < A_2/2|Y_{T_1} < 0, D) \Pr(Y_{T_1} < 0|D)) \\ &\leq A_3^2 \mathcal{Q}(A_2/2) (\Pr(Y_{T_1} \geq 0|D) + \Pr(Y_{T_1} < 0|D)) \\ &= A_3^2 \mathcal{Q}(A_2/2) \\ &\leq \frac{A_3^2}{2} e^{-A_2^2/8} \\ &= \alpha_3 nP. \end{aligned} \quad (30)$$

Letting B denote the event that $T_1 < n - 1$ and conditioning on the complement of D , we have

$$\begin{aligned} & \mathbb{E}\left[\sum_{k=1}^n X_k^2 \middle| D^c\right] \\ &= \mathbb{E}[X_n^2|D^c] + \mathbb{E}\left[\sum_{k=1}^{n-1} X_k^2 \middle| D^c\right] \\ &= \tilde{A}_3^2 + (n-1)A_1^2 \Pr(B^c|D^c) + A_1^2 \mathbb{E}[T_1|D^c, B] \Pr(B|D^c) \\ & \quad + A_2^2 \mathbb{E}[n-1-T_1|D^c, B, Y_{T_1} < 0] \Pr(B, Y_{T_1} < 0|D^c) \\ &\leq \tilde{A}_3^2 + (2n-3)A_1^2 + (n-2)A_2^2 \Pr(Y_{T_1} < 0|B, D^c) \\ &= \tilde{A}_3^2 + (2n-3)A_1^2 + (n-2)A_2^2 \mathcal{Q}(A_1). \end{aligned} \quad (31)$$

Noting that

$$\begin{aligned} \Pr(D^c) &= (1 - \rho)^{n-1} + (n-1)\rho(1 - \rho)^{n-2} \\ &= e^{-n(\log(\frac{1}{1-\rho}) - \delta_n)}, \end{aligned} \quad (32)$$

where $\delta_n \rightarrow 0$ as $n \rightarrow \infty$, it follows from (17), (18), (31), and (1) that

$$\lim_{n \rightarrow \infty} \mathbb{E}\left[\sum_{k=1}^n X_k^2 \middle| D^c\right] \Pr(D^c) = 0. \quad (33)$$

Combining (28), (29), (30), and (33) shows that

$$\mathbb{E}\left[\sum_{k=1}^n X_k^2\right] \leq (\alpha_1 + \alpha_2 + \alpha_3)nP + \varepsilon_n, \quad (34)$$

where $\varepsilon_n \rightarrow 0$ as $n \rightarrow \infty$. Thus, in view of (19), the power constraint is satisfied for all sufficiently large blocklengths n . \square

IV. POSITIVE RATES

Consider the problem of transmitting one of e^{nR} equiprobable messages in n uses of the channel described in Section III. Formally, a rate- R blocklength- n code for this channel comprises e^{nR} random n -tuples $X^n(1), \dots, X^n(e^{nR})$ such that each $X_k(m)$ is \mathcal{F}_{k-1} -measurable, and a (Borel-measurable) decoder mapping

$$\varphi: \mathbb{R}^n \times \{0, 1\}^n \rightarrow \{1, \dots, e^{nR}\}. \quad (35)$$

The probability of error is

$$\Pr(\text{error}) = \Pr(\varphi(Y^n, V^n) \neq M), \quad (36)$$

where conditional on $M = m$ the input to the channel is $X^n(m)$, and $\Pr(M = m) = e^{-nR}$ for all m . We impose the power constraint

$$\mathbb{E}\left[\sum_{k=1}^n X_k(m)^2\right] \leq nP, \quad m \in \{1, \dots, e^{nR}\}, \quad (37)$$

where P is some given positive number. Let $C(P)$ denote the capacity of the channel (in nats per channel use) subject to the power constraint (37), i.e.,

$$C(P) = \frac{1}{2} \log(1 + P), \quad (38)$$

and let $P_e(n, R)$ denote the least probability of error of any rate- R blocklength- n code satisfying (37).

Theorem IV.1. For any $\rho \in (0, 1)$ and any $P > 0$,

- 1) if $R > \rho C(P)$, then $\overline{\lim}_{n \rightarrow \infty} -\frac{1}{n} \log P_e(n, R) < \infty$;
- 2) if $R < \rho C(P)$, then $\underline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log(-\log P_e(n, R)) > 0$.

Note that the second-order error exponent in the case $0 < R < \rho C(P)$ is finite. Indeed, it is upper-bounded by $-\log(1 - \rho)$ on account of Theorem III.1.

Proof. We begin with the case $R > \rho C(P)$. Select $\varepsilon > 0$ small enough so that $R > (\rho + \varepsilon)C(P)$. Assume that the transmitter knows the positions of the feedback symbols in advance, i.e., V^n is revealed noncausally. This can only decrease the probability of error. It is shown in [4] that when a fixed fraction f of the output symbols is fed back to the transmitter (the positions of the symbols that are fed back are fixed and known to the transmitter and receiver), and R is larger than f times the capacity, then the probability of error cannot decay faster than exponentially. To apply this result here, let K denote the event that the number of symbols fed back in the first n channel uses is no more than $(\rho + \varepsilon)n$, i.e., $K = \{V_1 + \dots + V_n \leq (\rho + \varepsilon)n\}$. The maximum amount of power we can use conditional on K is $P_K = P/\Pr(K)$, and $P_K \rightarrow P$ as $n \rightarrow \infty$ because $\Pr(K) \rightarrow 1$ by the weak law of large numbers. Thus, by the continuity of $C(P)$ in P , we have $R > (\rho + \varepsilon)C(P_K)$ for all sufficiently large n , and for such n the conditional probability of error $\Pr(\text{error}|K)$ cannot decay faster than exponentially on account of the aforementioned result in [4]. The desired result now follows by noting that

$$\Pr(\text{error}) \geq \Pr(\text{error}|K) \Pr(K). \quad (39)$$

Next, consider the case where $R < \rho C(P)$. We exhibit a coding scheme that satisfies the power constraint (37) and achieves a positive second-order error exponent. Select $\varepsilon > 0$ small enough so that $R < (\rho - \varepsilon)C(P)$. The total blocklength of the scheme is $n + 1 + \alpha n$, where $\alpha > 0$ is arbitrary, and the rate is $R/(1 + 1/n + \alpha)$. (Later on we let α tend to zero.) Choose a codebook \mathcal{C} of size e^{nR} and blocklength $(\rho - \delta)n$, where δ is arbitrary in the interval $(0, \varepsilon)$, that satisfies the power constraint

$$\sum_{k=1}^{(\rho-\delta)n} x_k^2 \leq (\rho - \delta)nP, \quad x^{(\rho-\delta)n} \in \mathcal{C}. \quad (40)$$

Let m be the message to be transmitted. In the first n channel uses we repeat each component of the m -th codeword in \mathcal{C} until feedback occurs and pad the remaining channel uses (if any) with zeros. Let A be the event that at least $(\rho - \delta)n$ symbols are fed back in the first n channel uses, i.e., $A = \{V_1 + \dots + V_n \geq (\rho - \delta)n\}$.

If A occurs, then the transmitter decodes the codeword based on the symbols that were fed back using an optimal decoder for \mathcal{C} . If the decoding is successful, i.e., if the message produced by an optimal decoder for \mathcal{C} is m , then

the remaining $1 + \alpha n$ channel uses are padded with zeros. Otherwise a flag is sent in the $(n + 1)$ -st channel use:

$$X_{n+1} = \sqrt{P/(2P_e(\mathcal{C}))}, \quad (41)$$

where $P_e(\mathcal{C})$ is the maximum probability of error of the code \mathcal{C} . After sending the flag, the transmitter retransmits the message using a code $\tilde{\mathcal{C}}$ of size e^{nR} and blocklength αn that satisfies the power constraint

$$\sum_{k=1}^{\alpha n} x_k^2 \leq \frac{\alpha n P}{2P_e(\tilde{\mathcal{C}})}, \quad x^{\alpha n} \in \tilde{\mathcal{C}}. \quad (42)$$

If A does not occur, i.e., if fewer than $(\rho - \delta)n$ symbols are fed back in the first n channels uses, then the transmitter retransmits the message using a third code \mathcal{C}' of size e^{nR} and blocklength $1 + \alpha n$ that satisfies the power constraint

$$\sum_{k=1}^{1+\alpha n} x_k^2 \leq \frac{(1 + \alpha n)P}{2\Pr(A^c)}, \quad x^{1+\alpha n} \in \mathcal{C}'. \quad (43)$$

The decoding procedure is described next. If A occurs, then the receiver compares Y_{n+1} to the threshold

$$\Upsilon = \frac{1}{2} \sqrt{P/(2P_e(\mathcal{C}))}. \quad (44)$$

If $Y_{n+1} > \Upsilon$, then the receiver ignores the first n outputs and decodes the message based on the last αn outputs using an optimal decoder for $\tilde{\mathcal{C}}$. If $Y_{n+1} \leq \Upsilon$, then the receiver decodes the message based on the first $(\rho - \delta)n$ outputs that were fed back to the transmitter using an optimal decoder for \mathcal{C} . If A does not occur, then the receiver decodes the message based on the last $1 + \alpha n$ outputs using an optimal decoder for \mathcal{C}' .

To analyze the probability of error, we expand it as

$$\Pr(\text{error}) = \Pr(\text{error}|A) \Pr(A) + \Pr(\text{error}|A^c) \Pr(A^c). \quad (45)$$

Let E be the event that A occurs and that decoding \mathcal{C} based on the first $(\rho - \delta)n$ feedback symbols fails. Then

$$\begin{aligned} \Pr(\text{error}|A) &= \Pr(\text{error}|E) \Pr(E|A) \\ &\quad + \Pr(\text{error}|A \setminus E) \Pr(A \setminus E|A). \end{aligned} \quad (46)$$

Observe that

$$\begin{aligned} \Pr(\text{error}|E) &= \Pr(\text{error}|Y_{n+1} > \Upsilon, E) \Pr(Y_{n+1} > \Upsilon|E) \\ &\quad + \Pr(\text{error}|Y_{n+1} \leq \Upsilon, E) \Pr(Y_{n+1} \leq \Upsilon|E) \\ &\leq \Pr(\text{error}|Y_{n+1} > \Upsilon, E) + \Pr(Y_{n+1} \leq \Upsilon|E) \\ &\leq P_e(\tilde{\mathcal{C}}) + \mathcal{Q}(\Upsilon), \end{aligned} \quad (47)$$

where $P_e(\tilde{\mathcal{C}})$ denotes the maximal probability of error of the code $\tilde{\mathcal{C}}$. Observe further that

$$\begin{aligned} \Pr(\text{error}|A \setminus E) &= \Pr(\text{error}|A \setminus E, Y_{n+1} > \Upsilon) \Pr(Y_{n+1} > \Upsilon|A \setminus E) \\ &\leq \Pr(Y_{n+1} > \Upsilon|A \setminus E) = \mathcal{Q}(\Upsilon). \end{aligned} \quad (48)$$

Substituting (48) and (47) into (46), it follows that

$$\Pr(\text{error}|A) \leq P_e(\tilde{\mathcal{C}}) + \mathcal{Q}(\Upsilon). \quad (49)$$

If A does not occur, then an error occurs only if decoding of \mathcal{C}' fails and hence

$$\Pr(\text{error}|A^c) \leq P_e(\mathcal{C}'), \quad (50)$$

where $P_e(\mathcal{C}')$ is the maximal probability of error of \mathcal{C}' . Substituting (49) and (50) into (45), it follows that

$$\Pr(\text{error}) \leq P_e(\tilde{\mathcal{C}}) + P_e(\mathcal{C}') + \mathcal{Q}(\Upsilon). \quad (51)$$

We next show that each of the three probabilities on the RHS of (51) decay double-exponentially in n . We begin with $\mathcal{Q}(\Upsilon)$. Note that the rate of \mathcal{C} is $R/(\rho - \delta)$, and

$$\begin{aligned} \frac{R}{\rho - \delta} &< \frac{\rho - \varepsilon}{\rho - \delta} C(P) \\ &< C(P), \end{aligned} \quad (52)$$

where the last inequality holds because $0 < \delta < \varepsilon$. Since by (40) each codeword in \mathcal{C} can have power up to P , the rate that we want for \mathcal{C} is strictly below capacity. Consequently, we can find a \mathcal{C} (more precisely, a sequence of \mathcal{C} 's indexed by n) with

$$P_e(\mathcal{C}) \leq e^{-n\zeta}, \quad (53)$$

for some $\zeta > 0$. From (44) and (1) it thus follows that

$$\mathcal{Q}(\Upsilon) \leq \frac{1}{2} e^{-\frac{P}{16} e^{n\zeta}}. \quad (54)$$

From [3, Eq. (77)] it follows that both $P_e(\tilde{\mathcal{C}})$ and $P_e(\mathcal{C}')$ decay double-exponentially with n provided that the power constraints on $\tilde{\mathcal{C}}$ and \mathcal{C}' grow exponentially with n . This is true for $\tilde{\mathcal{C}}$ by (42) and (53), and for \mathcal{C}' by (43) and the fact that

$$\Pr(A^c) \leq e^{-2\delta^2 n}, \quad (55)$$

which follows from Hoeffding's Inequality. We conclude that the overall probability of error decays double-exponentially in n for every $\alpha > 0$. Since the rate of the scheme approaches $R/(1 + \alpha)$, and since $\alpha > 0$ is arbitrary, the proof is complete once we show that the power constraint is satisfied.

To this end, we define

$$T_\ell = \min\{k > T_{\ell-1} : V_k = 1\}, \quad \ell \geq 1, \quad (56)$$

where T_0 is taken to be zero. Denote the m -th codeword in \mathcal{C} by $x^n(m)$. The power in the first n channel uses can be bounded as

$$\begin{aligned} \mathbb{E}\left[\sum_{k=1}^n X_k^2(m)\right] &\leq \mathbb{E}\left[\sum_{\ell=1}^{(\rho-\delta)n} (T_\ell - T_{\ell-1}) x_\ell^2(m)\right] \\ &= \sum_{\ell=1}^{(\rho-\delta)n} \mathbb{E}[T_\ell - T_{\ell-1}] x_\ell^2(m) \\ &= \frac{1}{\rho} \sum_{\ell=1}^{(\rho-\delta)n} x_\ell^2(m) < nP, \end{aligned} \quad (57)$$

where the second inequality follows from (40). To bound the power in the remaining $1 + \alpha n$ channel uses, we first condition on A to obtain

$$\begin{aligned} &\mathbb{E}\left[\sum_{k=n+1}^{n+1+\alpha n} X_k^2(m) \middle| A\right] \\ &= \mathbb{E}\left[X_{n+1}^2(m) \middle| A\right] + \mathbb{E}\left[\sum_{k=n+2}^{n+1+\alpha n} X_k^2(m) \middle| A\right] \\ &\leq \Pr(E|A) \left(\frac{P}{2P_e(\mathcal{C})} + \frac{\alpha n P}{2P_e(\mathcal{C})}\right) \\ &\leq \frac{P}{2}(1 + \alpha n). \end{aligned} \quad (58)$$

Conditioning on the complement of A ,

$$\mathbb{E}\left[\sum_{k=n+1}^{n+1+\alpha n} X_k^2(m) \middle| A^c\right] \leq \frac{(1 + \alpha n)P}{2\Pr(A^c)}. \quad (59)$$

Combining (59), (58), and (57) it follows that

$$\mathbb{E}\left[\sum_{k=1}^{n+1+\alpha n} X_k^2(m)\right] < (n + 1 + \alpha n)P, \quad (60)$$

i.e., the power constraint is indeed satisfied. \square

V. CONCLUDING REMARKS

We conclude this paper with the following result for the case when the receiver does not know which symbols are fed back to the transmitter. A proof is given in [9].

Theorem V.1. *If $\rho > 1/2$ and $P > 0$, then one bit can be transmitted over the channel described in Section III with probability of error decaying double-exponentially in the blocklength even when the receiver does not know which output symbols are fed back to the transmitter.*

REFERENCES

- [1] R. G. Gallager and B. Nakiboğlu, "Variations on a theme by Schalkwijk and Kailath," *IEEE Trans. on Inf. Theory*, vol. 56, no. 1, pp. 6–18, Jan. 2010.
- [2] A. Khisti and A. Lapidoth, "Multiple access channels with intermittent feedback and side information," in Proceedings IEEE International Symposium on Information Theory (ISIT), Istanbul, Turkey, July 7–12, 2013.
- [3] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell System Techn. J.*, vol. 38, pp. 611–656, 1959.
- [4] M. Agarwal, D. Guo, M. L. Honig, "Error exponent for Gaussian channels with partial sequential feedback," *IEEE Trans. on Inf. Theory*, vol. 59, no. 8, pp. 4757–4766, Aug. 2013.
- [5] R. Mirghaderi, A. J. Goldsmith, T. Weissman, "Achievable error exponents in the Gaussian channel with rate-limited feedback," *IEEE Trans. on Inf. Theory*, vol. 59, no. 12, pp. 8144–8156, Dec. 2013.
- [6] Y.-H. Kim, A. Lapidoth, T. Weissman, "Error Exponents for the Gaussian channel with active noisy feedback," *IEEE Trans. on Inf. Theory*, vol. 57, no. 3, pp. 1223–1236, Mar. 2011.
- [7] Y.-H. Kim, A. Lapidoth, T. Weissman, "On the Reliability of Gaussian Channels with Noisy Feedback," Forty-Fourth Annual Allerton Conference, Illinois, USA, Sep. 2006.
- [8] Y.-H. Kim, A. Lapidoth, and T. Weissman, "Bounds on the error exponent of the AWGN channel with AWGN-corrupted feedback," Proceedings of the 24th IEEE Convention of Electrical and Electronics Engineers in Israel, pp. 184–188, Israel, Nov. 2006.
- [9] L. Palzer, *MSc Thesis*, ETH Zurich, July 2014.