Signaling over the Gaussian Channel with Intermittent Feedback

Lars Palzer
Technische Universität München
Email: lars.palzer@tum.de

***Abstract*—Optimal signaling is studied over a power-limited Gaussian channel with intermittent feedback, where a random mechanism—whose outcome is unknown to the receiver—determines whether or not the output symbol is fed back to the encoder. If the output symbols are fed back with probability smaller than one half, then even the two-messages error probability cannot decay faster than exponentially in the blocklength. But if this probability is greater than one half, then, even for some positive rates, a doubly exponential decay is achievable.**

I. Introduction

There has recently been increasing interest in the asymptotic behavior of the best achievable error probability over the memoryless Gaussian channel with imperfect feedback. Different feedback models have been studied, such as rate-limited [1], noisy [2]–[4] or partial feedback [5]. In [6] we considered the case introduced in [7] where the feedback is "intermittent," i.e., where each channel output is fed back with probability ρ. Unlike the present paper, [6] assumes that the receiver knows which outputs were fed back. Under this assumption, the optimal error probability for the transmission of a single bit decays double-exponentially in the blocklength with a second order error exponent of $-\log\rho$. For positive rates, a double-exponential decay is possible for $R < \rho C$ and is impossible for $R > \rho C$. Here R is the transmission rate and C the capacity of the forward channel.

In this paper we study the case where the receiver is not told which output symbols were fed back. We show that for $\rho < 1/2$, the two-messages probability of error cannot decay faster than exponentially in the blocklength, and we provide an upper bound on the best achievable error exponent (Theorem 1 ahead). For $\rho > 1/2$, however, a double-exponential decay is achievable: see Theorem 5, which also provides a lower bound on the iterated error exponent. Surprisingly, the error exponent does not tend to infinity as ρ increases to 1/2.

We also extend our result to positive rates and show that for $\rho > 1/2$ a double-exponential decay is achievable for some positive rates (Theorem 6 ahead).

II. Notation and Preliminaries

Boldface letters denote tuples, and subscripts denote their elements, e.g., $\mathbf{x} = (x_1, \ldots, x_n)$. The set $\{0, 1\}$ is denoted $\mathcal{S}$, and if $\mathbf{s} \in \mathcal{S}^n$ is a binary n-tuple, then $\mathsf{w}(\mathbf{s})$ denotes the number of ones among its elements. If $\mathbf{x}$ and $\mathbf{s}$ are as above, then $\mathbf{x}_{\mathbf{s}}$ is the tuple whose $\mathsf{w}(\mathbf{s})$ elements comprise those elements of $\mathbf{x}$ whose index k is such that $s_k = 1$. For example, if $\mathbf{x} = (1, 2, 3, 4, 5)$ and $\mathbf{s} = (0, 1, 1, 0, 1)$, then $\mathbf{x}_{\mathbf{s}} = (2, 3, 5)$. The all-ones tuple is denoted $\mathbf{1}$, and if $\mathbf{s}$ is a binary tuple then $\mathbf{s}^{\mathrm{c}}$ stands for the componentwise difference $\mathbf{1} - \mathbf{s}$. The Euclidean norm and inner product are denoted $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$. The complement of a set or event E is denoted by E^{c}, and the cardinality of a finite set $\mathcal{A}$ is denoted by $|\mathcal{A}|$. The expectation with respect to a probability measure P is denoted $\mathsf{E}_{\mathsf{P}}[\cdot]$. All logarithms are natural logarithms.

We denote by $Q(\xi)$ the probability that a standard Gaussian exceeds ξ. It is bounded by

$$\frac{1}{\sqrt{2\pi\xi^2}}\, e^{-\xi^2/2}\Big(1 - \frac{1}{\xi^2}\Big) < Q(\xi) \le \frac{1}{2}\, e^{-\xi^2/2}, \quad \xi > 0. \tag{1}$$

III. Problem Statement

A message m, which is drawn uniformly from the set $\mathcal{M}$, is to be transmitted over a channel whose time-k output is

$$Y_k = x_k + Z_k, \quad k \in \{1, 2, \ldots\} \tag{2}$$

where x_k is the time-k channel input and $\{Z_k\}_{k=1}^{\infty}$ are independent and identically distributed (IID) Gaussian noise samples of zero mean and variance $\sigma^2 > 0$. The transmitter receives intermittent feedback in the sense that Y_k is revealed strictly-causally to the encoder if, and only if, $S_k = 1$, where $\{S_k\}_{k=1}^{\infty}$ is independent of the message and channel noise; it is IID Bernoulli with $\Pr(S_k = 1) = \rho$ for $\rho \in (0, 1)$; and it is not revealed to the receiver.

More precisely, let $(\Omega, \mathcal{F}, \mathrm{P})$ 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 $s^k \in \mathcal{S}^k$ we have $\{S^k = s^k\} \cap A \in \sigma(S^k, Y_{i(1)}, \ldots, Y_{i(\ell)})$ where $i(1), \ldots, i(\ell)$ are those indices $i \in \{1, \ldots, k\}$ for which $s_i = 1$. A $|\mathcal{M}|$-message blocklength-n code for this channel comprises $|\mathcal{M}|$ random n-tuples $\mathbf{X}(m)$, $m \in \mathcal{M}$, such that each $X_k(m)$ is $\mathcal{F}_{k-1}$-measurable. We write $\mathbf{X}(m, \mathbf{Y}_{\mathbf{S}}, \mathbf{S})$ for the n-tuple of channel inputs that the encoder produces to convey the message m when the symbols $\mathbf{Y}_{\mathbf{S}}$ are fed back and $\mathbf{Y}_{\mathbf{S}^{\mathrm{c}}}$ are not.

A decoder φ is a (Borel-measurable) mapping

$$\varphi\colon \mathbb{R}^n \to \mathcal{M}. \tag{3}$$

We denote the decision region for each message $m \in \mathcal{M}$ by $\mathcal{D}_m = \{\mathbf{y} \in \mathbb{R}^n \colon \varphi(\mathbf{y}) = m\}$, and note that by (3), $\bigcup_{m\in\mathcal{M}} \mathcal{D}_m = \mathbb{R}^n$. Let P denote the joint law of $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{S}$ induced by the blocklength-n coding scheme and let $p(\mathbf{y}|\mathbf{s})$ and $p(\mathbf{y}_{\mathbf{s}}|\mathbf{s})$ denote the conditional densities of $\mathbf{Y}$ and $\mathbf{Y}_{\mathbf{s}}$

(the feedback) given that $\mathbf{S} = \mathrm{s}$, respectively. We denote the conditional versions of the joint law and conditional densities conditioned on $M = m$ by P_m, $p_m(\mathbf{y}|\mathrm{s})$ and $p_m(\mathbf{y}_\mathrm{s}|\mathrm{s})$, e.g., $\mathsf{P}_m(\cdot) = \mathsf{P}(\cdot|M = m)$. The probability of error is

$$\mathsf{P}(\text{error}) = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \mathsf{P}_m(\mathbf{Y} \notin \mathcal{D}_m). \tag{4}$$

We impose the average power constraint for some $P > 0$:

$$\mathsf{E}\left[\sum_{k=1}^{n} X_k(m)^2\right] \le nP, \quad m \in \mathcal{M}. \tag{5}$$

IV. Inachievability for $\rho < 1/2$

Theorem 1. *Let $\mathcal{M} = \{0, 1\}$ and let $p_\mathrm{e}\left(P/\sigma^2, n\right)$ denote the least probability of error of any blocklength-n coding scheme satisfying* (5). *For $\rho \in (0, 1/2)$,*

$$\limsup_{n\to\infty} -\frac{1}{n} \log p_\mathrm{e}\left(P/\sigma^2, n\right) \le \frac{2\left(\sqrt{\rho P + \sigma^2/2} + \sqrt{(1-\rho)P}\right)^2}{\sigma^2}. \tag{6}$$

The intuition is that for $\rho < 1/2$, with high probability less than half of the outputs are fed back. In this case, it may happen that the part of the sequence which is fed back contains little noise and the part of the sequence which is not contains a large amount of noise, e.g., on the order of the signal power. The probability for this amount of noise decreases exponentially in the blocklength. The noisy part of the received sequence can then look similar to a part of the codeword for a different message. Since less than half of the symbols are fed back, these two different parts can be of the same length. It seems reasonable that an error occurs in such a situation and thus that the error probability only decreases exponentially in the blocklength.

In the remainder of this section, we provide a proof for Theorem 1. Fix some $\alpha > 0$, $\beta > 0$, and $0 < \delta < 1/2 - \rho$, and define

$$\mathcal{T}_y := \left\{\mathbf{y} \in \mathbb{R}^n \colon \|\mathbf{y}\|^2 < n\alpha^2\right\}, \tag{7}$$

$$\mathcal{G} := \left\{\mathrm{s} \in \mathcal{S}^n \colon \left|\frac{\mathsf{w}(\mathrm{s})}{n} - \rho\right| \le \delta\right\}. \tag{8}$$

where all $\mathrm{s} \in \mathcal{G}$ satisfy $\mathsf{w}(\mathrm{s}) < n/2$. For any $\mathrm{s} \in \mathcal{G}$, define

$$\mathcal{B}(\mathrm{s}) := \left\{\mathrm{s}' \in \mathcal{S}^n \colon \mathsf{w}(\mathrm{s}') = \mathsf{w}(\mathrm{s}) \text{ and } \langle \mathrm{s}, \mathrm{s}' \rangle = 0\right\}. \tag{9}$$

And for any $k \le n/2$, define

$$\mathcal{G}_k := \left\{(\mathrm{s}, \mathrm{s}') \in \mathcal{S}^n \times \mathcal{S}^n \colon \mathsf{w}(\mathrm{s}) = \mathsf{w}(\mathrm{s}') = k, \langle \mathrm{s}, \mathrm{s}' \rangle = 0\right\}. \tag{10}$$

Note that if $\mathsf{w}(\mathrm{s})$ and $\mathsf{w}(\mathrm{s}')$ are equal to such a k, then

$$\mathrm{s}' \in \mathcal{B}(\mathrm{s}) \iff \mathrm{s} \in \mathcal{B}(\mathrm{s}') \iff (\mathrm{s}, \mathrm{s}') \in \mathcal{G}_k. \tag{11}$$

Finally, given some encoder, define for every $\mathrm{s} \in \mathcal{S}^n$,

$$\mathcal{T}_m(\mathrm{s}) := \left\{\mathbf{y} \in \mathbb{R}^n \colon \|\mathbf{x}_{\mathrm{s}^\mathrm{c}}(m, \mathbf{y}_\mathrm{s}, \mathrm{s})\|^2 < n\beta^2\right\}. \tag{12}$$

We next define a new probability density;

Definition 2. For any $\mathrm{s} \in \mathcal{G}$, any $\tilde{\mathrm{s}} \in \mathcal{B}(\mathrm{s})$, and $\mathbf{y} \in \mathbb{R}^n$,

$$q(\mathbf{y}|\mathrm{s}, \tilde{\mathrm{s}}) := p_0(\mathbf{y}_\mathrm{s}|\mathrm{s})\, p_1(\mathbf{y}_{\tilde{\mathrm{s}}}|\tilde{\mathrm{s}})\, g(\mathbf{y}_\mathbf{r}|\mathbf{r}), \tag{13}$$

where $\mathbf{r} := \mathbf{1} - \mathrm{s} - \tilde{\mathrm{s}}$, where $g(\mathbf{y}_\mathbf{r}|\mathbf{r}) := (2\pi\sigma^2)^{-j/2}\, e^{-\frac{\|\mathbf{y}_\mathbf{r}\|^2}{2\sigma^2}}$, and where $j = \mathsf{w}(\mathbf{r})$.

The following lemma, whose proof is omitted, lower-bounds the conditional channel output densities $p_0(\mathbf{y}|\mathrm{s})$ and $p_1(\mathbf{y}|\tilde{\mathrm{s}})$.

Lemma 3. *For any pair $(\mathrm{s}, \tilde{\mathrm{s}})$ with $\mathrm{s} \in \mathcal{G}$ and $\tilde{\mathrm{s}} \in \mathcal{B}(\mathrm{s})$,*

$$p_0(\mathbf{y}|\mathrm{s}) \ge q(\mathbf{y}|\mathrm{s}, \tilde{\mathrm{s}})\; e^{-n\frac{(\alpha+\beta)^2}{2\sigma^2}}, \quad \mathbf{y} \in \mathcal{T}_y \cap \mathcal{T}_0(\mathrm{s}), \tag{14}$$

$$p_1(\mathbf{y}|\tilde{\mathrm{s}}) \ge q(\mathbf{y}|\mathrm{s}, \tilde{\mathrm{s}})\; e^{-n\frac{(\alpha+\beta)^2}{2\sigma^2}}, \quad \mathbf{y} \in \mathcal{T}_y \cap \mathcal{T}_1(\tilde{\mathrm{s}}). \tag{15}$$

Fixing $\mathrm{s} \in \mathcal{G}$ and averaging (14) over all $\tilde{\mathrm{s}} \in \mathcal{B}(\mathrm{s})$, yields

$$p_0(\mathbf{y}|\mathrm{s}) \ge \frac{1}{|\mathcal{B}(\mathrm{s})|} \sum_{\tilde{\mathrm{s}} \in \mathcal{B}(\mathrm{s})} q(\mathbf{y}|\mathrm{s}, \tilde{\mathrm{s}})\, e^{-n\frac{(\alpha+\beta)^2}{2\sigma^2}}, \qquad \mathbf{y} \in \mathcal{T}_y \cap \mathcal{T}_0(\mathrm{s}). \tag{16}$$

Hence,

$$\begin{aligned}\mathsf{P}_0(\mathbf{Y} \notin \mathcal{D}_0) &\ge \sum_{\mathrm{s} \in \mathcal{G}} \mathsf{P}(\mathbf{S} = \mathrm{s}) \int_{\mathbf{y} \in \mathcal{D}_1 \cap \mathcal{T}_y \cap \mathcal{T}_0(\mathrm{s})} p_0(\mathbf{y}|\mathrm{s})\, \mathrm{d}\mathbf{y} \\ &\ge e^{-n\frac{(\alpha+\beta)^2}{2\sigma^2}} \sum_{\mathrm{s} \in \mathcal{G}} \sum_{\tilde{\mathrm{s}} \in \mathcal{B}(\mathrm{s})} \frac{\mathsf{P}(\mathbf{S} = \mathrm{s})}{|\mathcal{B}(\mathrm{s})|} \\ &\qquad \cdot \int_{\mathbf{y} \in \mathcal{D}_1 \cap \mathcal{T}_y \cap \mathcal{T}_0(\mathrm{s}) \cap \mathcal{T}_1(\tilde{\mathrm{s}})} q(\mathbf{y}|\mathrm{s}, \tilde{\mathrm{s}})\, \mathrm{d}\mathbf{y},\end{aligned} \tag{17}$$

where the intersection with $\mathcal{T}_1(\tilde{\mathrm{s}})$ can only reduce the integral. Likewise,

$$\begin{aligned}\mathsf{P}_1(\mathbf{Y} \notin \mathcal{D}_1) &\ge e^{-n\frac{(\alpha+\beta)^2}{2\sigma^2}} \sum_{\tilde{\mathrm{s}} \in \mathcal{G}} \sum_{\mathrm{s} \in \mathcal{B}(\tilde{\mathrm{s}})} \frac{\mathsf{P}(\tilde{\mathbf{S}} = \tilde{\mathrm{s}})}{|\mathcal{B}(\tilde{\mathrm{s}})|} \\ &\qquad \cdot \int_{\mathbf{y} \in \mathcal{D}_0 \cap \mathcal{T}_y \cap \mathcal{T}_0(\mathrm{s}) \cap \mathcal{T}_1(\tilde{\mathrm{s}})} q(\mathbf{y}|\mathrm{s}, \tilde{\mathrm{s}})\, \mathrm{d}\mathbf{y}.\end{aligned} \tag{18}$$

Noting that when $\tilde{\mathrm{s}} \in \mathcal{G}$ and $\mathrm{s} \in \mathcal{B}(\tilde{\mathrm{s}})$ we have

$$\mathsf{P}(\tilde{\mathbf{S}} = \tilde{\mathrm{s}}) = \mathsf{P}(\mathbf{S} = \mathrm{s}), \quad |\mathcal{B}(\tilde{\mathrm{s}})| = |\mathcal{B}(\mathrm{s})|, \tag{19}$$

and using (11) one can prove the following identity whose RHS does not depend on $\mathcal{D}_0$ and $\mathcal{D}_1$:

Lemma 4. *The following identity holds true:*

$$\begin{aligned}&\sum_{\mathrm{s} \in \mathcal{G}} \sum_{\tilde{\mathrm{s}} \in \mathcal{B}(\mathrm{s})} \frac{\mathsf{P}(\mathbf{S} = \mathrm{s})}{|\mathcal{B}(\mathrm{s})|} \int_{\mathbf{y} \in \mathcal{D}_1 \cap \mathcal{T}_y \cap \mathcal{T}_0(\mathrm{s}) \cap \mathcal{T}_1(\tilde{\mathrm{s}})} q(\mathbf{y}|\mathrm{s}, \tilde{\mathrm{s}})\, \mathrm{d}\mathbf{y} \\ &+ \sum_{\tilde{\mathrm{s}} \in \mathcal{G}} \sum_{\mathrm{s} \in \mathcal{B}(\tilde{\mathrm{s}})} \frac{\mathsf{P}(\tilde{\mathbf{S}} = \tilde{\mathrm{s}})}{|\mathcal{B}(\tilde{\mathrm{s}})|} \int_{\mathbf{y} \in \mathcal{D}_0 \cap \mathcal{T}_y \cap \mathcal{T}_0(\mathrm{s}) \cap \mathcal{T}_1(\tilde{\mathrm{s}})} q(\mathbf{y}|\mathrm{s}, \tilde{\mathrm{s}})\, \mathrm{d}\mathbf{y} \\ &= \sum_{\mathrm{s} \in \mathcal{G}} \sum_{\tilde{\mathrm{s}} \in \mathcal{B}(\mathrm{s})} \frac{\mathsf{P}(\mathbf{S} = \mathrm{s})}{|\mathcal{B}(\mathrm{s})|} \int_{\mathbf{y} \in \mathcal{T}_y \cap \mathcal{T}_0(\mathrm{s}) \cap \mathcal{T}_1(\tilde{\mathrm{s}})} q(\mathbf{y}|\mathrm{s}, \tilde{\mathrm{s}})\, \mathrm{d}\mathbf{y}.\end{aligned} \tag{20}$$

Combining (17) and (18) using Lemma 4, we obtain that the total probability of error

$$\mathsf{P}(\text{error}) = 1/2 \cdot \mathsf{P}_0(\mathbf{Y} \notin \mathcal{D}_0) + 1/2 \cdot \mathsf{P}_1(\mathbf{Y} \notin \mathcal{D}_1)$$

is lower-bounded by

$$\frac{e^{-n\frac{(\alpha+\beta)^2}{2\sigma^2}}}{2} \sum_{\substack{\mathbf{s}\in\mathcal{G},\\ \tilde{\mathbf{s}}\in\mathcal{B}(\mathbf{s})}} \frac{\mathsf{P}(\mathbf{S}=\mathbf{s})}{|\mathcal{B}(\mathbf{s})|} \int_{\mathbf{y}\in\mathcal{T}_y\cap\mathcal{T}_0(\mathbf{s})\cap\mathcal{T}_1(\tilde{\mathbf{s}})} q(\mathbf{y}|\mathbf{s},\tilde{\mathbf{s}})\,\mathrm{d}\mathbf{y}. \tag{21}$$

To lower bound (21), we define a new joint probability law Q on $\mathbf{Y}$, $\mathbf{S}$, $\tilde{\mathbf{S}}$:

$$\mathsf{Q}(\mathbf{S}=\mathbf{s},\tilde{\mathbf{S}}=\tilde{\mathbf{s}}) \coloneqq \frac{\mathsf{P}(\mathbf{S}=\mathbf{s})1\{\mathbf{s}\in\mathcal{G},\tilde{\mathbf{s}}\in\mathcal{B}(\mathbf{s})\}}{|\mathcal{B}(\mathbf{s})|\mathsf{P}(\mathbf{S}\in\mathcal{G})}, \tag{22}$$

$$\mathsf{Q}(\mathbf{Y}\in\mathcal{A}|\mathbf{S}=\mathbf{s},\tilde{\mathbf{S}}=\tilde{\mathbf{s}}) \coloneqq \int_{\mathbf{y}\in\mathcal{A}} q(\mathbf{y}|\mathbf{s},\tilde{\mathbf{s}})\,\mathrm{d}\mathbf{y},\ \ \mathcal{A}\subseteq\mathbb{R}^n. \tag{23}$$

Note that $\mathbf{S}$ and $\tilde{\mathbf{S}}$ have the same marginal distributions under Q. Rewriting the lower bound in terms of Q,

$$\begin{aligned}
&\mathsf{P}(\text{error}) \\
&\geq \frac{1}{2}e^{-n\frac{(\alpha+\beta)^2}{2\sigma^2}} \sum_{\substack{\mathbf{s}\in\mathcal{G}\\ \tilde{\mathbf{s}}\in\mathcal{B}(\mathbf{s})}} \frac{\mathsf{P}(\mathbf{S}=\mathbf{s})}{|\mathcal{B}(\mathbf{s})|} \\
&\qquad\cdot\int_{\mathbf{y}\in\mathcal{T}_y\cap\mathcal{T}_0(\mathbf{s})\cap\mathcal{T}_1(\tilde{\mathbf{s}})} q(\mathbf{y}|\mathbf{s},\tilde{\mathbf{s}})\,\mathrm{d}\mathbf{y} \\
&= \frac{1}{2}e^{-n\frac{(\alpha+\beta)^2}{2\sigma^2}}\mathsf{P}(\mathbf{S}\in\mathcal{G}) \sum_{\mathbf{s}\in\mathcal{S}^n}\sum_{\tilde{\mathbf{s}}\in\mathcal{S}^n} \mathsf{Q}(\mathbf{S}=\mathbf{s},\tilde{\mathbf{S}}=\tilde{\mathbf{s}}) \\
&\quad\cdot\mathsf{Q}\big(\mathbf{Y}\in\mathcal{T}_y\cap\mathcal{T}_0(\mathbf{S})\cap\mathcal{T}_1(\tilde{\mathbf{S}})\,\big|\,\mathbf{S}=\mathbf{s},\tilde{\mathbf{S}}=\tilde{\mathbf{s}}\big) &(24)\\
&= \frac{1}{2}e^{-n\frac{(\alpha+\beta)^2}{2\sigma^2}}\mathsf{P}(\mathbf{S}\in\mathcal{G})\mathsf{Q}\big(\mathbf{Y}\in\mathcal{T}_y\cap\mathcal{T}_0(\mathbf{S})\cap\mathcal{T}_1(\tilde{\mathbf{S}})\big). &(25)
\end{aligned}$$

It remains to lower-bound (25). By the definition of $\mathcal{G}$ (8) and the law of large numbers, $\mathsf{P}(\mathbf{S}\in\mathcal{G})\geq 1-\varepsilon$ for sufficiently large n. And using

$$\mathsf{Q}(\mathbf{Y}\in\mathcal{A}\cap\mathcal{B}\cap\mathcal{C}) \geq 1-\mathsf{Q}(\mathbf{Y}\notin\mathcal{A})-\mathsf{Q}(\mathbf{Y}\notin\mathcal{B})-\mathsf{Q}(\mathbf{Y}\notin\mathcal{C}),$$

$$\begin{aligned}
&\mathsf{Q}(\mathbf{Y}\in\mathcal{T}_y\cap\mathcal{T}_0(\mathbf{S})\cap\mathcal{T}_1(\tilde{\mathbf{S}})) \\
&\quad\geq 1-\mathsf{Q}(\mathbf{Y}\notin\mathcal{T}_y)-\mathsf{Q}(\mathbf{Y}\notin\mathcal{T}_0(\mathbf{S}))-\mathsf{Q}(\mathbf{Y}\notin\mathcal{T}_1(\tilde{\mathbf{S}})). \quad (26)
\end{aligned}$$

Let $\mathcal{T}'_m(\mathbf{s}) \coloneqq \{\mathbf{Y_s}\in\mathbb{R}^{\mathsf{w}(\mathbf{s})} : \mathbf{Y}\in\mathcal{T}_m(\mathbf{s})\}$. Noting that for every $\mathbf{s}\in\mathcal{S}^n$, $\{\mathbf{Y_s}\notin\mathcal{T}'_m(\mathbf{s})\}$ is equivalent to $\{\mathbf{Y}\notin\mathcal{T}_m(\mathbf{s})\}$, and using the definitions of $q(\cdot)$ (13) and Q (22-23),

$$\begin{aligned}
\mathsf{Q}(\mathbf{Y}\notin\mathcal{T}_0(\mathbf{S})) &= \sum_{\mathbf{s}\in\mathcal{S}^n}\mathsf{Q}(\mathbf{S}=\mathbf{s})\underbrace{\mathsf{Q}(\mathbf{Y_s}\notin\mathcal{T}'_0(\mathbf{s})|\mathbf{S}=\mathbf{s})}_{=\mathsf{P}_0(\mathbf{Y_s}\notin\mathcal{T}'_0(\mathbf{s})|\mathbf{S}=\mathbf{s})} \\
&\qquad\cdot\underbrace{\mathsf{Q}(\mathbf{Y}\notin\mathcal{T}_0(\mathbf{s})|\mathbf{S}=\mathbf{s},\mathbf{Y_s}\notin\mathcal{T}'_0(\mathbf{s}))}_{=1} \\
&= \sum_{\mathbf{s}\in\mathcal{S}^n}\mathsf{Q}(\mathbf{S}=\mathbf{s})\mathsf{P}_0(\mathbf{Y_s}\notin\mathcal{T}'_0(\mathbf{s})|\mathbf{S}=\mathbf{s}) \\
&= \mathsf{P}_0(\mathbf{Y_S}\notin\mathcal{T}'_0(\mathbf{S})|\mathbf{S}\in\mathcal{G}) \\
&= \mathsf{P}_0(\|\mathbf{X}_{\mathbf{S}^c}(0,\mathbf{Y_S},\mathbf{S})\|^2\geq n\beta^2|\mathbf{S}\in\mathcal{G}),
\end{aligned}$$

$$\mathsf{Q}(\mathbf{Y}\notin\mathcal{T}_1(\tilde{\mathbf{S}})) = \mathsf{P}_1(\|\mathbf{X}_{\tilde{\mathbf{S}}^c}(1,\mathbf{Y}_{\tilde{\mathbf{S}}},\tilde{\mathbf{S}})\|^2\geq n\beta^2|\tilde{\mathbf{S}}\in\mathcal{G}),$$

and by Markov's inequality we lower bound (26) by

$$\begin{aligned}
&1-\frac{\mathsf{E}_{\mathsf{Q}}[\|\mathbf{Y}\|^2]}{n\alpha^2}-\frac{\mathsf{E}_{\mathsf{P}_0}\big[\|\mathbf{X}_{\mathbf{S}^c}(0,\mathbf{Y_S},\mathbf{S})\|^2\,\big|\,\mathbf{S}\in\mathcal{G}\big]}{n\beta^2} \\
&-\frac{\mathsf{E}_{\mathsf{P}_1}\big[\|\mathbf{X}_{\tilde{\mathbf{S}}^c}(1,\mathbf{Y}_{\tilde{\mathbf{S}}},\tilde{\mathbf{S}})\|^2\,\big|\,\tilde{\mathbf{S}}\in\mathcal{G}\big]}{n\beta^2}. \qquad (27)
\end{aligned}$$

Thus, we need to compute the expectations in (27) in order to make good choices for α and β. It is not hard to show that

$$\mathsf{E}_{\mathsf{Q}}[\|\mathbf{Y}\|^2] \leq \frac{2\rho nP+n(1+2\delta)\sigma^2}{(1-\varepsilon)}, \tag{28}$$

$$\mathsf{E}_{\mathsf{P}_0}\big[\|\mathbf{X}_{\mathbf{S}^c}(0,\mathbf{Y_S},\mathbf{S})\|^2\,\big|\,\mathbf{S}\in\mathcal{G}\big] \leq \frac{(1-\rho)nP}{(1-\varepsilon)}, \tag{29}$$

$$\mathsf{E}_{\mathsf{P}_1}\big[\|\mathbf{X}_{\tilde{\mathbf{S}}^c}(1,\mathbf{Y}_{\tilde{\mathbf{S}}},\tilde{\mathbf{S}})\|^2\,\big|\,\tilde{\mathbf{S}}\in\mathcal{G}\big] \leq \frac{(1-\rho)nP}{(1-\varepsilon)}. \tag{30}$$

Recall (5). The intuition for (28) is that under Q, about $2\rho n$ outputs are drawn according to $p_m(\cdot)$ with second moment $\approx 2\rho n(P+\sigma^2)$; the remaining are noise with energy $\approx (1-2\rho)n\sigma^2$. In (29-30), we calculate the second moment of a random selection of about $(1-\rho)n$ channel inputs, which is approximately $(1-\rho)nP$. Now, we (suboptimally) choose:

$$\alpha = \sqrt{\frac{2(2\rho P+(1+2\delta)\sigma^2)}{(1-\varepsilon)^2}}, \quad \beta = \sqrt{\frac{4(1-\rho)P}{(1-\varepsilon)^2}}. \tag{31}$$

Combining (27), (28)–(30) and (31), we arrive at

$$\mathsf{Q}(\mathbf{Y}\in\mathcal{T}_y\cap\mathcal{T}_0(\mathbf{S})\cap\mathcal{T}_1(\tilde{\mathbf{S}})) \geq \varepsilon. \tag{32}$$

Plugging this into (25) yields

$$\mathsf{P}(\text{error}) \geq \frac{1}{2}(1-\varepsilon)\varepsilon e^{-n\frac{4\left(\sqrt{(\rho P+(1+2\delta)\sigma^2/2}+\sqrt{(1-\rho)P}\right)^2}{2\sigma^2(1-\varepsilon)^2}}, \tag{33}$$

and computing the error exponent for $\varepsilon,\delta\to 0$ proves (6). ■

V. Achievability for $\rho > 1/2$

A. Two-Message Case

Theorem 5. *Let $\mathcal{M}=\{0,1\}$. For any $P/\sigma^2>0$ and any $\rho\in(1/2,1)$,*

$$\begin{aligned}
&\liminf_{n\to\infty}\frac{1}{n}\log\big(-\log p_{\mathrm{e}}(P/\sigma^2,n)\big) \\
&\quad\geq \max_{\alpha\in(0,1)} \alpha D\Big(1/2\,\Big\|\,\rho\Big(1-Q\Big(\sqrt{P/(\alpha\sigma^2)}\Big)\Big)\Big), \qquad (34)
\end{aligned}$$

where the maximization is subject to $Q\big(\sqrt{P/(\alpha\sigma^2)}\big)<\frac{\rho-1/2}{\rho}$ and we denote $D(a\|b)\coloneqq a\log\frac{a}{b}+(1-a)\log\frac{1-a}{1-b}$.

Note that the second order error exponent in (34) is upper-bounded by $-\log(1-\rho)$ as shown in [6].

To prove Theorem 5, we present a coding scheme achieving (34). As it will be similar for both messages with the obvious reversal of signs and inequalities, we describe the code construction for $M=0$. Let us fix the blocklength at n and let $\alpha\in(0,1)$ be such that $\alpha n\in\mathbb{N}$. The scheme is divided into three phases of length αn, $(1-\alpha)n-1$ and 1.

1) Sketch of the Coding Scheme: In the first phase, the transmitter uses a binary repetition code and performs hard-decisions on each of the feedback symbols. If the number of correct feedback symbols in phase one exceeds $\alpha n/2$, a majority decision of the hard-decision symbols cannot fail any more. Thus, the transmitter can be certain that symbol-wise hard-decisions followed by minimum distance decoding of the first αn symbols will succeed. The second phase is dedicated to saving enough power to ensure that the symbol error probability in phase one is sufficiently small. In the last channel use, the transmitter is then able to retransmit the message using large power if it did not observe more than $\alpha n/2$ correct symbols in phase one (i.e., there are either too many symbol errors or too few feedback symbols). We can now describe the scheme explicitly.

2) Encoder: During the first phase, the transmitter sends $X_1 = \cdots = X_{\alpha n} = A$; A is chosen below. In the second phase, it is silent to save power: $X_{\alpha n+1} = \cdots = X_{n-1} = 0$. For the third phase, let K denote the event that more than $\alpha n/2$ positive output symbols are fed back in phase one:

$$K = \left\{ \sum_{k=1}^{\alpha n} S_k \, 1\{Y_k > 0\} > \alpha n/2 \right\}. \tag{35}$$

The transmitter sends at time-n:

$$X_n = \begin{cases} 0 & \text{if } K \text{ occurs,} \\ \tilde{A} & \text{if } K^{\mathrm{c}} \text{ occurs.} \end{cases} \tag{36}$$

Next, we describe the choice of A and $\tilde{A}$. Choose $\delta, \varepsilon > 0$ and let $P' = P - \varepsilon$. We take $0 < \alpha < 1$ small enough to satisfy

$$\begin{aligned} & 1/2 < \rho\Big(1 - Q\Big(\sqrt{P'/(\alpha\sigma^2)}\Big)\Big) \\ \iff & Q\Big(\sqrt{P'/(\alpha\sigma^2)}\Big) < (\rho - 1/2)/\rho, \end{aligned} \tag{37}$$

and choose

$$A = \sqrt{P'/\alpha}, \quad \tilde{A} = e^{\frac{\alpha n}{2}\left(D\left(\frac{1}{2}\middle\|\rho\left(1-Q\left(\sqrt{\frac{P'}{\alpha\sigma^2}}\right)\right)\right)-\delta\right)}. \tag{38}$$

The choice of amplitudes will be justified later by showing that the power constraint is satisfied.

3) Decoder: The decoder makes a tentative decision after phase one based on the sign of the majority of the symbols. If the magnitude of the last symbol is less than $\tilde{A}/2$, it takes its tentative as the final decision, otherwise it decides based on the sign of Y_n. Accordingly, we define the decoding function:

$$\varphi(\mathbf{Y}) = \begin{cases} 0 & \text{if } \left(\sum_{k=1}^{\alpha n} 1\{Y_k > 0\} > \frac{\alpha n}{2}, Y_n > -\frac{\tilde{A}}{2}\right) \\ & \text{or } \left(\sum_{k=1}^{\alpha n} 1\{Y_k > 0\} \le \frac{\alpha n}{2}, Y_n \ge \frac{\tilde{A}}{2}\right), \\ 1 & \text{otherwise.} \end{cases} \tag{39}$$

4) Error Probability: We expand the probability of error:

$$\begin{aligned} \mathsf{P}_0(\text{error}) &= \mathsf{P}_0(K)\mathsf{P}_0(\text{error}|K) + \mathsf{P}_0(K^{\mathrm{c}})\mathsf{P}_0(\text{error}|K^{\mathrm{c}}) \\ &\le \mathsf{P}_0(\text{error}|K) + \mathsf{P}_0(\text{error}|K^{\mathrm{c}}). \end{aligned} \tag{40}$$

Starting with the event K, we have

$$\mathsf{P}_0(\text{error}|K) = \mathsf{P}_0(Y_n \le -\tilde{A}/2 | X_n = 0) = Q(\tilde{A}/2). \tag{41}$$

Conditioned on K^{c}, the decoder errs if the retransmission fails:

$$\mathsf{P}_0(\text{error}|K^{\mathrm{c}}) = \mathsf{P}_0(Y_n < \tilde{A}/2 | X_n = \tilde{A}) = Q(\tilde{A}/2). \tag{42}$$

Inserting (41) and (42) into (40) yields

$$\mathsf{P}_0(\text{error}) \le 2Q(\tilde{A}/2), \tag{43}$$

from which we conclude using (38) and the symmetry of the coding scheme for the two messages:

$$\begin{aligned} & \liminf_{n\to\infty} \frac{1}{n} \log\left(-\log p_{\mathrm{e}}\left(P/\sigma^2, n\right)\right) \\ & \quad \ge \alpha n \left(D\left(\frac{1}{2}\middle\|\rho\left(1 - Q\left(\sqrt{\frac{P-\varepsilon}{\alpha\sigma^2}}\right)\right)\right) - \delta \right). \end{aligned} \tag{44}$$

Letting δ, ε tend to zero as n tends to infinity and maximizing over α then yields (34).

5) Power Consumption: In the first phase, we have $\mathsf{E}\left[\sum_{k=1}^{\alpha n} X_k^2\right] = nP'$. The second phase and the third phase in the event K use no power. To get an expression for $\mathsf{P}_0(K^{\mathrm{c}})$, we observe that at each channel use in phase one, the probability that a symbol turns out to be correct and is fed back to the receiver is for $1 \le k \le \alpha n$:

$$\begin{aligned} \mathsf{P}_0(S_k = 1, Y_k > 0) &= \mathsf{P}_0(S_k = 1)\,\mathsf{P}_0(Y_k > 0) \\ &= \rho\Big(1 - Q\Big(\sqrt{P'/(\alpha\sigma^2)}\Big)\Big), \end{aligned} \tag{45}$$

since S_k is independent of Y_k at any time. Thus, $\mathsf{P}_0(K^{\mathrm{c}})$ is the probability that the sum of αn IID Bernoulli random variables with success probability $\rho\big(1-Q\big(\sqrt{P'/(\alpha\sigma^2)}\big)\big)$ does not exceed $\alpha n/2$. Hence, given that (37) holds, we can apply Sanov's theorem to find that

$$\mathsf{P}_0(K^{\mathrm{c}}) \le e^{-\alpha n\left(D\left(\frac{1}{2}\middle\|\rho\left(1-Q\left(\sqrt{\frac{P'}{\alpha\sigma^2}}\right)\right)\right)-\delta_n\right)}, \tag{46}$$

where $\delta_n \to 0$ as $n \to \infty$. It follows from (38) and (46) that

$$\lim_{n\to\infty} \mathsf{E}_{\mathsf{P}_0}[X_n^2 \,|\, K^{\mathrm{c}}]\, \mathsf{P}_0(K^{\mathrm{c}}) = 0. \tag{47}$$

The total power usage is thus

$$\mathsf{E}_{\mathsf{P}_0}\left[\sum_{k=1}^{n} X_k^2\right] \le nP' + \varepsilon_n = n(P - \varepsilon) + \varepsilon_n, \tag{48}$$

where $\varepsilon_n \to 0$ as $n \to \infty$. Therefore, the power constraint is indeed satisfied. ■

B. Positive Rates

Let $\mathcal{M} = \{1, \ldots, e^{nR}\}$ and let $p_{\mathrm{e}}(P/\sigma^2, R, n)$ denote the least probability of error of any blocklength-n coding scheme satisfying (5).

Theorem 6. *For any P/σ^2, there exists an $R_0 > 0$ such that for any $R \in (0, R_0)$, we can find a blocklength-n coding scheme transmitting at rate R with*

$$\liminf_{n\to\infty} \frac{1}{n} \log\big(-\log p_{\mathrm{e}}\big(P/\sigma^2, R, n\big)\big) > 0. \tag{49}$$

Sketch of the Proof: We sketch the construction of a sequence of blocklength-n coding schemes transmitting at a positive rate and achieving an error probability decaying double-exponentially in n; by similarity we restrict ourselves to $M = m$. Let $\alpha > 0$ so that $\alpha n \in \mathbb{N}$. We split the block into three phases of length $n' = \alpha n$, 1 and $n'' = (1 - \alpha)n - 1$.

First Phase: Let $R' > 0$. We use a q-ary block code $\mathcal{V}$ of size $e^{n'R'}$. By the Gilbert-Varshamov bound [8], we can find such a code with minimum Hamming distance $\delta n'$ if $R' \leq (1 - H_q(\delta)) \log q$, where $H_q(\delta) : [0,1] \to [0,1]$ is the q-ary entropy function which is concave and attains its maximum of 1 at $\delta = 1 - 1/q$. Hence, any $\delta \in (0, 1 - 1/q)$ allows $R' > 0$ for sufficiently large n'. To transmit a codeword over the channel, we use a mapping $\phi : \mathcal{V} \to \mathcal{X}^{n'}$, where $\mathcal{X}$ is a q-ary pulse amplitude modulation (PAM) signal constellation with a Euclidean distance of $d(q, \alpha)$ between two signal points; and $d(q, \alpha)$ is decreasing in q and α. The mapping ϕ generates the set of $e^{n'R'}$ channel input sequences $\mathcal{C}$ and is chosen to maximize $d(q, \alpha)$ while satisfying for some $\varepsilon > 0$:

$$\sum_{k=1}^{n'} x_k^2 \leq P - \varepsilon, \quad x^{n'} \in \mathcal{C}. \tag{50}$$

Second Phase: The transmitter performs optimal hard-decisions on all feedback symbols and counts the number of correct symbols. If more than $n'(1 - \delta/2)$ symbols are fed back and correct, it is clear that decoding via hard-decisions for the PAM symbols followed by inverting ϕ and applying minimum Hamming distance decoding cannot fail as the amount of possible symbol errors is less than half the minimum Hamming distance of $\mathcal{V}$. We denote this event by K. If K occurs, the transmitter remains silent. Otherwise, it sends a flag:

$$X_{n'+1} = \sqrt{\varepsilon / (2\,\mathsf{P}_m(K^{\mathrm{c}}))}. \tag{51}$$

Third Phase: If the transmitter was silent in phase two, it also remains silent in phase three. Otherwise, it retransmits the intended message in the remaining n'' time slots using a code $\tilde{\mathcal{C}}$ of size $e^{n'R'}$ that satisfies

$$\sum_{k=1}^{n''} x_k^2 \leq \frac{\varepsilon}{2\,\mathsf{P}_m(K^{\mathrm{c}})}, \quad x^{n''} \in \tilde{\mathcal{C}}. \tag{52}$$

Note that $\tilde{\mathcal{C}}$ is not a q-ary block code but any good non-feedback code for a Gaussian channel satisfying (52).

Decoder: The decoder first compares $Y_{n'+1}$ to the threshold

$$\Upsilon = 1/2 \cdot \sqrt{\varepsilon / (2\,\mathsf{P}_m(K^{\mathrm{c}}))}. \tag{53}$$

If $Y_{n'+1} > \Upsilon$, the receiver decodes the message based on the last n'' channel outputs using an optimal decoder for $\tilde{\mathcal{C}}$. Otherwise, it decodes based on the first n' symbols using symbol-wise hard-decisions, inversion of ϕ and a minimum Hamming distance decoder for $\mathcal{V}$.

Error Probability: Trivially,

$$\mathsf{P}_m(\text{error}) \leq \mathsf{P}_m(\text{error}|K) + \mathsf{P}_m(\text{error}|K^{\mathrm{c}}). \tag{54}$$

We first bound the latter probability. Let A_k, $1 \leq k \leq n'$, be the outcome of an optimal hard-decision for a symbol Y_k in phase one. We bound $\mathsf{P}_m(A_k = X_k)$ neglecting the lower symbol error probability of the two outer symbols:

$$\mathsf{P}_m(A_k = X_k) \geq 1 - 2Q\big(d(q, \alpha)/2\sigma\big). \tag{55}$$

Using the independence of S_k and Y_k,

$$\mathsf{P}_m(S_k = 1, A_k = X_k) = \rho\,\mathsf{P}_m(A_k = X_k). \tag{56}$$

Thus, K^{c} is the event that the outcome of n' independent Bernoulli random variables of success probability $\rho\,\mathsf{P}_m(A_k = X_k)$ yields at most $(1 - \delta/2)n'$ successes. Taking q, δ large and α small enough to satisfy

$$\rho\,\mathsf{P}_m(A_k = X_k) > (1 - \delta/2), \tag{57}$$

we can apply Sanov's theorem to find for some $\gamma > 0$ that

$$\mathsf{P}_m(K^{\mathrm{c}}) \leq e^{-n'\gamma}. \tag{58}$$

If K occurs, the decoder errs if it incorrectly thinks that a flag was sent. The probability is by (53) and (58):

$$\mathsf{P}_m(Y_{n'+1} > \Upsilon | K) = Q(\Upsilon/\sigma) \leq Q\Big(\sqrt{\varepsilon/(8\sigma^2)} \cdot e^{n'\gamma/2}\Big), \tag{59}$$

which decays double-exponentially in n. If K^{c} occurs, the decoder errs if either the receiver incorrectly thinks that no flag was sent or if decoding based on the last n'' symbols fails:

$$\mathsf{P}_m(\text{error}|K^{\mathrm{c}}) \leq \mathsf{P}_m(Y_{n'+1} \leq \Upsilon | K^{\mathrm{c}}) + p_{\mathrm{e}}(\tilde{\mathcal{C}}), \tag{60}$$

where $p_{\mathrm{e}}(\tilde{C})$ is the maximal error probability of the code $\tilde{\mathcal{C}}$. The first probability in (60) decays double-exponentially in n for the same reason as in (59). The second probability also decays double-exponentially in n on account of [9, Eq. (77)] since the power constraint for $\tilde{\mathcal{C}}$ (52) grows exponentially in n by (58) provided that (57) is satisfied. Hence, the total probability of error decays double-exponentially in n. It is easy to show that the power constraint (5) is satisfied.

Rate: The condition (57) can be satisfied for any $\rho > 1/2$ by choosing δ and $\mathsf{P}_m(A_k = X_k)$ sufficiently close to one (i.e., q large enough and α small enough). The overall transmission rate is then $R = \alpha R'$ which is positive since R' and α are positive. This concludes the proof. ■

Acknowledgment

The author would like to express his gratitude to Christoph Bunte and Amos Lapidoth for helpful discussions.

References

[1] R. Mirghaderi, A. J. Goldsmith, T. Weissman, "Achievable error exponents in the Gaussian channel with rate-limited feedback," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8144–8156, Dec. 2013

[2] Y.-H. Kim, A. Lapidoth, T. Weissman, "On the Reliability of Gaussian Channels with Noisy Feedback," Forty-Fourth Annual Allerton Conf., Illinois, USA, Sep. 2006

[3] Y.-H. Kim, A. Lapidoth, T. Weissman, "Error Exponents for the Gaussian channel with active noisy feedback," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1223–1236, Mar. 2011

[4] Y.-H. Kim, A. Lapidoth, and T. Weissman, "Bounds on the error exponent of the AWGN channel with AWGN-corrupted feedback," Proc. 24th IEEE Conv. of Elec. and Elect. Eng. in Israel, pp. 184-188, Israel, Nov. 2006.

[5] M. Agarwal, D. Guo, M. L. Honig, "Error exponent for Gaussian channels with partial sequential feedback," *IEEE Trans. Inf. Theory*, vol. 59, no. 8, pp. 4757–4766, Aug. 2013

[6] C. Bunte, A. Lapidoth and L. Palzer, "Coding for the Gaussian channel with intermittent feedback," *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, USA, Jun. 2014.

[7] A. Khisti and A. Lapidoth, "Multiple access channels with intermittent feedback and side information," *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013

[8] R. M. Roth, *Introduction to coding theory*, Camebridge Univ. Press, 2006.

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

Appendix

Proof of Lemma 3

Recall that $\mathbf{s}^{\mathrm{c}} = \mathbf{1} - \mathbf{s}$. Then, for any $\mathbf{s} \in \mathcal{G}$, $\tilde{\mathbf{s}} \in \mathcal{B}(\mathbf{s})$:

$$\begin{aligned}
&p_0(\mathbf{y}|\mathbf{s}) \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\|\mathbf{y}-\mathbf{x}(0,\mathbf{y}_{\mathbf{s}},\mathbf{s})\|^2}{2\sigma^2}} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\|\mathbf{y}_{\mathbf{s}}-\mathbf{x}_{\mathbf{s}}(0,\mathbf{y}_{\mathbf{s}},\mathbf{s})\|^2}{2\sigma^2}} e^{-\frac{\|\mathbf{y}_{\mathbf{s}^{\mathrm{c}}}-\mathbf{x}_{\mathbf{s}^{\mathrm{c}}}(0,\mathbf{y}_{\mathbf{s}},\mathbf{s})\|^2}{2\sigma^2}} \\
&\geq (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\|\mathbf{y}_{\mathbf{s}}-\mathbf{x}_{\mathbf{s}}(0,\mathbf{y}_{\mathbf{s}},\mathbf{s})\|^2}{2\sigma^2}} e^{-\frac{(\|\mathbf{y}_{\mathbf{s}^{\mathrm{c}}}\|+\|\mathbf{x}_{\mathbf{s}^{\mathrm{c}}}(0,\mathbf{y}_{\mathbf{s}},\mathbf{s})\|)^2}{2\sigma^2}} && (61) \\
&\geq (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\|\mathbf{y}_{\mathbf{s}}-\mathbf{x}_{\mathbf{s}}(0,\mathbf{y}_{\mathbf{s}},\mathbf{s})\|^2}{2\sigma^2}} e^{-\frac{n(\alpha+\beta)^2}{2\sigma^2}} && (62) \\
&\geq (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\|\mathbf{y}_{\mathbf{s}}-\mathbf{x}_{\mathbf{s}}(0,\mathbf{y}_{\mathbf{s}},\mathbf{s})\|^2}{2\sigma^2}} e^{-\frac{\|\mathbf{y}_{\tilde{\mathbf{s}}}-\mathbf{x}_{\tilde{\mathbf{s}}}(1,\mathbf{y}_{\tilde{\mathbf{s}}},\tilde{\mathbf{s}})\|^2}{2\sigma^2}} \\
&\qquad \cdot e^{-\frac{\|\mathbf{y}_{\mathbf{r}}\|^2}{2\sigma^2}} e^{-n\frac{(\alpha+\beta)^2}{2\sigma^2}} && (63) \\
&= p_0(\mathbf{y}_{\mathbf{s}}|\mathbf{s})\, p_1(\mathbf{y}_{\tilde{\mathbf{s}}}|\tilde{\mathbf{s}})\, g(\mathbf{y}_{\mathbf{r}}|\mathbf{r})\, e^{-n\frac{(\alpha+\beta)^2}{2\sigma^2}} \\
&= q(\mathbf{y}|\mathbf{s},\tilde{\mathbf{s}})\, e^{-n\frac{(\alpha+\beta)^2}{2\sigma^2}}, \qquad \mathbf{y} \in \mathcal{T}_y \cap \mathcal{T}_0(\mathbf{s})\,. && (64)
\end{aligned}$$

Here,

(61) follows from the triangle inequality,

(62) follows from $\|\mathbf{y}_{\mathbf{s}^{\mathrm{c}}}\| \leq \|\mathbf{y}\| \leq \sqrt{n}\alpha$ and $\|\mathbf{x}_{\mathbf{s}^{\mathrm{c}}}(0,\mathbf{y}_{\mathbf{s}},\mathbf{s})\| \leq \sqrt{n}\beta$ for $\mathbf{y} \in \mathcal{T}_y \cap \mathcal{T}_0(\mathbf{s})$,

(63) follows since $1 \geq e^{-\frac{\|\mathbf{y}_{\tilde{\mathbf{s}}}-\mathbf{x}_{\tilde{\mathbf{s}}}(1,\mathbf{y}_{\tilde{\mathbf{s}}},\tilde{\mathbf{s}})\|^2}{2\sigma^2}} e^{-\frac{\|\mathbf{y}_{\mathbf{r}}\|^2}{2\sigma^2}}$,

(64) follows from the definition of $q(\cdot)$ (13). ■

Computing the Expectations (28-30)

We prove the upper bounds (28-30). Recall that $\mathbf{r} = \mathbf{1} - \mathbf{s} - \tilde{\mathbf{s}}$. For (28),

$$\begin{aligned}
&\mathsf{E}_{\mathsf{Q}}[\|\mathbf{Y}\|^2] \\
&= \sum_{\mathbf{s},\tilde{\mathbf{s}}\in\mathcal{S}^n} \mathsf{Q}(\mathbf{S}=\mathbf{s},\tilde{\mathbf{S}}=\tilde{\mathbf{s}}) \\
&\qquad \cdot\ \mathsf{E}_{\mathsf{Q}}\big[\|\mathbf{Y}_{\mathbf{s}}\|^2 + \|\mathbf{Y}_{\tilde{\mathbf{s}}}\|^2 + \|\mathbf{Y}_{\mathbf{r}}\|^2 \,\big|\, \mathbf{S}=\mathbf{s},\tilde{\mathbf{S}}=\tilde{\mathbf{s}}\big] \\
&= \mathsf{E}_{\mathsf{Q}}[\|\mathbf{Y}_{\mathbf{S}}\|^2] + \mathsf{E}_{\mathsf{Q}}[\|\mathbf{Y}_{\tilde{\mathbf{S}}}\|^2] + \mathsf{E}_{\mathsf{Q}}[\|\mathbf{Y}_{\mathbf{R}}\|^2]. && (65)
\end{aligned}$$

Using the definition of Q (22-23), we can write

$$\begin{aligned}
\mathsf{E}_{\mathsf{Q}}\big[\|\mathbf{Y}_{\mathbf{S}}\|^2\big] &= \mathsf{E}_{\mathsf{P}_0}\big[\|\mathbf{Y}_{\mathbf{S}}\|^2 \,\big|\, \mathbf{S}\in\mathcal{G}\big] && (66) \\
&\leq \frac{\mathsf{E}_{\mathsf{P}_0}\big[\|\mathbf{Y}_{\mathbf{S}}\|^2\big]}{\mathsf{P}(\mathbf{S}\in\mathcal{G})} && (67) \\
&= \frac{\mathsf{E}_{\mathsf{P}_0}[\sum_{k=1}^n 1\{S_k=1\}Y_k^2]}{\mathsf{P}(\mathbf{S}\in\mathcal{G})} \\
&= \frac{\sum_{k=1}^n \mathsf{E}_{\mathsf{P}}[1\{S_k=1\}]\mathsf{E}_{\mathsf{P}_0}[Y_k^2]}{\mathsf{P}(\mathbf{S}\in\mathcal{G})} && (68) \\
&= \frac{\rho\,\mathsf{E}_{\mathsf{P}_0}[\|\mathbf{Y}\|^2]}{\mathsf{P}(\mathbf{S}\in\mathcal{G})} \\
&\leq \frac{\rho n(P+\sigma^2)}{\mathsf{P}(\mathbf{S}\in\mathcal{G})} && (69) \\
&\leq \frac{\rho n(P+\sigma^2)}{1-\varepsilon}, && (70)
\end{aligned}$$

where

(66) follows from the definition of Q,

(67) follows from the nonnegativity of $\|\mathbf{Y}_{\mathbf{s}}\|^2$,

(68) follows from the independence of Y_k and S_k,

(69) follows from the power constraint on $\mathbf{X}$ (5), the distribution of $\mathbf{Z}$ and the independence of X_k and Z_k,

(70) follows from the law of large numbers.

Applying a similar analysis to the second term in (65),

$$\mathsf{E}_{\mathsf{Q}}[\|\mathbf{Y}_{\tilde{\mathbf{S}}}\|^2] \leq \frac{\rho n(P+\sigma^2)}{1-\varepsilon}. \tag{71}$$

For the third term, we have

$$\begin{aligned}
&\mathsf{E}_{\mathsf{Q}}\big[\|\mathbf{Y}_{\mathbf{R}}\|^2\big] \\
&= \sum_{\mathbf{s},\tilde{\mathbf{s}}\in\mathcal{S}^n} \mathsf{Q}(\mathbf{S}=\mathbf{s},\tilde{\mathbf{S}}=\tilde{\mathbf{s}})\mathsf{E}_{\mathsf{Q}}\big[\|\mathbf{Y}_{\mathbf{r}}\|^2 \,\big|\, \mathbf{S}=\mathbf{s},\tilde{\mathbf{S}}=\tilde{\mathbf{s}}\big] \\
&= \sum_{\mathbf{s},\tilde{\mathbf{s}}\in\mathcal{S}^n} \mathsf{Q}(\mathbf{S}=\mathbf{s},\tilde{\mathbf{S}}=\tilde{\mathbf{s}})(n-\mathsf{w}(\mathbf{s})-\mathsf{w}(\tilde{\mathbf{s}}))\sigma^2 && (72) \\
&= \sigma^2(n - 2\,\mathsf{E}_{\mathsf{Q}}[\mathsf{w}(\mathbf{S})]) && (73) \\
&\leq \sigma^2(n - 2n(\rho-\delta)), && (74)
\end{aligned}$$

where

(72) holds since $\mathbf{Y}_{\mathbf{r}}$ has $(n-\mathsf{w}(\mathbf{s})-\mathsf{w}(\tilde{\mathbf{s}}))$ elements each of which is drawn independently from a Gaussian distribution with zero mean and variance σ^2 (13),

(73) is true since $\mathbf{S}$ and $\tilde{\mathbf{S}}$ have the same marginals under Q,

(74) follows from the definition of Q (22) and $\mathcal{G}$ (8).

Combining (65), (70), (71) and (74) proves (28). ■

For (29),

$$\begin{aligned}
&\mathsf{E}_{\mathsf{P}_0}\big[\|\mathbf{X}_{\mathbf{S}^{\mathrm{c}}}(0,\mathbf{Y}_{\mathbf{S}},\mathbf{S})\|^2 \,\big|\, \mathbf{S}\in\mathcal{G}\big] \\
&\leq \frac{\mathsf{E}_{\mathsf{P}_0}\big[\sum_{k=1}^n 1\{S_k=0\}X_k(0,\mathbf{Y}_{S^{k-1}},S^{k-1})^2\big]}{\mathsf{P}(\mathbf{S}\in\mathcal{G})} && (75) \\
&\leq \frac{\sum_{k=1}^n \mathsf{E}_{\mathsf{P}_0}\big[1\{S_k=0\}\big]\,\mathsf{E}_{\mathsf{P}_0}\big[X_k(0,\mathbf{Y}_{S^{k-1}},S^{k-1})^2\big]}{\mathsf{P}(\mathbf{S}\in\mathcal{G})} \\
& && (76) \\
&= \frac{\rho\,\mathsf{E}_{\mathsf{P}_0}\big[\|\mathbf{X}(0,\mathbf{Y}_{\mathbf{S}},\mathbf{S})\|^2\big]}{\mathsf{P}(\mathbf{S}\in\mathcal{G})} \\
&\leq \frac{(1-\rho)nP}{1-\varepsilon}, && (77)
\end{aligned}$$

where

(75) follows from the nonnegativity of $\|\mathbf{X}_{\mathbf{S}^{\mathrm{c}}}(0,\mathbf{Y}_{\mathbf{S}},\mathbf{S})\|^2$,

(76) follows from the independence of S_k and $X_k(\cdot)$,

(77) follows from the power constraint (5) and the law of large numbers. ■

A similar calculation proves (30). ■

Proof of Lemma 4

It remains to prove Lemma 4. In the following derivation,

(78) uses the fact that $(\rho-\delta)n \leq \mathsf{w}(\mathbf{s}) \leq (\rho+\delta)n$ for all $\mathbf{s} \in \mathcal{G}$ by (8) and reorders the summation with (11),

(79) is true on account of (19),

(80) holds by (3) since $\mathcal{D}_0 \cup \mathcal{D}_1 = \mathbb{R}^n$. ■

$$
\begin{align}
&\sum_{\mathbf{s}\in\mathcal{G}}\sum_{\tilde{\mathbf{s}}\in\mathcal{B}(\mathbf{s})}\frac{\mathsf{P}(\mathbf{S}=\mathbf{s})}{|\mathcal{B}(\mathbf{s})|}\int_{\mathbf{y}\in\mathcal{D}_1\cap\mathcal{T}_y\cap\mathcal{T}_0(\mathbf{s})\cap\mathcal{T}_1(\tilde{\mathbf{s}})} q(\mathbf{y}|\mathbf{s},\tilde{\mathbf{s}})\,\mathrm{d}\mathbf{y}+\sum_{\tilde{\mathbf{s}}\in\mathcal{G}}\sum_{\mathbf{s}\in\mathcal{B}(\tilde{\mathbf{s}})}\frac{\mathsf{P}(\tilde{\mathbf{S}}=\tilde{\mathbf{s}})}{|\mathcal{B}(\tilde{\mathbf{s}})|}\int_{\mathbf{y}\in\mathcal{D}_0\cap\mathcal{T}_y\cap\mathcal{T}_0(\mathbf{s})\cap\mathcal{T}_1(\tilde{\mathbf{s}})} q(\mathbf{y}|\mathbf{s},\tilde{\mathbf{s}})\,\mathrm{d}\mathbf{y} \notag\\
&=\sum_{k=\lceil(\rho-\delta)n\rceil}^{\lfloor(\rho+\delta)n\rfloor}\sum_{(\mathbf{s},\tilde{\mathbf{s}})\in\mathcal{G}_k}\frac{\mathsf{P}(\mathbf{S}=\mathbf{s})}{|\mathcal{B}(\mathbf{s})|}\int_{\mathbf{y}\in\mathcal{D}_1\cap\mathcal{T}_y\cap\mathcal{T}_0(\mathbf{s})\cap\mathcal{T}_1(\tilde{\mathbf{s}})} q(\mathbf{y}|\mathbf{s},\tilde{\mathbf{s}})\,\mathrm{d}\mathbf{y} \notag\\
&\quad+\sum_{k=\lceil(\rho-\delta)n\rceil}^{\lfloor(\rho+\delta)n\rfloor}\sum_{(\mathbf{s},\tilde{\mathbf{s}})\in\mathcal{G}_k}\frac{\mathsf{P}(\tilde{\mathbf{S}}=\tilde{\mathbf{s}})}{|\mathcal{B}(\tilde{\mathbf{s}})|}\int_{\mathbf{y}\in\mathcal{D}_0\cap\mathcal{T}_y\cap\mathcal{T}_0(\mathbf{s})\cap\mathcal{T}_1(\tilde{\mathbf{s}})} q(\mathbf{y}|\mathbf{s},\tilde{\mathbf{s}})\,\mathrm{d}\mathbf{y} \tag{78}\\
&=\sum_{k=\lceil(\rho-\delta)n\rceil}^{\lfloor(\rho+\delta)n\rfloor}\sum_{(\mathbf{s},\tilde{\mathbf{s}})\in\mathcal{G}_k}\left(\frac{\mathsf{P}(\mathbf{S}=\mathbf{s})}{|\mathcal{B}(\mathbf{s})|}\int_{\mathbf{y}\in\mathcal{D}_1\cap\mathcal{T}_y\cap\mathcal{T}_0(\mathbf{s})\cap\mathcal{T}_1(\tilde{\mathbf{s}})} q(\mathbf{y}|\mathbf{s},\tilde{\mathbf{s}})\,\mathrm{d}\mathbf{y}+\frac{\mathsf{P}(\tilde{\mathbf{S}}=\tilde{\mathbf{s}})}{|\mathcal{B}(\tilde{\mathbf{s}})|}\int_{\mathbf{y}\in\mathcal{D}_0\cap\mathcal{T}_y\cap\mathcal{T}_0(\mathbf{s})\cap\mathcal{T}_1(\tilde{\mathbf{s}})} q(\mathbf{y}|\mathbf{s},\tilde{\mathbf{s}})\,\mathrm{d}\mathbf{y}\right) \notag\\
&=\sum_{k=\lceil(\rho-\delta)n\rceil}^{\lfloor(\rho+\delta)n\rfloor}\sum_{(\mathbf{s},\tilde{\mathbf{s}})\in\mathcal{G}_k}\frac{\mathsf{P}(\mathbf{S}=\mathbf{s})}{|\mathcal{B}(\mathbf{s})|}\left(\int_{\mathbf{y}\in\mathcal{D}_1\cap\mathcal{T}_y\cap\mathcal{T}_0(\mathbf{s})\cap\mathcal{T}_1(\tilde{\mathbf{s}})} q(\mathbf{y}|\mathbf{s},\tilde{\mathbf{s}})\,\mathrm{d}\mathbf{y}+\int_{\mathbf{y}\in\mathcal{D}_0\cap\mathcal{T}_y\cap\mathcal{T}_0(\mathbf{s})\cap\mathcal{T}_1(\tilde{\mathbf{s}})} q(\mathbf{y}|\mathbf{s},\tilde{\mathbf{s}})\,\mathrm{d}\mathbf{y}\right) \tag{79}\\
&=\sum_{k=\lceil(\rho-\delta)n\rceil}^{\lfloor(\rho+\delta)n\rfloor}\sum_{(\mathbf{s},\tilde{\mathbf{s}})\in\mathcal{G}_k}\frac{\mathsf{P}(\mathbf{S}=\mathbf{s})}{|\mathcal{B}(\mathbf{s})|}\int_{\mathbf{y}\in\mathcal{T}_y\cap\mathcal{T}_0(\mathbf{s})\cap\mathcal{T}_1(\tilde{\mathbf{s}})} q(\mathbf{y}|\mathbf{s},\tilde{\mathbf{s}})\,\mathrm{d}\mathbf{y} \tag{80}\\
&=\sum_{\mathbf{s}\in\mathcal{G}}\sum_{\tilde{\mathbf{s}}\in\mathcal{B}(\mathbf{s})}\frac{\mathsf{P}(\mathbf{S}=\mathbf{s})}{|\mathcal{B}(\mathbf{s})|}\int_{\mathbf{y}\in\mathcal{T}_y\cap\mathcal{T}_0(\mathbf{s})\cap\mathcal{T}_1(\tilde{\mathbf{s}})} q(\mathbf{y}|\mathbf{s},\tilde{\mathbf{s}})\,\mathrm{d}\mathbf{y}. \notag
\end{align}
$$