Vazquez-Vilar, G. (2019). *On the Error Probability of Optimal Codes in Gaussian Channels under Maximal Power Constraint*. In: 2019 IEEE International Symposium on Information Theory (ISIT), 7-12 July 2019, pp. 2943-2946.

DOI: 10.1109/isit.2019.8849543

On the Error Probability of Optimal Codes in Gaussian Channels under Maximal Power Constraint

Gonzalo Vazquez-Vilar
Universidad Carlos III de Madrid, Spain
Email: gvazquez@ieee.org

***Abstract*—For an additive white Gaussian noise channel, we prove that Th. 41 in [Polyanskiy, Poor, Verdú 2010] is a lower bound to the error probability of any channel code satisfying the maximal power constraint. In contrast, the (tighter) lower bound to the error probability in Eq. (20) in [Shannon 1959] only holds under equal power constraint.**

I. Introduction

We consider the problem of transmitting M equiprobable messages over n uses of an additive white Gaussian noise (AWGN) channel. In [1], Shannon derived a lower bound on the error probability for codes subject to a certain power constraint Γ. Using geometrical arguments, Shannon lower-bounded the error probability of a code with all the codewords lying on the n-dimensional sphere with squared radius $n\Gamma$ (*equal power constraint*) [1, Eq. (20)]. Then, he considered a length-n code such that the codeword energy is not larger than $n\Gamma$ (*maximal power constraint*). He argued that such code can be transformed by adding an extra $(n+1)$-th coordinate to equalize the codeword energy to $n\Gamma$. As a result, the lower bound in [1, Eq. (20)], evaluated for the blocklength $n+1$, also holds for any length-n maximal power constrained code.

More recently, Polyanskiy, Poor and Verdú proved that a surrogate binary hypothesis test can be used to lower bound the error probability of a channel code [2, Th. 27]. Particularizing this bound for the additive white Gaussian noise (AWGN) channel under equal power constraint yields [2, Th. 41]. As discussed above, evaluating [2, Th. 41] for a blocklength $n+1$ yields a converse bound for a length-n code in the maximal power constraint setting.

While most of the analysis in [1] is focused in characterizing the asymptotics of [1, Eq. (20)], this bound is extremely accurate in the finite-length setting [3]. Indeed, in general, Shannon's approach yields tighter bounds than [2, Th. 41] under equal power constraint. In this work, we prove that [2, Th. 41] is directly a lower bound to the error probability of a length-n maximal power constrained code (with no $n+1$ extension required). In contrast, Shannon lower bound only holds under equal power constraint, and the $n+1$ extension argument is needed in the maximal power constraint setting.

G. Vazquez-Vilar is also with the Gregorio Marañón Health Research Institute, Madrid, Spain. This work has been funded in part by the European Research Council (ERC) under grant 714161, and by the Spanish Ministry of Economy and Competitiveness under grants IJCI-2015-27020 and TEC2016-78434-C3 (AEI/FEDER, EU).

II. System Model and Preliminaries

We consider the problem of transmitting M equiprobable messages over n uses of an AWGN channel W with noise power σ^2. Specifically, for the input $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and output $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ the channel $W = P_{\boldsymbol{Y}|\boldsymbol{X}}$ has a probability density function (pdf) given by

$$w(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} \varphi_{x_i,\sigma}(y_i), \tag{1}$$

where $\varphi_{\mu,\sigma}(\cdot)$ denotes the pdf of the Gaussian distribution,

$$\varphi_{\mu,\sigma}(x) \triangleq \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{2}$$

The encoder maps a message $v \in \{1, \ldots, M\}$ to the channel as $\boldsymbol{x} = \boldsymbol{c}_v$ using the codebook $\mathcal{C} \triangleq \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_M\}$. Based on the channel output $\boldsymbol{y}$, the decoder guesses the transmitted message $\hat{v} \in \{1, \ldots, M\}$. The error probability is thus given by $P_{\mathrm{e}}(\mathcal{C}) \triangleq \Pr\{\hat{V} \neq V\}$ where the underlying probability is induced by the chain of source, encoder, channel and decoder. We consider codebooks satisfying a certain power constraint:

- Equal-power constrained codes,
$$\mathcal{L}_{\mathrm{e}}(\Gamma) \triangleq \Big\{ \mathcal{C} \;\Big|\; \|\boldsymbol{c}_i\|^2 = n\Gamma, \quad i = 1, \ldots, M \Big\}. \tag{3}$$
- Maximal-power constrained codes,
$$\mathcal{L}_{\mathrm{m}}(\Gamma) \triangleq \Big\{ \mathcal{C} \;\Big|\; \|\boldsymbol{c}_i\|^2 \leq n\Gamma, \quad i = 1, \ldots, M \Big\}. \tag{4}$$
- Average-power constrained codes,
$$\mathcal{L}_{\mathrm{a}}(\Gamma) \triangleq \Big\{ \mathcal{C} \;\Big|\; \tfrac{1}{M} \textstyle\sum_{i=1}^{M} \|\boldsymbol{c}_i\|^2 \leq n\Gamma \Big\}. \tag{5}$$

Clearly, $\mathcal{L}_{\mathrm{e}}(\Gamma) \subset \mathcal{L}_{\mathrm{m}}(\Gamma) \subset \mathcal{L}_{\mathrm{a}}(\Gamma)$. While the equal-power constraint is easier to analyze, the maximal and average-power constraints are more useful in practice. Here, we present lower bounds on $P_{\mathrm{e}}(\mathcal{C})$ under equal and maximal-power constraints.

A. Shannon'59 lower bound

Let θ be the half-angle of a n-dimensional cone with vertex at the origin and with axis going through the vector $\boldsymbol{x} = (1, \ldots, 1)$. We denote by $\Phi_n(\theta, \sigma^2)$ the probability that such vector is moved outside this cone by effect of the i.i.d. Gaussian noise with variance σ^2 in each dimension.

Theorem 1 ([1, Eq. (33)]): Let $\mathcal{C} \in \mathcal{L}_{\mathrm{e}}(\Gamma)$ be a length-n code of cardinality M satisfying an equal power constraint.

Let $\theta_{n,M}$ denote the half-angle of a cone with solid angle equal to Ω_n/M, where Ω_n is the surface of the n-dimensional hypersphere. Then,

$$P_{\mathrm{e}}(\mathcal{C}) \geq \Phi_n\left(\theta_{n,M}, \frac{\sigma^2}{\Gamma}\right). \tag{6}$$

While this bound is conceptually simple and accurate for relatively short codes [3], it is difficult to evaluate. The computation of this bound is treated, e.g., in [4], [5].

B. PPV'10 lower bound

In [2], Polyanskiy *et al.* proved that the error probability of a binary hypothesis test with certain parameters can be used to lower bound the error probability $P_{\mathrm{e}}(\mathcal{C})$ for a certain channel $P_{\boldsymbol{Y}|\boldsymbol{X}}$. In particular, [2, Th. 27] shows that

$$P_{\mathrm{e}}(\mathcal{C}) \geq \inf_{P_{\boldsymbol{X}}} \sup_{Q_{\boldsymbol{Y}}} \left\{ \alpha_{\frac{1}{M}}\left(P_{\boldsymbol{X}} P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{X}} \times Q_{\boldsymbol{Y}}\right)\right\}, \tag{7}$$

where $\alpha_\beta(P,Q)$ is the minimum type-I error for a maximum type-II error $\beta \in [0,1]$ in a binary hypothesis testing problem between the distributions P and Q.

The bound (7) is usually referred to as the *meta-converse bound* since several converse bounds in the literature can be recovered from it via relaxation. While it is possible to restrict the set of distributions $Q_{\boldsymbol{Y}}$ over which the bound is maximized and still obtain a lower bound, the minimization over $P_{\boldsymbol{X}}$ needs to be carried out over all the n-dimensional probability distributions (not necessarily product) satisfying the power constraint considered.

For the Gaussian channel, Polyanskiy *et al.* fixed $Q_{\boldsymbol{Y}}$ to be zero-mean Gaussian distributed with variance θ^2 and independent entries, i.e., $Q_{\boldsymbol{Y}} = Q$ with pdf

$$q(\boldsymbol{y}) = \prod_{i=1}^{n} \varphi_{0,\theta}(y_i). \tag{8}$$

Particularizing (7) for this channel and fixing $Q_{\boldsymbol{Y}} = Q$, yields

$$P_{\mathrm{e}}(\mathcal{C}) \geq \inf_{P \in \mathcal{P}_\Gamma} \left\{ \alpha_{\frac{1}{M}}\left(PW, P \times Q\right)\right\}, \tag{9}$$

where the minimization is over all input distributions P satisfying a certain power constraint Γ, denoted by $\mathcal{P}_\Gamma$. For this choice of Q, $\alpha_{\frac{1}{M}}(\cdot,\cdot)$ presents spherical symmetry. Then, restricting the input codebook to lie on the surface of a n-dimensional hyper-sphere of squared radius $n\Gamma$ (equal power constraint), setting $\theta^2 = \Gamma + \sigma^2$, the following result follows.

Theorem 2 ([2, Th. 41]): Let $\mathcal{C} \in \mathcal{L}_{\mathrm{e}}(\Gamma)$ be a length-n code of cardinality M satisfying an equal power constraint. Then,

$$P_{\mathrm{e}}(\mathcal{C}) \geq \alpha_{\frac{1}{M}}\left(\varphi^n_{\sqrt{\Gamma},\sigma}, \varphi^n_{0,\theta}\right), \tag{10}$$

where $\theta^2 = \Gamma + \sigma^2$.

This expression can be evaluated via the probability of two noncentral χ^2 distributions (see Appendix A for details). However, for fixed rate $R \triangleq \frac{1}{n}\log_2 M$, the term $\frac{1}{M} = 2^{-nR}$ decreases exponentially with the block-length and traditional series series expansions of the noncentral χ^2 fail even for moderate values of n (see discussion in [2, p. 2326]).

(a) (b)

Fig. 1: Induced integration regions by (a) the Shannon'59 lower bound (6), and (b) the PPV'10 lower bound (10).

C. Comparison between Shannon'59 and PPV'10

Shannon'59 lower bound in Theorem 1 corresponds to the probability that the additive Gaussian noise moves a given codeword out of the n-dimensional cone centered at the codeword (cone that roughly covers $1/M$-th of the output space). We show next that the PPV'10 lower bound in Theorem 2 admits an analogous geometrical interpretation.

Let $\boldsymbol{x} = \left(\sqrt{\Gamma}, \ldots, \sqrt{\Gamma}\right)$ and let $\theta > \sigma$. For the hypothesis test on the right-hand side of (10), the condition

$$\frac{\varphi^n_{\sqrt{\Gamma},\sigma}(\boldsymbol{y})}{\varphi^n_{0,\theta}(\boldsymbol{y})} = \frac{\theta^n}{\sigma^n} \exp\left[\frac{\|\boldsymbol{y}\|^2}{2\theta^2} - \frac{\|\boldsymbol{y}-\boldsymbol{x}\|^2}{2\sigma^2}\right] = \gamma \tag{11}$$

for some $\gamma > 0$, defines the boundary of the decision region induced by the optimal Neyman-Pearson test. We next study the shape of this region. To this end, we note that

$$\frac{\|\boldsymbol{y}\|^2}{2\theta^2} - \frac{\|\boldsymbol{y}-\boldsymbol{x}\|^2}{2\sigma^2} = -\frac{\theta^2-\sigma^2}{2\sigma^2\theta^2}\left(\|\boldsymbol{y}\|^2 - 2a\langle\boldsymbol{x},\boldsymbol{y}\rangle + a\|\boldsymbol{x}\|^2\right) \tag{12}$$

$$= -\frac{\theta^2-\sigma^2}{2\sigma^2\theta^2}\left(\|\boldsymbol{y}-a\boldsymbol{x}\|^2 + (a-a^2)\|\boldsymbol{x}\|^2\right), \tag{13}$$

where $a = \frac{\theta^2}{\theta^2-\sigma^2} \geq 0$ for $\theta^2 \geq \sigma^2$, and where $\langle\boldsymbol{x},\boldsymbol{y}\rangle$ denotes the inner product between $\boldsymbol{x}$ and $\boldsymbol{y}$.

Using (13) with $\|\boldsymbol{x}\|^2 = n\Gamma$ and $\theta^2 = \Gamma + \sigma^2$, we obtain that the boundary of the decision region (11) becomes

$$\left\|\boldsymbol{y} - \left(1+\tfrac{\sigma^2}{\Gamma}\right)\boldsymbol{x}\right\|^2 = \bar{\gamma}, \tag{14}$$

where $\bar{\gamma} = n\sigma^2\left(1+\frac{\sigma^2}{\Gamma}\right)\left(1+\log\left(1+\frac{\Gamma}{\sigma^2}\right)+\frac{2}{n}\log(\gamma)\right)$.

As (14) corresponds to the equation of an n-dimensional sphere, we can alternatively describe the PPV'10 lower bound in Theorem 2 as the probability that the additive Gaussian noise moves the codeword $\boldsymbol{x}$ out of the n-dimensional sphere centered at $\left(1+\frac{\sigma^2}{\Gamma}\right)\boldsymbol{x}$ (that covers $1/M$-th of the output space). Note that the "regions" induced by Theorem 1 correspond to cones, while those induced by Theorem 2 correspond to spheres (see Fig. 1). Cones are close to the optimal ML decoding regions for codewords evenly distributed on surface of an n-dimensional sphere with squared radius $n\Gamma$.[1] On the other hand, "spherical regions" allow different configurations of the codewords inside the sphere. Then, the meta-converse bound may hold beyond the equal-power constraint.

This intuition is proven to be right in the next section.

[1]Indeed, in $n = 2$ dimensions Shannon'59 lower bound yields the exact error probability of an M-PSK constellation. See Section III-A for details.

III. Lower Bound for Maximal-power Constraints

In order to lower bound the error probability of a maximal-power constrained codebook we start by considering the general meta-converse in (7). In order to make the minimization over $P_{\boldsymbol{X}}$ in (7) tractable we shall use the following result.

Lemma 1 ([6, Lem. 25]): Let $P_{\boldsymbol{X}} = \sum_j \lambda_j P_{\boldsymbol{X}_j}$ with $\lambda_j > 0$, $\sum_j \lambda_j = 1$, be a convex combination of the distributions $P_{\boldsymbol{X}_j}$ and let $\{P_{\boldsymbol{X}_j}\}$ have pairwise disjoint supports. Then, the hypothesis testing error trade-off function satisfies

$$\alpha_\beta\big(P_{\boldsymbol{X}} P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{X}} \times Q_{\boldsymbol{Y}}\big) = \min_{\substack{\{\beta_j\}:\\ \beta=\sum_j \lambda_j\beta_j}} \sum_j \lambda_j \alpha_{\beta_j}\big(P_{\boldsymbol{X}_j} P_{\boldsymbol{Y}|\boldsymbol{X}}, P_{\boldsymbol{X}_j} \times Q_{\boldsymbol{Y}}\big). \quad (15)$$

This lemma asserts that it is possible to express the test (7) as a convex combination of disjoint sub-tests provided that the type-II error is optimally distributed among them. Applying this decomposition in (9) for the Gaussian channel under maximal power constraint, we obtain the following result.

Theorem 3 (Maximal power constraint): Let $\mathcal{C} \in \mathcal{L}_{\mathrm{m}}(\Gamma)$ be a length-n code of cardinality M satisfying a maximal power constraint and let $n \geq 1$. Then, for any $\theta > \sigma$,

$$P_{\mathrm{e}}(\mathcal{C}) \geq \alpha_{\frac{1}{M}}\big(\varphi^n_{\sqrt{\Gamma},\sigma}, \varphi^n_{0,\theta}\big). \quad (16)$$

Proof: For any $0 \leq \rho \leq \sqrt{\Gamma}$, we define the input set $\mathcal{S}_\rho \triangleq \{\boldsymbol{x} \,|\, \|\boldsymbol{x}\|^2 = n\rho^2\}$. Then, any input distribution $P_{\boldsymbol{X}}$ induces a distribution over the parameter ρ, $P_\rho \triangleq \Pr\{\mathcal{S}_\rho\}$. We consider the conditional distribution

$$\mathrm{d}P_{\boldsymbol{X}|\rho}(\boldsymbol{x}) = \begin{cases} \frac{\mathrm{d}P_{\boldsymbol{X}}(\boldsymbol{x})}{\mathrm{d}P_\rho}, & \boldsymbol{x} \in \mathcal{S}_\rho,\\ 0, & \text{otherwise.}\end{cases} \quad (17)$$

It follows that $P_{\boldsymbol{X}}(\boldsymbol{x}) = \int P_{\boldsymbol{X}|\rho}(\boldsymbol{x})\,\mathrm{d}P_\rho$ with $\mathrm{d}P_\rho$ satisfying $\mathrm{d}P_\rho \geq 0$, $\int \mathrm{d}P_\rho = 1$. Then, we apply Lemma 1 to the right-hand side of (9) to obtain

$$\inf_{P\in\mathcal{P}_\Gamma} \Big\{\alpha_{\frac{1}{M}}\big(PW, P\times Q\big)\Big\} = \inf_{\substack{\{P_\rho,\beta_\rho\}:\\ \int\beta_\rho \,\mathrm{d}P_\rho = \frac{1}{M}}} \left\{\int \alpha_{\beta_\rho}\big(P_\rho W, P_\rho \times Q\big)\,\mathrm{d}P_\rho\right\} \quad (18)$$

$$= \inf_{\substack{\{P_\rho,\beta_\rho\}:\\ \int\beta_\rho \,\mathrm{d}P_\rho = \frac{1}{M}}} \left\{\int \alpha_{\beta_\rho}\big(\varphi^n_{\rho,\sigma}, \varphi^n_{0,\theta}\big)\,\mathrm{d}P_\rho\right\}, \quad (19)$$

where the last step follows from the spherical symmetry of each of the sub-tests in (18) and since $\boldsymbol{x} = (\rho,\ldots,\rho) \in \mathcal{S}_\rho$.

To solve the optimization in (19) we resort in the following lemma, which is then proven in the appendices.

Lemma 2: Let $\sigma < \theta$, with $\sigma, \theta \in \mathbb{R}^+$ and $n \geq 1$. Then, $\alpha_\beta\big(\varphi^n_{\rho,\sigma}, \varphi^n_{0,\theta}\big)$ is non-increasing in ρ for any fixed $\beta \in [0,1]$.

According to Lemma 2, for any $0 \leq \rho \leq \sqrt{\Gamma}$, it holds that $\alpha_\beta\big(\varphi^n_{\rho,\sigma}, \varphi^n_{0,\theta}\big) \geq \alpha_\beta\big(\varphi^n_{\sqrt{\Gamma},\sigma}, \varphi^n_{0,\theta}\big)$. As any maximal-power constrained input distribution $P \in \mathcal{P}_\Gamma$ satisfies $P_\rho = 0$ for $\rho > \sqrt{\Gamma}$, we conclude that

$$\inf_{\substack{\{P_\rho,\beta_\rho\}:\\ \int\beta_\rho \,\mathrm{d}P_\rho = \frac{1}{M}}} \left\{\int \alpha_{\beta_\rho}\big(\varphi^n_{\rho,\sigma}, \varphi^n_{0,\theta}\big)\,\mathrm{d}P_\rho\right\}$$

$$\geq \inf_{\substack{\{P_\rho,\beta_\rho\}:\\ \int\beta_\rho \,\mathrm{d}P_\rho = \frac{1}{M}}} \left\{\int \alpha_{\beta_\rho}\big(\varphi^n_{\sqrt{\Gamma},\sigma}, \varphi^n_{0,\theta}\big)\,\mathrm{d}P_\rho\right\} \quad (20)$$

$$\geq \alpha_{\frac{1}{M}}\big(\varphi^n_{\sqrt{\Gamma},\sigma}, \varphi^n_{0,\theta}\big), \quad (21)$$

where in (21) we used that the function $\alpha_\beta(\cdot,\cdot)$ is convex with respect to β, hence, $\int \alpha_{\beta_\rho}(\cdot,\cdot)\,\mathrm{d}P_\rho \geq \alpha_{\int \beta_\rho \mathrm{d}P_\rho}(\cdot,\cdot)$.

Then, using (9), (19) and (21) the result follows. ■

Setting $\theta^2 = \Gamma + \sigma^2$ in Theorem 3, we recover the bound in Theorem 2. We conclude that the bound in Theorem 2 also holds for maximal power constraint. This is not the case however for the Shannon'59 lower bound in Theorem 1, as we show next with an example.

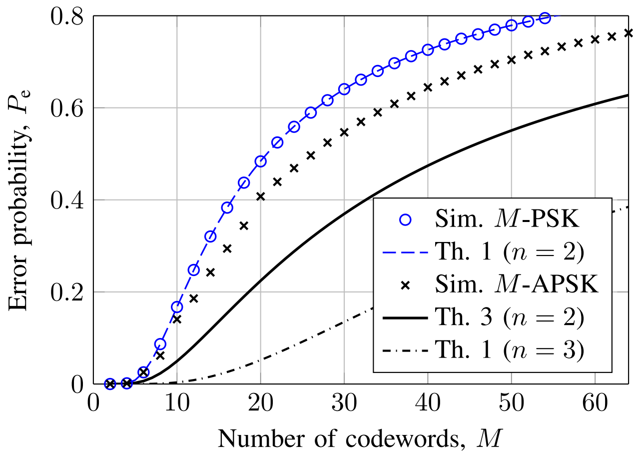

Fig. 2: Lower bounds to the channel coding error probability over an AWGN channel with $n = 2$ and SNR= 10 dB.

A. Example: 2-dimensional constellations

We consider the problem of transmitting $M \geq 2$ codewords over a additive Gaussian noise channel with $n = 2$ dimensions. Figure 2 compares the bounds in Theorem 1 (evaluated for $n = 2$ and $n = 3$) and Theorem 3 with $\theta^2 = \Gamma + \sigma^2$. For reference, we include the simulated ML decoding error probability of an M-PSK (phase-shift keying) and M-APSK (amplitude-phase-shift keying) constellations satisfying the maximal power constraint. For $n = 2$, Shannon'59 lower bound in Theorem 1 coincides with the ML decoding error probability of the M-PSK constellation (as the 2-dimensional cones are precisely the ML decoding regions of the M-PSK constellation). Theorem 1 only applies for codebooks (or constellations) satisfying the equal power constraint. Indeed, the M-APSK simulated error probability violates the bound evaluated for $n = 2$. Theorem 3 applies to both equal and maximal power constraints, as it does Theorem 1 evaluated for $n = 3$. We can see that Theorem 3 is tighter in this setting.

Appendix A
Proof of Lemma 2

Let $\sigma, \theta > 0$ and $n \geq 1$, be fixed parameters. We define

$$\jmath_\rho(\boldsymbol{y}) \triangleq \log \frac{\varphi^n_{\rho,\sigma}(\boldsymbol{y})}{\varphi^n_{0,\theta}(\boldsymbol{y})} \tag{22}$$

$$= \log \frac{\theta}{\sigma} + \frac{1}{2} \sum_{i=1}^{n} \frac{\theta^2 (y_i - \rho)^2 - \sigma^2 y_i^2}{\sigma^2 \theta^2}. \tag{23}$$

The trade-off $\alpha_\beta\big(\varphi^n_{\rho,\sigma}, \varphi^n_{0,\theta}\big)$ admits the parametric form

$$\alpha(\rho, \gamma) = \Pr\big[\jmath_\rho(\boldsymbol{Y}_0) \leq \gamma\big] = \Pr\big[\jmath_{0,\rho}(\boldsymbol{Z}) \leq \gamma\big], \tag{24}$$

$$\beta(\rho, \gamma) = \Pr\big[\jmath_\rho(\boldsymbol{Y}_1) > \gamma\big] = \Pr\big[\jmath_{1,\rho}(\boldsymbol{Z}) > \gamma\big], \tag{25}$$

in terms of the auxiliary parameter $\gamma \in \mathbb{R}$. Here, $\boldsymbol{Y}_0 \sim \varphi^n_{\rho,\sigma}$, $\boldsymbol{Y}_1 \sim \varphi^n_{0,\theta}$ and, for $\boldsymbol{Z} \sim \varphi^n_{0,1}$ and $\delta \triangleq \theta^2 - \sigma^2$, we defined

$$\jmath_{0,\rho}(\boldsymbol{z}) \triangleq \log \frac{\theta}{\sigma} - \frac{n}{2} \frac{\rho^2}{\delta} + \frac{1}{2} \frac{\delta}{\sigma^2} \sum_{i=1}^{n} \left(z_i - \frac{\sigma \rho}{\delta} \right)^2, \tag{26}$$

$$\jmath_{1,\rho}(\boldsymbol{z}) \triangleq \log \frac{\theta}{\sigma} - \frac{n}{2} \frac{\rho^2}{\delta} + \frac{1}{2} \frac{\delta}{\theta^2} \sum_{i=1}^{n} \left(z_i - \frac{\theta \rho}{\delta} \right)^2. \tag{27}$$

The equivalence between the 1st and 2nd identities in (24) and (25) follows from (23), (26) and (27) via a change of variables.

Given (26) and (27), since $\boldsymbol{Z} \sim \varphi^n_{0,1}$, we conclude that $\jmath_{0,\rho}(\boldsymbol{Z})$ and $\jmath_{1,\rho}(\boldsymbol{Z})$ follow a (shifted and scaled) noncentral χ^2 distribution with n degrees of freedom and non-centrality parameters $n\sigma^2\rho^2/\delta^2$ and $n\theta^2\rho^2/\delta^2$, respectively. The cdf of a noncentral χ^2 distribution can be written in terms of the generalized Marcum Q-function $Q_m(a, b)$ defined in (37). Then, using (24), (25), (26) and (27), we characterize $\alpha_\beta\big(\varphi^n_{\rho,\sigma}, \varphi^n_{0,\theta}\big)$ as a function of an auxiliary parameter $\tilde{\gamma} \geq 0$ as

$$\alpha(\rho, \tilde{\gamma}) = Q_{\frac{n}{2}} \left(\sqrt{n} \frac{\sigma \rho}{\delta}, \frac{\tilde{\gamma}}{\sigma} \right), \tag{28}$$

$$\beta(\rho, \tilde{\gamma}) = 1 - Q_{\frac{n}{2}} \left(\sqrt{n} \frac{\theta \rho}{\delta}, \frac{\tilde{\gamma}}{\theta} \right). \tag{29}$$

To prove that $\alpha_\beta\big(\varphi^n_{\rho,\sigma}, \varphi^n_{0,\theta}\big)$ is non-increasing in ρ, we need to show that its derivative with respect to ρ is non-positive. To this end, we could invert (29) to obtain the dependence of $\tilde{\gamma}$ with ρ for fixed β and substitute this $\tilde{\gamma}(\rho)$ in (28) before taking the derivative. However, given the nature of the functions involved, there is no closed-form expression for $\tilde{\gamma}(\rho)$. Instead, we use the chain rule for total derivatives to write

$$\frac{\partial \beta(\rho, \tilde{\gamma})}{\partial \rho} = \frac{\partial \beta(\rho, \tilde{\gamma})}{\partial \rho} + \frac{\partial \beta(\rho, \tilde{\gamma})}{\partial \tilde{\gamma}} \frac{\partial \tilde{\gamma}}{\partial \rho}. \tag{30}$$

As β is fixed, we set (30) equal to 0 and solve for $\frac{\partial \tilde{\gamma}}{\partial \rho}$. Then,

$$\frac{\partial \tilde{\gamma}}{\partial \rho} = - \frac{\frac{\partial}{\partial \rho} \beta(\rho, \tilde{\gamma})}{\frac{\partial}{\partial \tilde{\gamma}} \beta(\rho, \tilde{\gamma})} = \frac{I_{\frac{n}{2}}\big(\sqrt{n}\frac{\rho\tilde{\gamma}}{\delta}\big)\sqrt{n}\frac{\theta}{\delta}}{I_{\frac{n}{2}-1}\big(\sqrt{n}\frac{\rho\tilde{\gamma}}{\delta}\big)\frac{1}{\theta}}, \tag{31}$$

where $I_m(\cdot)$ is the m-th order modified Bessel function of the first kind and where we used that (see Appendix B)

$$\frac{\partial Q_m(a, b)}{\partial a} = \frac{b^m}{a^{m-1}} e^{-\frac{a^2+b^2}{2}} I_m(ab), \tag{32}$$

$$\frac{\partial Q_m(a, b)}{\partial b} = - \frac{b^m}{a^{m-1}} e^{-\frac{a^2+b^2}{2}} I_{m-1}(ab). \tag{33}$$

We now evaluate the derivative of $\frac{\partial \alpha}{\partial \rho}$ for fixed β. By applying the chain rule for total derivatives and using (31), (32) and (33), we obtain

$$\frac{\partial \alpha(\rho, \tilde{\gamma})}{\partial \rho} = \frac{\partial \alpha(\rho, \tilde{\gamma})}{\partial \rho} + \frac{\partial \alpha(\rho, \tilde{\gamma})}{\partial \tilde{\gamma}} \frac{\partial \tilde{\gamma}}{\partial \rho} \tag{34}$$

$$= - \frac{\sqrt{n}}{\sigma} \frac{b^{\frac{n}{2}}}{a^{\frac{n}{2}-1}} e^{-\frac{a^2+b^2}{2}} I_{\frac{n}{2}}\big(\sqrt{n}\tfrac{\rho\tilde{\gamma}}{\delta}\big) \tag{35}$$

$$= - \frac{n\rho}{\delta} \Big(\frac{\tilde{\gamma}\delta}{\sqrt{n}\sigma^2\rho} \Big)^{\frac{n}{2}} e^{-\frac{n\sigma^4\rho^2 + \delta^2\tilde{\gamma}^2}{2\delta^2\sigma^2}} I_{\frac{n}{2}}\big(\sqrt{n}\tfrac{\rho\tilde{\gamma}}{\delta}\big) \tag{36}$$

where $a = \sqrt{n}\frac{\sigma\rho}{\delta}$ and $b = \frac{\tilde{\gamma}}{\sigma}$ in (35). As (36) is non-positive for $\delta = \theta^2 - \sigma^2 > 0$, then Lemma 2 follows.

Appendix B
Derivatives of the Marcum-Q function

For $a > 0$ and $b > 0$, the Marcum-Q function is defined as

$$Q_m(a, b) \triangleq \int_b^\infty \frac{t^m}{a^{m-1}} e^{-\frac{t^2+a^2}{2}} I_{m-1}(at) \, \mathrm{d}t. \tag{37}$$

The derivative (33) then follows directly from (37). For (32) we make use of the series representation [7, Eq. (4.62)]

$$Q_m(a, b) = e^{-\frac{t^2+a^2}{2}} \sum_{r=1-m}^{\infty} \Big(\frac{a}{b}\Big)^r I_{-r}(ab) \tag{38}$$

and we write its derivative with respect to a to obtain

$$\frac{\partial Q_m(a, b)}{\partial a} = e^{-\frac{t^2+a^2}{2}} \sum_{1-m}^{\infty} \Big(\frac{a}{b}\Big)^r \bigg(\Big(\frac{r}{a} - a \Big) I_{-r}(ab) + b I'_{-r}(ab) \bigg). \tag{39}$$

Using the identity $I'_m(x) = \frac{m}{x} I_m(x) + I_{m+1}(x)$ [8, Sec. 8.486] and canceling terms we obtain (32). To the best of our knowledge, the form of the derivative in (32) does not appear in the literature for non-integer values of m. For integer values of m, (32) can be easily obtained from (37) by using the identities $Q_m(a, b) = 1 - Q_{1-m}(b, a)$ and $I_m(x) = I_{-m}(x)$.

Acknowledgment

Fruitful discussions with Barış Nakiboğlu, Tobias Koch and David Morales-Jimenez are gratefully acknowledged.

References

[1] C. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell System Technical Journal*, vol. 38, p. 611656, 1959.

[2] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[3] I. Sason and S. Shamai (Shitz), *Performance analysis of linear codes under maximum-likelihood decoding: a tutorial*. Foundations and Trends Commun. and Inf. Theory, now Publishers, 2006.

[4] A. Valembois and M. Fossorier, "Sphere-packing bounds revisited for moderate block lengths," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 2998–3014, 2004.

[5] G. Wiechman and I. Sason, "An improved sphere-packing bound for finite-length codes over symmetric memoryless channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1962–1990, May 2008.

[6] Y. Polyanskiy, "Saddle point in the minimax converse for channel coding," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2576–2595, May 2013.

[7] M.-S. A. Marvin K. Simon, *Digital Communication over Fading Channels*, 2nd ed. New Jersey: Wiley-IEEE Press, 2004.

[8] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. London: Elsevier, 2007.