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

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

***Abstract*—This paper studies the performance of block coding on an additive white Gaussian noise channel under different power limitations at the transmitter. Lower bounds are presented for the minimum error probability of codes satisfying an average power constraint. These bounds are tighter than previous results in the literature, and yield a better understanding on the structure of good codes under an average power limitation.**

I. Introduction

We consider the problem of transmitting equiprobable messages over several uses of an additive white Gaussian noise (AWGN) channel. We consider different power constraints: *equal power constraint* (all the codewords in the transmission code have equal energy); *maximal power constraint* (the energy of all the codewords is below a certain threshold); and *average power constraint* (while some codewords may violate the threshold, the energy budget is satisfied in average).

In his 1959 paper, Shannon derived a lower bound to the error probability of any equal power constrained codebook via geometrical arguments [1, Eq. (20)]. Following a different approach, Polyanskiy, Poor and Verdú applied a particular instance of a binary hypothesis test to lower bound the same error probability [2, Th. 41]. While [2, Th. 41] was derived originally under an equal power constraint, it was recently shown to also hold under a maximal power constraint [3, Th. 3]. Other connections among the system performance under the three power constraints are studied in [1, Sec. XIII] (see also [2, Lem. 39]).

In this work, we establish direct lower bounds for codes satisfying an average power limitation at the transmitter. Our analysis is based on the meta-converse bound [2, Th. 27] evaluated for auxiliary Gaussian distributions. We characterize the error probability of the binary hypothesis test appearing in this bound for the AWGN channel, and use its properties to avoid the optimization over input distributions. Our results show that, if the cardinality of the codebook is below a certain threshold, [2, Th. 41] and [3, Th. 3] hold under an average power limitation without any modifications. The resulting bound is tighter than previous results in the literature for the same power constraint and provide an accurate characterization of the error probability for a wide range of system parameters.

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 grant TEC2016-78434-C3 (AEI/FEDER, EU).

II. System Model

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

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

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

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

In our communications system, a source produces a certain message $v \in \{1, \dots, M\}$ randomly with equal probability. This message is mapped by the encoder to a codeword $\boldsymbol{c}_v$ according to a codebook $\mathcal{C} \triangleq \{\boldsymbol{c}_1, \dots, \boldsymbol{c}_M\}$, and the sequence $\boldsymbol{x} = \boldsymbol{c}_v$ is transmitted over the channel. Then, based on the channel output $\boldsymbol{y}$, the decoder guesses the transmitted message $\hat{v} \in \{1, \dots, M\}$. We define the average error probability

$$P_{\mathrm{e}}(\mathcal{C}) \triangleq \Pr\{\hat{V} \neq V\}, \tag{3}$$

where the underlying probability is induced by the chain of source, encoder, channel and decoder.

We consider codebooks satisfying the following constraints:

- Equal power constraint Υ:

$$\mathcal{L}_{\mathrm{e}}(n, M, \Upsilon) \triangleq \Big\{ \mathcal{C} \;\Big|\; \|\boldsymbol{c}_i\|^2 = n\Upsilon,\ i = 1, \dots, M \Big\} \tag{4}$$

- Maximal power constraint Υ:

$$\mathcal{L}_{\mathrm{m}}(n, M, \Upsilon) \triangleq \Big\{ \mathcal{C} \;\Big|\; \|\boldsymbol{c}_i\|^2 \le n\Upsilon,\ i = 1, \dots, M \Big\} \tag{5}$$

- Average power constraint Υ:

$$\mathcal{L}_{\mathrm{a}}(n, M, \Upsilon) \triangleq \Big\{ \mathcal{C} \;\Big|\; \tfrac{1}{M} \textstyle\sum_{i=1}^{M} \|\boldsymbol{c}_i\|^2 \le n\Upsilon \Big\} \tag{6}$$

Clearly, $\mathcal{L}_{\mathrm{e}} \subseteq \mathcal{L}_{\mathrm{m}} \subseteq \mathcal{L}_{\mathrm{a}}$. In the following, we study lower bounds on the error probability $P_{\mathrm{e}}(\mathcal{C})$ under equal, maximal and average power constraints. While derivation of converse bounds is easier under an equal power constraint, the maximal power and average power constraints are more relevant for practical applications.

III. Meta-converse Bound for Equal and Maximal Power Constraints

In [2], Polyanskiy *et al.* proved that the error probability of a binary hypothesis test with certain parameters lower bounds the error probability $P_e(\mathcal{C})$ for a certain channel W. In particular, [2, Th. 27] shows that

$$P_e(\mathcal{C}) \geq \inf_{P\in\mathcal{P}} \sup_{Q} \left\{\alpha_{\frac{1}{M}}\left(PW, P\times Q\right)\right\}, \tag{7}$$

where $\mathcal{P}$ is the set of distributions over the input alphabet $\mathcal{X}^n$ satisfying a certain constraint and Q is an auxiliary distribution over the output alphabet $\mathcal{Y}^n$ which is not allowed to depend on the input $\boldsymbol{x}$. Here $\alpha_\beta\left(A,B\right)$ denotes the minimum type-I error for a maximum type-II error $\beta\in[0,1]$ in a binary hypothesis test between the distributions A and B. Specifically, for two distributions A and B defined over an alphabet $\mathcal{Z}$, the function $\alpha_\beta\left(A,B\right)$ is given by

$$\alpha_\beta(A,B) \triangleq \inf_{\substack{0\leq T\leq 1:\\ \mathrm{E}_B[T(Z)]\leq\beta}} \Big\{1-\mathrm{E}_A[T(Z)]\Big\}, \tag{8}$$

where $T:\mathcal{Z}\to[0,1]$ and $\mathrm{E}_P[\cdot]$ is the expectation operator with respect to the random variable $Z\sim P$.

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

For the Gaussian channel W, Polyanskiy *et al.* fixed Q to be zero-mean Gaussian distributed with variance θ^2 and independent entries, i.e., with pdf

$$q(\boldsymbol{y}) = \prod_{i=1}^{n} \varphi_{0,\theta}(y_i). \tag{9}$$

For this choice of Q, $\alpha_{\frac{1}{M}}(\cdot,\cdot)$ presents spherical symmetry. Then, restricting the input distribution to lie on the surface of a n-dimensional hyper-sphere of squared radius $n\Upsilon$ and setting $\theta^2=\Upsilon+\sigma^2$, they obtained the following result.

Theorem 1 (Converse, equal power constraint [2, Th. 41]): Let $\mathcal{C}\in\mathcal{L}_e(n,M,\Upsilon)$ be a length-n code of cardinality M satisfying an equal power constraint. Then, for $\theta^2=\Upsilon+\sigma^2$,

$$P_e(\mathcal{C}) \geq \alpha_{\frac{1}{M}}\left(\varphi^n_{\sqrt{\Upsilon},\sigma}, \varphi^n_{0,\theta}\right). \tag{10}$$

The bound in Theorem 1 can be extended to maximal and average power constraints using, e.g., [2, Lem. 39]. A direct lower bound under maximal power constraint is given next.

Theorem 2 (Converse, maximal power constraint [3, Th. 3]): Let $\mathcal{C}\in\mathcal{L}_m(n,M,\Upsilon)$ be a length-n code of cardinality M satisfying a maximal power constraint. For any $\theta\geq\sigma$, $n\geq 1$, the lower bound (10) holds for this code.

The bounds in Theorems 1 and 2 coincide for equal and maximal power constraints. Then, one may wonder if this is also the case for codes satisfying an average power constraint. In Section IV, we will show that the lower bound (10) holds in this setting under certain conditions (but not in general).

A. Computation of $\alpha_\beta\left(\varphi^n_{\sqrt{\gamma},\sigma}, \varphi^n_{0,\theta}\right)$

Computation of Theorems 1 and 2 require to evaluate

$$f(\beta,\gamma) \triangleq \alpha_\beta\left(\varphi^n_{\sqrt{\gamma},\sigma}, \varphi^n_{0,\theta}\right). \tag{11}$$

We next provide a parametric formulation of this function.

Proposition 1: Let $\sigma,\theta>0$ and $n\geq 1$, be fixed parameters, and define $\delta\triangleq\theta^2-\sigma^2$. The trade-off between α and β admits the following parametric formulation as a function of the auxiliary parameter $t\geq 0$,

$$\alpha(\gamma,t) = Q_{\frac{n}{2}}\left(\sqrt{n\gamma}\frac{\sigma}{\delta}, \frac{t}{\sigma}\right), \tag{12}$$

$$\beta(\gamma,t) = 1 - Q_{\frac{n}{2}}\left(\sqrt{n\gamma}\frac{\theta}{\delta}, \frac{t}{\theta}\right), \tag{13}$$

where $Q_m(x,y)$ denotes the generalized Marcum Q-function. Let t_b satisfy $\beta(\gamma,t_b)=b$ according to (13). Then, the function (11) is given by $f(b,\gamma)=\alpha(\gamma,t_b)$ according to (12).

Proof outline: Following the lines of the proof of [2, Th. 41], we obtain a parametric formulation in terms of two non-central χ^2 distributions. Then, to recover (12)-(13), we write the cumulative density function $F_{n,\nu}(x)$ of a non-central χ^2 distribution with n degrees of freedom and non-centrality parameter ν in terms of the generalized Marcum Q-function $Q_m(a,b)$ as $F_{n,\nu}(x)=1-Q_{\frac{n}{2}}\left(\sqrt{\nu},\sqrt{x}\right)$. ■

In Proposition 1, we need to invert the marcum-Q function in (13) to evaluate $f(\beta,\gamma)$. The following alternative expression is more adequate for implementation purposes, as it only requires to solve a one dimensional optimization problem.

Corollary 1: Let $\sigma,\theta>0$ and $n\geq 1$, be fixed parameters. The function $f(\beta,\gamma)=\alpha_\beta\left(\varphi^n_{\sqrt{\gamma},\sigma}, \varphi^n_{0,\theta}\right)$ is given by

$$f(\beta,\gamma) = \max_{t\geq 0}\left\{ Q_{\frac{n}{2}}\left(\sqrt{n\gamma}\frac{\sigma}{\delta}, \frac{t}{\sigma}\right) + \frac{\theta^n}{\sigma^n} e^{\frac{1}{2}\left(\frac{n\gamma}{\delta}-\frac{\delta t^2}{\sigma^2\theta^2}\right)} \times\left(1-\beta-Q_{\frac{n}{2}}\left(\sqrt{n\gamma}\frac{\theta}{\delta}, \frac{t}{\theta}\right)\right)\right\}. \tag{14}$$

Proof outline: We define

$$j(\boldsymbol{y}) \triangleq \log\frac{\varphi^n_{\sqrt{\gamma},\sigma}(\boldsymbol{y})}{\varphi^n_{0,\theta}(\boldsymbol{y})} \tag{15}$$

$$= n\log\frac{\theta}{\sigma} - \frac{1}{2}\sum_{i=1}^{n}\frac{\theta^2(y_i-\sqrt{\gamma})^2-\sigma^2 y_i^2}{\sigma^2\theta^2}. \tag{16}$$

According to the Neyman-Pearson lemma, the trade-off $\alpha_\beta\left(\varphi^n_{\sqrt{\gamma},\sigma}, \varphi^n_{0,\theta}\right)$ admits the parametric form

$$\alpha(t') = \Pr\left[j(\boldsymbol{Y}_0)\leq t'\right], \tag{17}$$

$$\beta(t') = \Pr\left[j(\boldsymbol{Y}_1)> t'\right], \tag{18}$$

in terms of $t'\in\mathbb{R}$ and where $\boldsymbol{Y}_0\sim\varphi^n_{\sqrt{\gamma},\sigma}$, $\boldsymbol{Y}_1\sim\varphi^n_{0,\theta}$.

We apply [4, Lem. 1] to the tail probabilities (17)-(18) and consider the change of variables $t'\leftrightarrow t$, which are related as $t^2=2\sigma^2\theta^2\frac{1}{\delta}\left(n\log\frac{\theta}{\sigma}+\frac{n}{2}\frac{\gamma}{\delta}-t'\right)$. Then, to obtain the desired result, we proceed as in the proof of Proposition 1 and use that $e^{t'}=\frac{\theta^n}{\sigma^n}\exp\left\{\frac{1}{2}\left(\frac{n\gamma}{\delta}-\frac{\delta t^2}{\sigma^2\theta^2}\right)\right\}$. ■

IV. Lower Bounds for Average-Power Constraint

The Legendre-Fenchel (LF) transform of a function g is

$$g^*(b) = \max_{a\in\mathcal{A}}\bigl\{\langle a,b\rangle - g(a)\bigr\}, \tag{19}$$

where $\mathcal{A}$ is the domain of the function g and $\langle a,b\rangle$ denotes the interior product between a and b.

The function g^* is usually referred to as Fenchel's conjugate (or convex conjugate) of g. If g is a convex function with closed domain, applying the LF transform twice recovers the original function, *i.e.*, $g^{**} = g$. If g is not convex, applying the LF transform twice returns the lower convex envelope of g, which is the largest lower semi-continuous convex function function majorized by g. For our problem, for $f(\beta,\gamma)$ in (11), we define

$$\underline{f}(\beta,\gamma) \triangleq f^{**}(\beta,\gamma), \tag{20}$$

and note that $\underline{f}(\beta,\gamma) \le f(\beta,\gamma)$ for any $\beta\in[0,1]$ and $\gamma\ge 0$.

The lower convex envelope (20) is a lower bound to the error probability in the average power constraint setting.

Theorem 3 (Converse, average power constraint): Let $\mathcal{C}\in\mathcal{L}_{\mathrm{a}}(n,M,\Upsilon)$ be a length-n code of cardinality M satisfying the average power constraint Υ. Then, for any $\theta\ge\sigma$, $n\ge 1$,

$$P_{\mathrm{e}}(\mathcal{C}) \ge \underline{f}\bigl(\tfrac{1}{M},\Upsilon\bigr), \tag{21}$$

where $\underline{f}(\beta,\gamma)$ is the lower convex envelope (20) of $f(\beta,\gamma)$ defined in (11).

Proof: We start by considering the general meta-converse bound in (7) with $\mathcal{P} = \mathcal{P}_{\mathrm{a}}(\Upsilon)$ corresponding to the set of distributions satisfying an average power constraint, i.e.,

$$\mathcal{P}_{\mathrm{a}}(\Upsilon) \triangleq \Bigl\{\boldsymbol{X}\sim P_{\boldsymbol{X}}\ \Bigm|\ \mathrm{E}\bigl[\|\boldsymbol{X}\|^2\bigr]\le n\Upsilon\Bigr\}. \tag{22}$$

To solve the minimization over P in (7) we shall use the following decomposition. For any $\gamma\ge 0$, we define the set $\mathcal{S}_\gamma \triangleq \bigl\{\boldsymbol{x}\,|\,\|\boldsymbol{x}\|^2 = n\gamma\bigr\}$. Then, any input distribution $P_{\boldsymbol{X}}$ induces a distribution over the parameter γ, $P_\gamma \triangleq \Pr\{\boldsymbol{X}\in\mathcal{S}_\gamma\}$, and a conditional distribution

$$\mathrm{d}P_{\boldsymbol{X}|\gamma}(\boldsymbol{x}) = \begin{cases}\frac{\mathrm{d}P_{\boldsymbol{X}}(\boldsymbol{x})}{\mathrm{d}P_\gamma}, & \boldsymbol{x}\in\mathcal{S}_\gamma,\\ 0, & \text{otherwise.}\end{cases} \tag{23}$$

It follows that $P_{\boldsymbol{X}}(\boldsymbol{x}) = \int P_{\boldsymbol{X}|\gamma}(\boldsymbol{x})\,\mathrm{d}P_\gamma$. Here, $\mathrm{d}P_\gamma\ge 0$ and $\int\mathrm{d}P_\gamma = 1$. Furthermore, the conditional distributions $P_{\boldsymbol{X}|\gamma}$ have disjoint support. Then, we apply [5, Lem. 25] to write

$$\begin{aligned}
&\inf_{P\in\mathcal{P}_{\mathrm{a}}(\Upsilon)}\Bigl\{\alpha_{\frac{1}{M}}\bigl(PW, P\times Q\bigr)\Bigr\}\\
&= \inf_{\substack{\{P_\gamma,\beta_\gamma\}:\\ \int\gamma\,\mathrm{d}P_\gamma=\Upsilon\\ \int\beta_\gamma\,\mathrm{d}P_\gamma=\frac{1}{M}}}\left\{\int\alpha_{\beta_\gamma}\bigl(P_\gamma W, P_\gamma\times Q\bigr)\,\mathrm{d}P_\gamma\right\} \qquad (24)\\
&= \inf_{\substack{\{P_\gamma,\beta_\gamma\}:\\ \int\gamma\,\mathrm{d}P_\gamma=\Upsilon\\ \int\beta_\gamma\,\mathrm{d}P_\gamma=\frac{1}{M}}}\left\{\int\alpha_{\beta_\gamma}\bigl(\varphi^n_{\sqrt{\gamma},\sigma},\varphi^n_{0,\theta}\bigr)\,\mathrm{d}P_\gamma\right\}, \qquad (25)
\end{aligned}$$

where the last step follows from the spherical symmetry of each of the tests in (24), using that $\boldsymbol{x} = (\sqrt{\gamma},\ldots,\sqrt{\gamma})\in\mathcal{S}_\gamma$.

Using that $\underline{f}(\beta,\gamma)\le f(\beta,\gamma) = \alpha_\beta\bigl(\varphi^n_{\sqrt{\gamma},\sigma},\varphi^n_{0,\theta}\bigr)$, we lower-bound the right-hand side of (25) as

$$\begin{aligned}
&\inf_{\substack{\{P_\gamma,\beta_\gamma\}:\\ \int\gamma\,\mathrm{d}P_\gamma=\Upsilon\\ \int\beta_\gamma\,\mathrm{d}P_\gamma=\frac{1}{M}}}\left\{\int f\bigl(\beta_\gamma,\gamma\bigr)\,\mathrm{d}P_\gamma\right\}\\
&\ge \inf_{\substack{\{P_\gamma,\beta_\gamma\}:\\ \int\gamma\,\mathrm{d}P_\gamma=\Upsilon\\ \int\beta_\gamma\,\mathrm{d}P_\gamma=\frac{1}{M}}}\left\{\int \underline{f}\bigl(\beta_\gamma,\gamma\bigr)\,\mathrm{d}P_\gamma\right\} \qquad (26)\\
&\ge \inf_{\substack{\{P_\gamma,\beta_\gamma\}:\\ \int\gamma\,\mathrm{d}P_\gamma=\Upsilon\\ \int\beta_\gamma\,\mathrm{d}P_\gamma=\frac{1}{M}}}\Bigl\{\underline{f}\bigl(\tfrac{1}{M},\Upsilon\bigr)\Bigr\} \qquad (27)\\
&= \underline{f}\bigl(\tfrac{1}{M},\Upsilon\bigr), \qquad (28)
\end{aligned}$$

where (27) follows by applying Jensen's inequality since $\underline{f}(\beta,\gamma)$ is jointly convex in both parameters and by using the constraints; and (28) holds since the objective of the optimization in (27) does not depend on $\{P_\gamma,\beta_\gamma\}$.

The lower bound (21) then follows from combining (7), (24)-(25) and the inequalities (26)-(28). ■

The function $\underline{f}(\beta,\gamma)$ can be evaluated numerically by considering a 2-dimensional grid of the parameters (β,γ), using (14) to compute $f(\beta,\gamma)$ over this grid, and obtaining the corresponding convex envelope. Nevertheless, sometimes $\underline{f}\bigl(\frac{1}{M},\Upsilon\bigr) = f\bigl(\frac{1}{M},\Upsilon\bigr) = \alpha_{\frac{1}{M}}\bigl(\varphi^n_{\sqrt{\Upsilon},\sigma},\varphi^n_{0,\theta}\bigr)$ and these steps can be avoided, as the next result shows.

Corollary 2: Let $\sigma,\theta>0$ and $n\ge 1$, be fixed parameters, and define $\delta\triangleq\theta^2-\sigma^2$. For $t\ge 0$, we define

$$\xi_1(t) \triangleq Q_{\frac{n}{2}}\left(\sqrt{n\Upsilon}\frac{\sigma}{\delta},\frac{t}{\sigma}\right) - Q_{\frac{n}{2}}\left(0,\sqrt{\bigl(\tfrac{t^2}{\sigma^2}-n\Upsilon\tfrac{\theta^2}{\delta^2}\bigr)_+}\right), \tag{29}$$

$$\xi_2(t) \triangleq \frac{\theta^n}{\sigma^n}e^{-\frac{1}{2}\left(\frac{t^2}{\sigma^2\theta^2}-\frac{n\Upsilon}{\delta}\right)}\left(Q_{\frac{n}{2}}\left(0,\sqrt{\bigl(\tfrac{t^2}{\theta^2}-n\Upsilon\tfrac{\sigma^2}{\delta^2}\bigr)_+}\right) - Q_{\frac{n}{2}}\left(\sqrt{n\Upsilon}\frac{\theta}{\delta},\frac{t}{\theta}\right)\right), \tag{30}$$

$$\xi_3(t) = \frac{n\Upsilon}{2\delta}\left(\frac{t\delta}{\sigma^2\sqrt{n\Upsilon}}\right)^{\frac{n}{2}}e^{-\frac{1}{2}\left(\frac{t^2}{\sigma^2}+n\Upsilon\frac{\sigma^2}{\delta^2}\right)}I_{\frac{n}{2}}\Bigl(\sqrt{n\Upsilon}\frac{t}{\delta}\Bigr), \tag{31}$$

where $(a)_+ = \max(0,a)$, $Q_m(a,b)$ is the Marcum Q-function and $I_m(\cdot)$ denotes the m-th order modified Bessel function of the first kind. Let t_0 be the solution to the implicit equation

$$\xi_1(t_0)+\xi_2(t_0)+\xi_3(t_0) = 0, \tag{32}$$

and let

$$\bar{M} \triangleq \bigl(1-Q_{\frac{n}{2}}\bigl(\sqrt{n\Upsilon}\theta/\delta,\ t_0/\theta\bigr)\bigr)^{-1}. \tag{33}$$

Then, for any code $\mathcal{C}\in\mathcal{L}_{\mathrm{a}}(n,M,\Upsilon)$ with cardinality $M\le\bar{M}$,

$$P_{\mathrm{e}}(\mathcal{C}) \ge \alpha_{\frac{1}{M}}\bigl(\varphi^n_{\sqrt{\Upsilon},\sigma},\varphi^n_{0,\theta}\bigr). \tag{34}$$

Proof: See the Appendix. ■

Corollary 2 implies that the bound from Theorems 1 and 2 holds in the average power constraint setting if the cardinality of the codebook is sufficiently small. Indeed, it follows that this condition is satisfied for typical communication systems. For transmission rates very close to capacity or above capacity, the bound (21) is needed instead (see the example in Fig. 2).

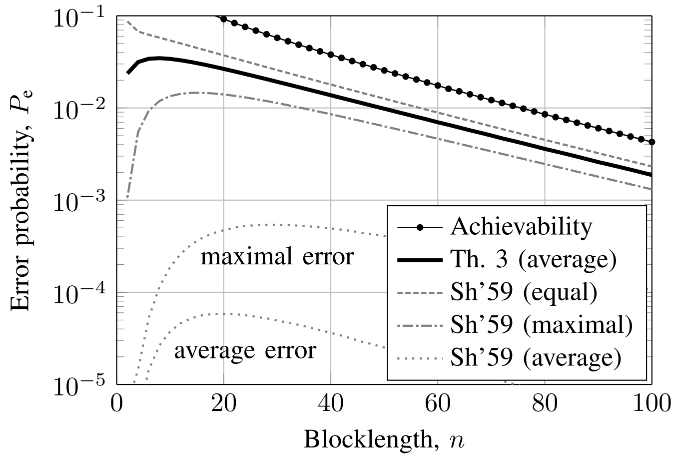

Fig. 1: Upper and lower bounds to the channel coding error probability over an AWGN channel with SNR $= 10$ dB and rate $R = 1.5$ bits/channel use.

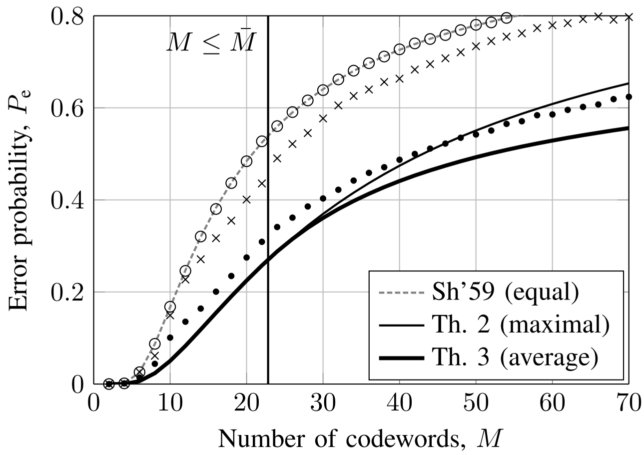

Fig. 2: Lower bounds to the channel coding error probability over an AWGN channel with $n = 2$ and SNR $= 10$ dB. Markers show the simulated error probability of a sequence of codes satisfying an equal ($\circ$), maximal ($\times$) and average ($\bullet$) power constraints. Vertical line corresponds to the boundary $M \leq \bar{M} \approx 22.8$ from Corollary 2.

V. Numerical Examples

A. Comparison with previous results

We consider the transmission of $M = 2^{nR}$ codewords over n uses of an AWGN channel with $R = 1.5$ bits/channel use and SNR $= 10 \log_{10} \frac{\Upsilon}{\sigma^2} = 10$ dB. The channel capacity is $C = \frac{1}{2} \log_2\bigl(1 + \frac{\Upsilon}{\sigma^2}\bigr) \approx 1.8$ bits/channel use.

Figure 1 compares the lower bound from Theorem 3 with previous results in the literature. In particular, we consider Shannon'59 achievability and converse bounds for equal power constraint [1, Eq. (20)], Shannon'59 converse bound for maximal power constraint [1, Eqs. (20) and (83)], and the lower bound for average power constraint that follows from combining [1, Eq. (20)] and [6, Lem. 65]. While the bounds in Figure 1 hold under the average probability of error formalism, for reference we also include the curve *Sh'59 (average)* for maximal error probability, which is tighter than that for average error probability (see [6, Lem. 65] for details).

As the transmission rate R is close to capacity C, the optimizing θ^2 in Theorem 3 is close to the variance of the capacity achieving output distribution. Then, for simplicity, we fix $\theta^2 = \Upsilon + \sigma^2$. For the system parameters considered, the condition $M \leq \bar{M}$ from Corollary 2 is satisfied for all n and Theorem 3 can be evaluated using (34). It thus follows that the bounds from Theorems 1, 2 and 3 coincide.

The results in Figure 1 show that that Shannon'59 lower bound is the tightest bound in the equal power constraint setting. However, under both maximal and average power constraints, Theorem 3 yields a tighter lower bound and presents a small constant gap to the achievability bound from [1, Eq. (20)].[1] Indeed, for an average power constraint and under the average probability of error formalism the advantage of Theorem 3 over previous results is significant in the finite blocklength regime, as shown in Figure 1.

[1]The rate considered here is above the critical rate of the channel, and therefore the error exponents of the achievability and converse bounds in Figure 1 coincide. This is not longer true for rates below the critical rate.

B. Constellation design under power constraints

We consider the problem of transmitting M codewords with $n = 2$ uses of an AWGN channel with SNR $= 10$ dB. This problem is analogous to determining the best 2-dimensional constellation for an uncoded communication system.

Figure 2 depicts Shannon'59 lower bound [1, Eq. (20)], and the bounds from Theorems 2 and 3, both with $\theta^2 = \Upsilon + \sigma^2$. The vertical line shows the boundary of the region $M \leq \bar{M}$ from Corollary 2, where the bounds from Theorems 2 and 3 coincide. With markers, we show the simulated ML decoding error probability of a sequence of M-PSK (phase-shift keying) constellations satisfying an equal power constraint ($\circ$), of a sequence of M-APSK (amplitude-phase-shift keying) constellations satisfying a maximal ($\times$) and average ($\bullet$) power constraints (both optimized using an stochastic algorithm).

As 2-dimensional cones coincide with the ML decoding regions of an M-PSK constellation, Shannon'59 curve is on top of the corresponding simulated probability ($\circ$). However, Shannon'59 lower bound does not apply to M-APSK constellations satisfying maximal ($\times$) and average ($\bullet$) power constraints. We can see that while Theorem 3 applies in both of these settings, this is not the case for Theorem 2, that in general only applies under maximal power constraint. As stated in Corollary 2, the bounds from Theorems 2 and 3 coincide for $M \leq \bar{M} \approx 22.8$.

An analysis of the average power constrained codes ($\bullet$) that violate Theorem 2 shows that they present several constellation points concentrated at the origin $(0, 0)$. As these symbols coincide, it is not possible to distinguish between them and they will often yield a decoding error. However, since the symbol $(0, 0)$ does not require any energy for its transmission, the average power for the remaining symbols is increased and this code yields an overall smaller probability of error.

ACKNOWLEDGMENT

The author gratefully acknowledges the insightful comments by the reviewers and fruitful discussions with Tobias Koch and David Morales-Jimenez.

APPENDIX

We characterize the region where $f(\beta,\gamma)$ and its convex envelope $\underline{f}(\beta,\gamma)$ coincide. We shall use the following result.

Proposition 2: Suppose g is differentiable with gradient ∇g. Let $\mathcal{A}$ denote the domain of g, and let $a \in \mathcal{A}$. If the inequality

$$g(\bar{a}) \geq g(a) + \nabla g(a)^T(\bar{a}-a), \tag{35}$$

is satisfied for all $\bar{a} \in \mathcal{A}$, then, it holds that $g(a) = g^{**}(a)$.

Proof: As g^{**} is the lower convex envelope of g, then $g(a) \geq g^{**}(a)$ trivially. It remains to show that (35) implies $g(a) \leq g^{**}(a)$. Fenchel's inequality [7, Sec. 3.3.2] yields

$$g^{**}(a) \geq \langle a, b\rangle - g^*(b), \tag{36}$$

for any b in the domain of g^*.

Setting $b = \nabla g(a)$ and using (19) in (36), we obtain

$$g^{**}(a) \geq \nabla g(a)^T a - \max_{\bar{a}\in\mathcal{A}}\bigl\{\nabla g(a)^T\bar{a} - g(\bar{a})\bigr\} \tag{37}$$

$$= \min_{\bar{a}\in\mathcal{A}}\bigl\{\nabla g(a)^T(a-\bar{a}) + g(\bar{a})\bigr\} \tag{38}$$

$$\geq \min_{\bar{a}\in\mathcal{A}}\bigl\{g(a)\bigr\}, \tag{39}$$

where in the last step we used (35) to lower bound $g(\bar{a})$. Since the objective of (39) does not depend on $\bar{a}$, we conclude from (37)-(39) that $g(a) \leq g^{**}(a)$ and the result follows. ■

We apply Proposition 2 to the function $f(\beta,\gamma)$. We recall that $f(\beta,\gamma)$ is differentiable for $\beta \in [0,1]$ and $\gamma \geq 0$ with derivatives given in [8, App. A]. We define the gradients

$$\nabla_\beta f(b,g) \triangleq \frac{\partial f(\beta,\gamma)}{\partial\beta}\Big|_{\beta=b,\gamma=g}, \tag{40}$$

$$\nabla_\gamma f(b,g) \triangleq \frac{\partial f(\beta,\gamma)}{\partial\gamma}\Big|_{\beta=b,\gamma=g}. \tag{41}$$

According to Proposition 2, the function $f(\beta_0,\gamma_0)$ and its convex envelope $\underline{f}(\beta_0,\gamma_0)$ coincide if

$$f(\beta,\gamma) \ \geq\ f(\beta_0,\gamma_0) + (\beta-\beta_0)\nabla_\beta f(\beta_0,\gamma_0) + (\gamma-\gamma_0)\nabla_\gamma f(\beta_0,\gamma_0). \tag{42}$$

is satisfied for all $\beta \in [0,1]$ and $\gamma \geq 0$. This condition implies that the first-order Taylor approximation of f at (β_0,γ_0) is a global under-estimator of the function $f(\beta,\gamma)$.

The derivatives of $f(\beta,\gamma)$, given in [8, App. A], show that the function is decreasing in both parameters, convex with respect to β for all $\beta \in [0,1]$, and jointly convex with respect to (β,γ) except for the neighborhood near the axis $\gamma = 0$. Using these properties, it can be shown that the the condition (42) only needs to be verified along the axis $\gamma = 0$.

Then, we conclude that $f(\beta_0,\gamma_0) = \underline{f}(\beta_0,\gamma_0)$ if (42) holds for every $\beta \in [0,1]$ and $\gamma = 0$, i.e., if

$$f(\beta_0,\gamma_0) - f(\beta,0) \geq (\beta_0-\beta)\nabla_\beta f(\beta_0,\gamma_0) + \gamma_0\nabla_\gamma f(\beta_0,\gamma_0). \tag{43}$$

Let $\theta \geq \sigma > 0$, $n \geq 1$. Let t_0 be the value such that $\beta(\gamma_0,t_0) = \beta_0$ and let $\bar{t}$ satisfy $\beta(0,\bar{t}) = \beta$, for $\beta(\gamma,t)$ defined in (13). Using (12) and the derivatives in [8, App. A], yields

$$f(\beta_0,\gamma_0) - f(\beta,0) = Q_{\frac{n}{2}}\left(\sqrt{n\gamma_0}\frac{\sigma}{\delta}, \frac{t_0}{\sigma}\right) - Q_{\frac{n}{2}}\left(0, \frac{\bar{t}}{\sigma}\right), \tag{44}$$

$$\nabla_\beta f(\beta_0,\gamma_0) = -\frac{\theta^n}{\sigma^n}e^{\frac{1}{2}\left(\frac{n\gamma_0}{\delta} - t_0^2\left(\frac{1}{\sigma^2}-\frac{1}{\theta^2}\right)\right)}, \tag{45}$$

$$\nabla_\gamma f(\beta_0,\gamma_0) = -\frac{n}{2\delta}\left(\frac{t_0\delta}{\sigma^2\sqrt{n\gamma_0}}\right)^{\frac{n}{2}} I_{\frac{n}{2}}\left(\frac{t_0\sqrt{n\gamma_0}}{\delta}\right) \times e^{-\frac{1}{2}\left(\frac{n\gamma_0\sigma^2}{\delta^2} + \frac{t_0^2}{\sigma^2}\right)}. \tag{46}$$

As $\beta(\gamma_0,t_0) = \beta_0$ and $\beta(0,\bar{t}) = \beta$, using (13), it follows that

$$\beta_0 - \beta = Q_{\frac{n}{2}}\left(0, \frac{\bar{t}}{\theta}\right) - Q_{\frac{n}{2}}\left(\sqrt{n\gamma_0}\frac{\theta}{\delta}, \frac{t_0}{\theta}\right). \tag{47}$$

Substituting (44) and (47) in (43), reorganizing terms, yields

$$Q_{\frac{n}{2}}\left(\sqrt{n\gamma_0}\frac{\sigma}{\delta}, \frac{t_0}{\sigma}\right) + \nabla_\beta f(\beta_0,\gamma_0) Q_{\frac{n}{2}}\left(\sqrt{n\gamma_0}\frac{\theta}{\delta}, \frac{t_0}{\theta}\right) - \gamma_0\nabla_\gamma f(\beta_0,\gamma_0) \ \geq\ Q_{\frac{n}{2}}\left(0, \frac{\bar{t}}{\sigma}\right) + \nabla_\beta f(\beta_0,\gamma_0) Q_{\frac{n}{2}}\left(0, \frac{\bar{t}}{\theta}\right). \tag{48}$$

The interval $\beta \in [0,1]$ corresponds to $\bar{t} \geq 0$. We maximize the right-hand side of (48) over $\bar{t} \geq 0$ and we only verify the condition (48) for this maximum value. To this end, we find the derivative of the right-hand side of (48) with respect to $\bar{t}$, we identify the resulting expression with zero, and we use (45). We conclude that the right-hand side of (48) is maximized for

$$\bar{t}_\star = \sqrt{\left(t_0^2 - n\gamma\sigma^2\theta^2/\delta^2\right)_+} \tag{49}$$

where the threshold $(a)_+ = \max(0,a)$ follows from the constraint $\bar{t} \geq 0$. By evaluating the second derivative of (48), it can be verified that $\bar{t}_\star$ in (49) is indeed a maximum.

Using (45), (46) and (49) in (48) we obtain the desired characterization for the region of interest. For the statement of the result in Corollary 2, we select the smallest t_0 that fulfills (48) (which satisfies the condition with equality) and invert the transformation $\beta(\gamma_0,t_0) = \beta_0$ for $\gamma_0 = \Upsilon$ and $\beta_0 = \frac{1}{M}$.

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] G. Vazquez-Vilar, "On the error probability of optimal codes in Gaussian channels under maximal power constraint," in *2019 IEEE Int. Symp. on Inf. Theory*, Paris, France, July 2019.

[4] G. Vazquez-Vilar, A. Guillén i Fàbregas, T. Koch, and A. Lancho, "Saddlepoint approximation of the error probability of binary hypothesis testing," in *2018 IEEE Int. Symp. on Inf. Theory*, June 2018, pp. 2306–2310.

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

[6] ——, "Channel coding: non-asymptotic fundamental limits," Ph.D. dissertation, Princeton University, Sep. 2010.

[7] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, US: Cambridge University Press, 2004.

[8] G. Vazquez-Vilar, "Error probability bounds for Gaussian channels under maximal and average power constraints," *preprint arXiv:1907.03163*, 2019.