Optimal Rate Profile for Random Sphere Codes in the Gaussian Channel

Josep Font-Segura[1], Alfonso Martinez[1], Mehdi Dabirnia[2], and Albert Guillén i Fàbregas[3,4]
[1]Universitat Pompeu Fabra, [2]CTTC, [3]University of Cambridge, [4]Universitat Politècnica de Catalunya
josep.font@ieee.org, alfonso.martinez@ieee.org, mehdi.dabirnia@cttc.cat, guillen@ieee.org

***Abstract*—This paper studies the optimal rate profile for codeword placement in the interior of a hypersphere of squared radius P for power-constrained random sphere codes in the Gaussian channel. While an inner shell may worsen the exponent of random i.i.d. codes, in accordance with the optimality of random shell codes, the rate allocated to inner shells need not be zero in order to attain the optimal exponent. The analysis is built on an intermediate characterization of the error exponent as the worst exponent over shells of squared radius αP, for $0 \leq \alpha \leq 1$.**

I. Introduction

Optimal coding over the power-constrained Gaussian channel involves codewords satisfying the power constraint with equality, e.g. Shannon's shell codes [1]. Indeed, Shannon's shell codes attain a larger error exponent [2, Th. 7.4.4] than cost-constrained i.i.d. codes [2, Th. 5.6.2]. As the error probability of a random coding ensemble is dominated by poor codes within the ensemble, the addition of constraints, e.g. restricting codewords to a shell, imposing cost constraints, or requiring constant composition, removes such poor codes and may lead to higher error exponents than in the absence of these constraints.

In this paper, we consider the Gaussian channel and study whether an exponential number of messages may be placed inside the hypersphere of squared radius P, the power constraint, without a loss of error exponent. We refer to these as sphere codes, as their codewords are inside a hypersphere. We first express the error exponent as the worst exponent over shells of squared radius αP, for $0 \leq \alpha \leq 1$. We then determine the optimal rate profile for codeword placement inside the hypersphere, so as to equalize the exponent across shells. Inner shells are allocated just enough rate that their exponent matches that of the outer shell of squared radius P, allowing nonzero rate in the interior without sacrificing the overall exponent. We begin by establishing notation and reviewing known results on sphere and shell random codes.

A. Channel Model, Rates, and Exponents

Given a rate R and a blocklength n, a source generates $M_n = \lceil e^{nR} \rceil$ equiprobable messages. A codebook $\mathcal{C}_n = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_{M_n}\}$ is a collection of M_n codewords, each of n symbols drawn from a discrete or continuous alphabet $\mathcal{X} \subseteq \mathbb{C}$.

For the complex-valued additive white Gaussian noise (AWGN) channel with variance σ^2, the channel output $\boldsymbol{y} \in \mathbb{C}^n$ is related to the channel input $\boldsymbol{x} \in \mathcal{X}^n$ by the conditional probability $W^n(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^n W(y_i|x_i)$, where

$$W(y|x) = \frac{1}{\pi\sigma^2} e^{-\frac{|y-x|^2}{\sigma^2}}. \tag{1}$$

Let $Q(x)$ be a probability distribution over the input alphabet $\mathcal{X}$ satisfying the average power constraint

$$\sum_{x \in \mathcal{X}} Q(x)|x|^2 = P. \tag{2}$$

One example is the zero-mean complex Gaussian distribution with variance P. Throughout the paper, the summation sign is interpreted as an integral when required.

Given $\boldsymbol{y}$, a decoder forms the maximum-likelihood estimate $\hat{m} = \arg\max_{m \in \{1,\dots,M_n\}} W^n(\boldsymbol{y}|\boldsymbol{x}_m)$. When $\boldsymbol{x}_m$ is sent, a decoding error occurs if the decoder outputs $\hat{m} \neq m$. We denote the average error probability by $\epsilon(\mathcal{C}_n)$. The rate R is achievable if there exists a sequence of codes $\mathcal{C}_n$ with at least e^{nR} messages and vanishing error probability, i.e. $\lim_{n\to\infty} \epsilon(\mathcal{C}_n) = 0$. Similarly, the exponent $E(R)$ is achievable if there exists a sequence of codes $\mathcal{C}_n$ with at least e^{nR} messages and an exponential rate of decay of the error probability that satisfies $\liminf_{n\to\infty} -\frac{1}{n}\log\epsilon(\mathcal{C}_n) \geq E(R)$.

B. Sphere and Shell Codes

We now define some quantities, sets, and ensembles that are instrumental to our analysis. Given an alphabet $\mathcal{X}$ and some non-negative numbers λ, δ, we denote by $\mathcal{S}^n(\lambda,\delta)$ the set of codewords in the n-shell of squared radius $n\lambda$ and width $n\delta$:

$$\mathcal{S}^n(\lambda,\delta) = \big\{\boldsymbol{x} \in \mathcal{X}^n \ : \ n(\lambda-\delta) \leq \|\boldsymbol{x}\|^2 \leq n\lambda\big\}. \tag{3}$$

With some abuse of notation, we define $\mathcal{S}^n(\lambda) = \mathcal{S}^n(\lambda,\lambda)$. Setting $\lambda = P$, $\mathcal{S}^n(P)$ gives the set of codewords with power at most P in the n-sphere of squared radius nP. The spherical random coding ensemble is the set of codes of M_n codewords drawn pairwise-independently from the distribution $Q^n_{\text{iid}}(\boldsymbol{x})$ under a maximum power constraint,

$$Q^n_{\text{iid}}(\boldsymbol{x}) = \frac{1}{\nu_n}\prod_{i=1}^n Q(x_i)\mathbf{1}\big\{\boldsymbol{x} \in \mathcal{S}^n(P)\big\}, \tag{4}$$

where ν_n is a normalization factor that quantifies the probability that a codeword drawn with independent symbols from $\prod_{i=1}^n Q(x_i)$ falls in $\mathcal{S}^n(P)$.

Similarly, setting $\lambda = P$ and $\delta = \delta_n = \delta_0/n$ gives the n-shell of (absolute) width δ_0 and squared radius nP. The shell

This work has been funded in part by the Spanish Ministry of Science, Innovation, and Universities under grant PID2024-159557OB-C22/MICIU/AEI/10.13039/501100011033/FEDER, UE and by the European Research Council under grant 101142747.

random coding ensemble is the set of codes of M_n codewords drawn pairwise-independently from the distribution $Q^n_{\text{shell}}(\boldsymbol{x})$,

$$Q^n_{\text{shell}}(\boldsymbol{x}) = \frac{1}{\mu_n}\prod_{i=1}^{n} Q(x_i)\mathbf{1}\{\boldsymbol{x}\in\mathcal{S}^n(P,\delta_n)\}, \tag{5}$$

where μ_n is the probability that a codeword drawn with independent symbols from $\prod_{i=1}^n Q(x_i)$ falls[1] in $\mathcal{S}^n(P,\delta_n)$. The shell random coding ensemble is equivalent to Gallager's constrained-input ensemble [2, Sec. 7.3]. When $Q(x)$ is the Gaussian distribution and $\delta = 0$, Shannon's spherical ensemble [1] draws codewords uniformly on the n-surface $\|\boldsymbol{x}\|^2 = nP$.

For both sphere and shell ensembles, codewords will most likely fall in the outer shell. This fact proves sufficient for both ensembles to achieve Shannon's capacity, but not for performance metrics such as error exponents.

The error probability of the optimal code with codewords in the sphere $\mathcal{S}^n(P)$ is lower bounded by the error probability of the optimal code with codewords in the shell $\mathcal{S}^n(P,\delta_n)$, see [1, Sec. XIII]. The error exponent $E_{\text{shell}}(R)$ of shell codes is given by the error exponent of cost-constrained codes with cost function $|x|^2 - P$, that is

$$E_{\text{shell}}(R) = E_{\text{cc}}(R), \tag{6}$$

where $E_{\text{cc}}(R)$ is given in [2, Th. 7.3.2] by

$$E_{\text{cc}}(R) = \max_{\rho\in[0,1]}\max_{r\in\mathbb{R}}\{E_1(\rho,r,P) - \rho R\}, \tag{7}$$

and $E_1(\rho,r,\beta)$ is the Gallager function

$$E_1(\rho,r,\beta) = -\log\sum_y\left(\sum_x Q(x)e^{r(|x|^2-\beta)}W(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}. \tag{8}$$

The spherical random coding ensemble with input distribution in (4) yields a lower bound to the error exponent $E_{\text{sphere}}(R)$ of sphere codes, that is

$$E_{\text{sphere}}(R) \geq E_{\text{iid}}(R), \tag{9}$$

where $E_{\text{iid}}(R)$ is given in [2, Th. 5.6.2] by

$$E_{\text{iid}}(R) = \max_{0\leq\rho\leq1}\{E_0(\rho) - \rho R\}, \tag{10}$$

and $E_0(\rho)$ is the Gallager function

$$E_0(\rho) = -\log\sum_y\left(\sum_x Q(x)W(y|x)^{\frac{1}{1+\rho}}\right)^{1+\rho}. \tag{11}$$

In this paper, we provide a multishell sphere ensemble whose error exponent coincides with $E_{\text{shell}}(R)$, even though the full hypersphere is used. This proves that $E_{\text{sphere}}(R) = E_{\text{shell}}(R)$, and that the suboptimality of i.i.d. codes stems from the rate profile induced by the random generation of codewords, rather than from the use of the interior of the hypersphere. A related conclusion for rates above the critical rate is implicit in the analysis of modulo-lattice codes in [4].

[1]Neither ν_n nor μ_n is an exponential term, since the central limit theorem [3, Sec. XV.5] implies that codewords drawn from a distribution satisfying (2) will fall in $\mathcal{S}^n(P)$ with probability $O(1)$ and in $\mathcal{S}^n(P,\delta_n)$ with probability $O\big(\frac{1}{\sqrt{n}}\big)$ as $n\to\infty$.

II. Multishell Random Coding

A. Ensemble Construction

In this section, we derive an achievable error exponent for codes in the sphere $\mathcal{S}^n(P)$ by partitioning it into shells and analyzing the contribution of each shell to the overall error probability. To this extent, we partition the sphere $\mathcal{S}^n(P)$ into L_n+1 disjoint shells $\mathcal{R}^n_\ell(P,L_n)$, indexed by $\ell\in\{0,1,\ldots,L_n\}$, and given by

$$\mathcal{R}^n_\ell(P,L_n) = \begin{cases}\mathcal{S}^n(0) & \ell = 0,\\ \mathcal{S}^n\Big(\frac{\ell P}{L_n},\frac{P}{L_n}\Big)\setminus\mathcal{S}^n\Big(\frac{(\ell-1)P}{L_n}\Big) & \ell > 0.\end{cases} \tag{12}$$

The message set $\mathcal{M} = \{1,\ldots,M_n\}$ is partitioned into L_n+1 disjoint sets $\mathcal{M}_\ell$ such that $\mathcal{M}_0\cup\cdots\cup\mathcal{M}_{L_n} = \mathcal{M}$, where $|\mathcal{M}_\ell| = M_{n,\ell}$. The total number of messages satisfies $\sum_{\ell=0}^{L_n} M_{n,\ell} = M_n$. To each shell ℓ, we assign $M_{n,\ell} = \lceil e^{nR_\ell}\rceil$ messages, where $R_\ell\leq R$. The rate profile $\{R_\ell\}_{\ell=0}^{L_n}$ is a design parameter to be optimized. A particular profile arises when codewords are drawn from the i.i.d. ensemble Q^n_{iid} and assigned to shells according to where they fall; we analyze this induced profile in Sec. III-A.

Codewords for messages assigned to shell ℓ are drawn pairwise-independently from the distribution

$$Q^n_\ell(\boldsymbol{x}) = \frac{1}{\mu_{n,\ell}}\prod_{i=1}^n Q(x_i)\mathbf{1}\{\boldsymbol{x}\in\mathcal{R}^n_\ell(P,L_n)\}, \tag{13}$$

where $Q(x)$ satisfies (2) and $\mu_{n,\ell}$ is the probability that a codeword drawn with independent symbols from $\prod_{i=1}^n Q(x_i)$ falls in $\mathcal{R}^n_\ell(P,L_n)$, i.e.,

$$\mu_{n,\ell} = \sum_{\boldsymbol{x}}\prod_{i=1}^n Q(x_i)\mathbf{1}\Big\{\tfrac{n(\ell-1)P}{L_n} < \|\boldsymbol{x}\|^2 \leq \tfrac{n\ell P}{L_n}\Big\}. \tag{14}$$

B. Error Probability Bound

Let $\overline{\epsilon_\circ}$ be the average error probability over the ensemble of codes described in Sec. II-A. Using random coding arguments, there must exist at least one code in the ensemble whose error probability is at most $\overline{\epsilon_\circ}$. In this section, we derive an upper bound on the ensemble-average error probability in three steps: first applying the random coding union (RCU) bound, then using Markov's inequality and Hölder's inequality to decouple inter-shell error events, and finally extracting the exponent.

The adaptation of the RCU bound [5, Th. 16] to our multishell construction is analogous to the multi-class scenario in [6]. Under maximum-likelihood decoding, the ensemble-average error probability $\overline{\epsilon_\circ}$ is upper bounded by [6, Eq. (25)]

$$\overline{\epsilon_\circ} \leq \frac{1}{M_n}\cdot \sum_{\ell=0}^{L_n} M_{n,\ell}\sum_{\boldsymbol{x},\boldsymbol{y}} Q^n_\ell(\boldsymbol{x})W^n(\boldsymbol{y}|\boldsymbol{x})\min\Big\{1,\sum_{\bar{\ell}=0}^{L_n} M_{n,\bar{\ell}}\,\mathrm{pep}_{\bar{\ell}}(\boldsymbol{x},\boldsymbol{y})\Big\}, \tag{15}$$

where $\mathrm{pep}_{\bar{\ell}}(\boldsymbol{x},\boldsymbol{y})$ is the pairwise-error probability

$$\mathrm{pep}_{\bar{\ell}}(\boldsymbol{x},\boldsymbol{y}) = \sum_{\bar{\boldsymbol{x}}} Q^n_{\bar{\ell}}(\bar{\boldsymbol{x}}) \mathbf{1}\big\{W^n(\boldsymbol{y}|\bar{\boldsymbol{x}}) \geq W^n(\boldsymbol{y}|\boldsymbol{x})\big\}. \quad (16)$$

We use Markov's inequality in (16) with parameter $s_{\bar{\ell}} \geq 0$, similarly to [2, Eq. (5.6.8)], to obtain

$$\mathrm{pep}_{\bar{\ell}}(\boldsymbol{x},\boldsymbol{y}) \leq \sum_{\bar{\boldsymbol{x}}} Q^n_{\bar{\ell}}(\bar{\boldsymbol{x}}) \bigg(\frac{W^n(\boldsymbol{y}|\bar{\boldsymbol{x}})}{W^n(\boldsymbol{y}|\boldsymbol{x})}\bigg)^{s_{\bar{\ell}}}. \quad (17)$$

Using (17) and the inequality $\min\{1, A+B\} \leq A^\rho + B^{\rho'}$ for $A, B \geq 0$ and $\rho, \rho' \in [0,1]$ in (15), we have

$$\overline{\epsilon_\circ} \leq \frac{1}{M_n} \sum_{\ell,\bar{\ell}=0}^{L_n} M_{n,\ell} \sum_{\boldsymbol{x},\boldsymbol{y}} Q^n_\ell(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) + \Bigg(M_{n,\bar{\ell}} \sum_{\bar{\boldsymbol{x}}} Q^n_{\bar{\ell}}(\bar{\boldsymbol{x}}) \bigg(\frac{W^n(\boldsymbol{y}|\bar{\boldsymbol{x}})}{W^n(\boldsymbol{y}|\boldsymbol{x})}\bigg)^{s_{\bar{\ell}}}\Bigg)^{\rho_{\ell\bar{\ell}}}, \quad (18)$$

where $\rho_{\ell\bar{\ell}} \in [0,1]$ and $s_{\bar{\ell}} \geq 0$ for $\ell, \bar{\ell} \in \{1, \dots, L_n\}$. We define the function $G_\ell(\boldsymbol{y})$ as

$$G_\ell(\boldsymbol{y}) = \bigg(M_{n,\ell} \sum_{\boldsymbol{x}} Q^n_\ell(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x})^{\frac{1}{1+\rho_\ell}}\bigg)^{1+\rho_\ell}, \quad (19)$$

where $\rho_\ell \in [0,1]$. Setting $\rho_{\ell\bar{\ell}} = \frac{\rho_\ell(1+\rho_{\bar{\ell}})}{1+\rho_\ell} \in [0,1]$ and $s_\ell = \frac{1}{1+\rho_\ell} \in \big[\frac{1}{2}, 1\big]$, we can write $\rho_{\ell\bar{\ell}} = \frac{1-s_\ell}{s_{\bar{\ell}}}$, and (18) becomes

$$\overline{\epsilon_\circ} \leq \frac{1}{M_n} \sum_{\ell=0}^{L_n} \sum_{\bar{\ell}=0}^{L_n} \sum_{\boldsymbol{y}} G_\ell(\boldsymbol{y})^{\frac{1}{1+\rho_\ell}} G_{\bar{\ell}}(\boldsymbol{y})^{\frac{\rho_\ell}{1+\rho_\ell}}. \quad (20)$$

This decomposition expresses the probability of the *inter-shell* error event between shells ℓ and $\bar{\ell}$ as the product of two terms corresponding to *intra-shell* error events. The right-hand side of (20) is further upper bounded using Hölder's inequality in its arithmetic-geometric mean form, $\|fg\|_1 \leq \|f\|_p \|g\|_q \leq \frac{1}{p}\|f\|_p^p + \frac{1}{q}\|g\|_q^q$, with $\frac{1}{p} = \frac{1}{1+\rho_\ell}$ and $\frac{1}{q} = \frac{\rho_\ell}{1+\rho_\ell}$, yielding

$$\overline{\epsilon_\circ} \leq \frac{1}{M_n} \sum_{\ell=0}^{L_n} \sum_{\bar{\ell}=0}^{L_n} \bigg(\frac{1}{1+\rho_\ell} \sum_{\boldsymbol{y}} G_\ell(\boldsymbol{y}) + \frac{\rho_\ell}{1+\rho_\ell} \sum_{\boldsymbol{y}} G_{\bar{\ell}}(\boldsymbol{y})\bigg). \quad (21)$$

Since $\frac{1}{1+\rho_\ell} \leq 1$ and $\frac{\rho_\ell}{1+\rho_\ell} \leq \frac{1}{2}$, we further bound (21) by

$$\overline{\epsilon_\circ} \leq \frac{1}{M_n} \sum_{\ell=0}^{L_n} \sum_{\boldsymbol{y}} G_\ell(\boldsymbol{y}) + \sum_{\ell=0}^{L_n} \sum_{\bar{\ell}=0, \bar{\ell}\neq\ell}^{L_n} \bigg(\sum_{\boldsymbol{y}} G_\ell(\boldsymbol{y}) + \frac{1}{2} \sum_{\boldsymbol{y}} G_{\bar{\ell}}(\boldsymbol{y})\bigg). \quad (22)$$

Counting the number of appearances of $G_\ell(\boldsymbol{y})$ in (22), we obtain

$$\overline{\epsilon_\circ} \leq \frac{3(L_n+1)-1}{2M_n} \sum_{\ell=0}^{L_n} \sum_{\boldsymbol{y}} G_\ell(\boldsymbol{y}) \quad (23)$$

$$\leq \frac{3(L_n+1)-1}{2} \sum_{\ell=0}^{L_n} \frac{M_{n,\ell}}{M_n} \overline{\epsilon_\ell}, \quad (24)$$

where, substituting $G_\ell(\boldsymbol{y})$ from (19) in (23), we defined

$$\overline{\epsilon_\ell} = M_{n,\ell}^{\rho_\ell} \sum_{\boldsymbol{y}} \bigg(\sum_{\boldsymbol{x}} Q^n_\ell(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x})^{\frac{1}{1+\rho_\ell}}\bigg)^{1+\rho_\ell} \quad (25)$$

as the contribution of the ℓ-th shell to the error probability $\overline{\epsilon_\circ}$. To obtain an achievable error exponent from (24), we study the exponential behavior of each term $\overline{\epsilon_\ell}$ as $n \to \infty$.

C. Per-Shell Error Exponent

We identify (25) as the upper bound to the random-coding error probability given in [2, Eq. (5.6.1)] for transmission over the channel $W^n(\boldsymbol{y}|\boldsymbol{x})$, with input distribution $Q^n_\ell(\boldsymbol{x})$ and $M_{n,\ell}$ codewords. Following Gallager's method, we bound the input distribution (13) in (25). Similarly to [2, Eq. (7.3.17)], we use the inequality $\mathbf{1}\{A \leq B\} \leq e^{r(B-A)}$ for $r \geq 0$ to bound the indicator function for both boundaries of $\mathcal{R}^n_\ell(P, L_n)$ in (12). Setting $L_n = n$, the bound for the upper (resp. lower) boundary $\|\boldsymbol{x}\|^2 \leq \ell P$ (resp. $\|\boldsymbol{x}\|^2 > (\ell-1)P$) with parameter $r \geq 0$ (resp. $\bar{r} \geq 0$) is given by

$$\mathbf{1}\big\{\boldsymbol{x} \in \mathcal{R}^n_\ell(P, n)\big\} \leq e^{r(\ell P - \|\boldsymbol{x}\|^2) + \bar{r}(\|\boldsymbol{x}\|^2 - (\ell-1)P)} = e^{r_\ell(\|\boldsymbol{x}\|^2 - \ell P) + \bar{r}P}, \quad (26)$$

where $r_\ell = \bar{r} - r \in \mathbb{R}$. Substituting (26) in (13) and then in (25), we obtain

$$\overline{\epsilon_\ell} \leq \frac{\kappa_\ell M_{n,\ell}^{\rho_\ell}}{\mu_{n,\ell}^{1+\rho_\ell}} \cdot \sum_{\boldsymbol{y}} \bigg(\sum_{\boldsymbol{x}} \prod_{i=1}^n Q(x_i) e^{r_\ell\left(|x_i|^2 - \frac{\ell}{n}P\right)} W(y_i|x_i)^{\frac{1}{1+\rho_\ell}}\bigg)^{1+\rho_\ell}, \quad (27)$$

where we used that the channel is memoryless, and defined $\kappa_\ell = e^{\bar{r}(1+\rho_\ell)P}$. Interchanging the product and the summations, we obtain the single-letter bound

$$\overline{\epsilon_\ell} \leq \frac{\kappa_\ell M_{n,\ell}^{\rho_\ell}}{\mu_{n,\ell}^{1+\rho_\ell}} \cdot \Bigg(\sum_{y} \bigg(\sum_{x} Q(x) e^{r_\ell\left(|x_i|^2 - \frac{\ell}{n}P\right)} W(y|x)^{\frac{1}{1+\rho_\ell}}\bigg)^{1+\rho_\ell}\Bigg)^n. \quad (28)$$

This expression resembles [2, Eq. (7.3.19)], with both $M_{n,\ell}$ and $\mu_{n,\ell}$ affecting the exponential dependence on n. We define $D_\ell = \frac{1}{n}\log \mu_{n,\ell}$, which captures the large-deviation rate for codewords falling in shell ℓ. Then, optimizing (28) over $\rho_\ell \in [0,1]$ and $r_\ell \in \mathbb{R}$, we obtain

$$\overline{\epsilon_\ell} \leq \kappa_\ell \, e^{-nE_\ell(R_\ell)}, \quad (29)$$

where $E_\ell(R_\ell)$ is given by

$$E_\ell(R_\ell) = \max_{\rho\in[0,1]} \max_{r\in\mathbb{R}} \big\{E_1\big(\rho, r, \tfrac{\ell}{n}P\big) + (1+\rho)D_\ell - \rho R_\ell\big\}, \quad (30)$$

and $E_1(\rho, r, \beta)$ is given by (8). Substituting (29) in (24) and writing the number of messages $M_{n,\ell}$ and M_n in terms of the rates R_ℓ and R, we obtain

$$\overline{\epsilon_{\circledcirc}} \leq \frac{3(n+1)-1}{2} \sum_{\ell=0}^{n} \kappa_\ell \, e^{-n(E_\ell(R_\ell)+R-R_\ell)}. \tag{31}$$

D. Achievable Exponent

We now let $n \to \infty$ to obtain an achievable exponent for the multishell ensemble. The factors $\frac{3(n+1)-1}{2}$ and κ_ℓ in (31) are sub-exponential. The shell width P/n vanishes as $n \to \infty$, and the discrete values ℓ/n become dense in $[0,1]$.

For $\alpha \in [0,1]$, let $\ell_n(\alpha) = \lceil \alpha n \rceil$ denote the shell index corresponding to α at blocklength n. We define the limiting rate profile and normalization exponent as

$$R(\alpha) = \lim_{n\to\infty} R_{\ell_n(\alpha)}, \tag{32}$$

$$D(\alpha) = \lim_{n\to\infty} \frac{1}{n} \log \mu_{n,\ell_n(\alpha)}. \tag{33}$$

The quantity $D(\alpha)$ in (33) is the asymptotic exponent of the probability $\mu_{n,\ell}$ in (14). Since $\|\boldsymbol{x}\|^2$ is the sum of n i.i.d. terms, the saddlepoint approximation [7, Ch. 2] shows that

$$D(\alpha) = \min_{\tau\in\mathbb{R}} \big\{ \omega(\tau) - \tau\alpha P \big\}, \tag{34}$$

where $\omega(\tau)$ is the cumulant-generating function of $|X|^2$ under $Q(x)$:

$$\omega(\tau) = \log \sum_x Q(x) e^{\tau|x|^2}. \tag{35}$$

Since $Q(x)$ satisfies (2), we have $\omega'(0) = P$, and the minimum in (34) is achieved at $\tau = 0$ when $\alpha = 1$, giving $D(1) = 0$. For $\alpha < 1$, the minimum occurs at some $\tau^* > 0$, and $D(\alpha) < 0$. Overall, the quantity $D(\alpha)$ in (33) depends on the codeword distribution induced by $Q(x)$ across shells.

As the sum $\sum_{\ell=0}^{n} e^{nR_\ell}$ is dominated by its largest term as $n \to \infty$, the constraint $\sum_{\ell=0}^{n} M_{n,\ell} = M_n$ becomes

$$\max_{\alpha\in[0,1]} R(\alpha) = R. \tag{36}$$

Taking the negative normalized logarithm of both sides of (31) and using that the sum over ℓ is similarly dominated by its largest term, we obtain the achievable exponent

$$\liminf_{n\to\infty} -\frac{1}{n} \log \overline{\epsilon_{\circledcirc}} \geq \min_{\alpha\in[0,1]} \big\{ E_\alpha(R(\alpha)) + R - R(\alpha) \big\} \tag{37}$$

$$= E_{\circledcirc}(R), \tag{38}$$

where $E_\alpha(R)$ is the limiting form of (30), i.e.

$$E_\alpha(R) = \max_{\rho\in[0,1]} \max_{r\in\mathbb{R}} \big\{ E_1(\rho, r, \alpha P) + (1+\rho)D(\alpha) - \rho R \big\}. \tag{39}$$

The term $R - R(\alpha)$ in (37) reflects the relative contribution of shell α to the total rate: when $R(\alpha) < R$, this term is positive and improves the exponent, since fewer messages in a shell means errors from that shell are less costly. Conversely, the term $D(\alpha) < 0$ for $\alpha < 1$ penalizes inner shells, capturing the cost of placing codewords in an atypical power region. The optimal rate profile balances these competing effects.

III. Rate Profiles and Error Exponents

The achievable exponent (37) depends on the choice of rate profile $R(\alpha)$. In this section, we characterize two profiles of interest: the profile induced by the i.i.d. ensemble, and the optimal profile that maximizes the achievable exponent.

A. The i.i.d.-Induced Profile

We interpret the sphere codes generated with distribution (4) as the multishell random coding ensemble described in Sec. II-A. When M_n codewords are drawn from $Q^n_{\text{iid}}(\boldsymbol{x})$ and assigned to shells according to where they fall, the number of codewords in each shell $(M_{n,0}, \ldots, M_{n,n})$ follow a multinomial distribution with M_n trials and probabilities $p_\ell = \mu_{n,\ell}/\nu_n$. Taking the expectation in (29) and (31), and using that $\mathrm{E}[M_{n,\ell}^{1+\rho}] \doteq (M_n p_\ell)^{1+\rho}$ for shells with $M_n p_\ell \to \infty$, while shells with $M_n p_\ell \to 0$ are asymptotically empty and do not contribute, the shell occupancies may be replaced by their expected values $M_n p_\ell = M_n \mu_{n,\ell}/\nu_n$ exponentially, leading to the rate profile

$$R^{\text{iid}}(\alpha) = R + D(\alpha). \tag{40}$$

Substituting (40) into the achievable exponent (37) and (39), we obtain

$$E_{\circledcirc}(R) = \min_{\alpha\in[0,1]} \max_{\rho\in[0,1]} \max_{r\in\mathbb{R}} \big\{ E_1(\rho, r, \alpha P) - \rho R \big\}. \tag{41}$$

Since (41) is linear in α through the term $\alpha P r(1+\rho)$ in $E_1(\rho, r, \alpha P)$ and the minimum is over the compact set $[0,1]$, Fan's minimax theorem [8] allows us to write

$$\min_{\alpha\in[0,1]} \max_{\rho\in[0,1]} \max_{r\in\mathbb{R}} \big\{ E_1(\rho, r, \alpha P) - \rho R \big\}$$
$$= \max_{\rho\in[0,1]} \max_{r\in\mathbb{R}} \min_{\alpha\in[0,1]} \big\{ E_1(\rho, r, \alpha P) - \rho R \big\}. \tag{42}$$

The minimum over α on the right-hand side of (42) is achieved at $\alpha \to 0$ when $r > 0$ and at $\alpha = 1$ when $r < 0$. Therefore, the objective function after minimizing over α is continuous but not differentiable at $r = 0$, and achieves its maximum at $r = 0$. Since $E_1(\rho, 0, \alpha P) = E_0(\rho)$ for all α, we conclude

$$E_{\circledcirc}(R) = \max_{\rho\in[0,1]} \big\{ E_0(\rho) - \rho R \big\} = E_{\text{iid}}(R). \tag{43}$$

This illustrates that the i.i.d. ensemble Q^n_{iid} achieves the random-coding exponent $E_{\text{iid}}(R)$, and that the multishell interpretation (41) provides an alternative derivation of this classical result.

B. The Optimal Profile

We now determine the rate profile that maximizes the achievable exponent (37)–(38). The optimization problem is

$$\max_{R(\alpha)} E_{\circledcirc}(R) = \max_{R(\alpha)} \min_{\alpha\in[0,1]} \big\{ E_\alpha(R(\alpha)) + R - R(\alpha) \big\}, \tag{44}$$

subject to the constraint in (36). The structure of (44) is a max-min problem. The optimal profile equalizes the exponent contribution across all shells: if some shell α' had a strictly lower value of $E_{\alpha'}(R(\alpha')) + R - R(\alpha')$ than others, we could improve the minimum by adjusting the rate allocation. Therefore, the optimal profile $R^\star(\alpha)$ satisfies

$$E_\alpha(R^\star(\alpha)) + R - R^\star(\alpha) = E^\star \tag{45}$$

for all $\alpha \in [0,1]$, where $E^\star$ is the common value.

For the outer shell $\alpha = 1$, we have $D(1) = 0$ and the constraint in (36) requires $R^\star(1) = R$. The exponent contribution from the outer shell is therefore

$$E_1(R) + R - R = E_1(R) = E_{\rm cc}(R), \tag{46}$$

where the last equality follows from (39), which recovers Gallager's cost-constrained exponent. Hence, $E^\star = E_{\rm cc}(R)$.

For inner shells $\alpha < 1$, the condition (45) becomes

$$E_\alpha(R^\star(\alpha)) = E_{\rm cc}(R) - R + R^\star(\alpha). \tag{47}$$

Since $E_\alpha(R(\alpha))$ is a decreasing function of $R(\alpha)$ and the right-hand side of (47) is increasing in $R^\star(\alpha)$, there exists a unique solution for each α. Therefore, the multishell ensemble with rate profile $R^\star(\alpha)$ in (45) achieves the error exponent

$$E_{\circledcirc}(R) = E_{\rm cc}(R). \tag{48}$$

This result shows that an exponential number of messages may be placed inside the sphere $\mathcal{S}^n(P)$, not just on its outer shell, without loss of error exponent. The optimal rate profile allocates fewer messages to inner shells, precisely compensating for their large-deviation penalty $D(\alpha) < 0$.

IV. Gaussian Signalling

We now specialize to Gaussian signalling, $Q(x) = \frac{1}{\pi P} e^{-\frac{|x|^2}{P}}$ over $\mathcal{X} = \mathbb{C}$. The signal-to-noise ratio is ${\rm snr} = P/\sigma^2$, and the capacity is $C = \log(1 + {\rm snr})$. The cumulant-generating function (35) becomes

$$\omega(\tau) = -\log(1 - \tau P), \tag{49}$$

valid for $\tau < 1/P$. The large-deviation exponent (34) evaluates to

$$D(\alpha) = 1 - \alpha + \log \alpha. \tag{50}$$

As anticipated, $D(1) = 0$ and $D(\alpha) < 0$ for $\alpha \in (0, 1)$.

The Gallager function (8) becomes

$$E_1(\rho, r, \beta) = \log(1 - rP) + \beta r(1+\rho) + \rho \log\left(1 - rP + \frac{\rm snr}{1+\rho}\right), \tag{51}$$

which recovers the complex-valued version of [2, Eq. (7.4.21)] when $\beta = P$.

From (40) and (50), the rate induced by the i.i.d. ensemble in shell α is

$$R^{\rm iid}(\alpha) = R + 1 - \alpha + \log \alpha. \tag{52}$$

This profile decreases monotonically from $R^{\rm iid}(1) = R$ at the outer shell to $R^{\rm iid}(\alpha) \to -\infty$ as $\alpha \to 0$. Negative values of a rate profile $R(\alpha)$ corresponds to shells α with an expected number of messages that vanishes asymptotically.

The optimal rate allocation $R^\star(\alpha)$ is determined implicitly by (47). Substituting (50) and (51) into (39), we obtain $E_\alpha(R(\alpha))$ for Gaussian signalling, and $R^\star(\alpha)$ is the unique solution to

$$\max_{\rho \in [0,1]} \max_{r \in \mathbb{R}} \bigl\{ E_1(\rho, r, \alpha P) + (1+\rho) D(\alpha) - \rho R^\star(\alpha) \bigr\} = E_{\rm cc}(R) - R + R^\star(\alpha). \tag{53}$$

While (53) does not admit a closed-form solution, it can be solved numerically for each α.

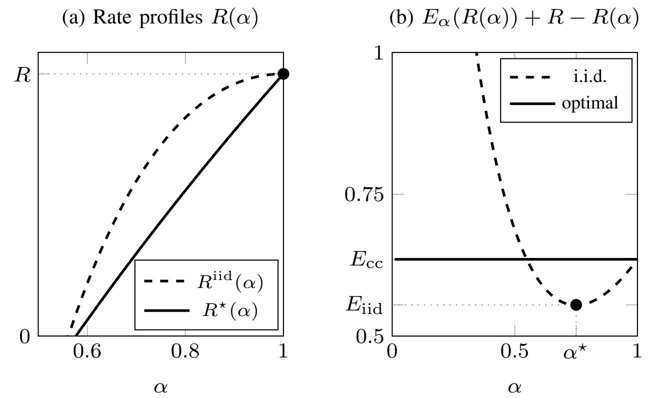

Fig. 1. (a) Rate profiles $R^{\rm iid}(\alpha) = R + D(\alpha)$ (dashed) and $R^\star(\alpha)$ (solid), and (b) per-shell exponent contribution $E_\alpha(R(\alpha)) + R - R(\alpha)$ for Gaussian signalling at ${\rm snr} = 2$ and $R = \frac{\log 2}{5}$ nats per channel use. The i.i.d.-induced profile yields a minimum exponent $E_{\rm iid}(R) = 0.555$ at $\alpha^\star = 0.75$. The optimal profile equalizes the exponent at $E_{\rm cc}(R) = 0.635$ for all $\alpha \in [0,1]$.

Fig. 1 compares the i.i.d.-induced and optimal rate profiles for Gaussian signalling at ${\rm snr} = 2$ and $R = \frac{\log 2}{5}$ nats per channel use. Compared to the profile induced by i.i.d. codes, the optimal profile $R^\star(\alpha)$ allocates less rate to shells near the outer shell than the i.i.d. induced profile, a redistribution that equalizes the per-shell exponent contribution $E_\alpha(R(\alpha)) + R - R(\alpha)$ across all shells. Under the i.i.d.-induced profile, the minimum occurs at $\alpha^\star = 0.75$, yielding $E_{\rm iid}(R) = 0.555$. Under the optimal profile, the exponent equals $E_{\rm cc}(R) = 0.635$ for all $\alpha \in [0,1]$. The gap $E_{\rm cc}(R) - E_{\rm iid}(R) = 0.080$ quantifies the suboptimality of the i.i.d. ensemble, which arises not from the use of the interior of the sphere, but from the particular rate allocation induced by i.i.d. codeword generation.

References

[1] C. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. Journal*, vol. 38, no. 3, pp. 611–656, May 1959.

[2] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, Inc., 1968.

[3] W. Feller, *An Introduction to Probability Theory and its Applications*. New York, NY: Wiley, 1971, vol. 2.

[4] T. Liu, P. Moulin, and R. Koetter, "On error exponents of modulo lattice additive noise channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 454–471, Feb. 2006.

[5] 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, May 2010.

[6] A. Tauste Campo, G. Vazquez-Vilar, A. Guillén i Fàbregas, T. Koch, and A. Martinez, "A derivation of the source-channel error exponent using nonidentical product distributions," *IEEE Trans. Inf. Th.*, vol. 60, no. 6, pp. 3209–3217, Jun. 2014.

[7] J. Jensen, *Saddlepoint Approximations*. Oxford University Press, 1995.

[8] K. Fan, "Minimax theorems," *Proc. National Academy of Sciences*, vol. 39, no. 1, pp. 42–47, 1953.