A Simple Derivation of the Refined SPB for the Constant Composition Codes

Barış Nakiboğlu
Department of Electrical and Electronics Engineering
Middle East Technical University (METU),
06800 Ankara Turkey
Email:bnakib@metu.edu.tr

***Abstract*—A judicious application of the Berry-Esseen theorem via the concepts of Augustin information and mean is demonstrated to be sufficient for deriving the sphere packing bound with a prefactor that is $\Omega\left(n^{-0.5(1-E'_{sp}(R,W,p))}\right)$ for the constant composition codes. The resulting non-asymptotic bounds have definite approximation error terms.**

I. Introduction

The decay of the optimal error probability with the block length for rates below the channel capacity has been studied since the early days of information theory. For certain channels and for certain values of the rate, sharp bounds were found early on. For the binary symmetric channel [1], for the Gaussian channel [2], and for the discrete channels with certain symmetries [3] —see the original publication in Russian so as to avoid typos in the translation—

$$P_{\mathbf{e}}^{(n)} = \Theta\left(n^{-\frac{1-E'_{sp}(R)}{2}} e^{-nE_{sp}(R)}\right) \qquad \forall R \in [R_{crit}, C] \quad (1)$$

where[1] $a_n = \Theta(b_n)$ iff $0 < \liminf_{n\to\infty} \left|\frac{a_n}{b_n}\right| \leq \limsup_{n\to\infty} \left|\frac{a_n}{b_n}\right| < \infty$, $E_{sp}(\cdot)$ is the sphere packing exponent of the channel, $E'_{sp}(\cdot)$ is its derivative, R_{crit} is the rate at which the slope of the sphere packing exponent curve is minus one, i.e. $E'_{sp}(R_{crit}) = -1$, and C is the capacity of the channel. For the binary erasure channels [1], on the other hand,

$$P_{\mathbf{e}}^{(n)} = \Theta\left(n^{-\frac{1}{2}} e^{-nE_{sp}(R)}\right) \qquad \forall R \in [R_{crit}, C]. \quad (2)$$

Neither (1), nor (2), holds for rates below the critical rate. If, however, we replace the equality sign with the greater than or equal to sign, then both (1) and (2) hold for all rates below the channel capacity. These lower bounds are customarily called sphere packing bounds (SPBs) because of the techniques used in their derivation.

Derivations of the SPB in [1]–[3] relied on the geometric structure of the output space of the channel and parameters that are defined only for certain models. The resulting bounds were expressed in terms of these parameters, as well. Thus it was not even clear that SPBs in [1]–[3] can be seen as special cases of a general bound. The evidence for such an understanding came not from a breakthrough about the lower bounds on the error probability but from a breakthrough about the upper bounds. Gallager's seminal work [4] unified and generalized the upper bounds on the error probability —at least in terms of the exponent— in all the previous works. It is only with Gallager's formulation in [4] that one can express the bounds in [1]–[3] as (1) and (2).

The first complete proof of the SPB for arbitrary discrete stationary product channels[2] (DSPCs) was presented in [5]. According to [5, Thm. 2]

$$P_{\mathbf{e}}^{(n)} \geq e^{-n\left[E_{sp}\left(R-O\left(n^{-1/2}\right)\right)+O\left(n^{-1/2}\right)\right]} \qquad \forall R \in (0, C) \quad (3)$$

where $a_n = O(b_n)$ iff there exists a $K \in \mathbb{R}_+$ such that $|a_n| \leq K b_n$ for large enough n. In the following two years, the SPB was proved first for stationary product channels with finite input sets in [6] and then for (possibly) non-stationary product channels in [7]. The SPB has been proven for various channel models [8]–[13], including certain quantum information theoretic ones. It is worth mentioning that a general proof that holds for both Gaussian channels [2] and for arbitrary DSPCs [5] was absent until recently, see [11]. These later works on the SPB [5]–[13] were primarily interested in establishing the right exponent; thus they were content with prefactors of the form $e^{-O\left(n^{1/2}\right)}$. Some authors did obtain prefactors of the form $e^{-O(\ln n)}$ but obtaining the best possible, if not tight, prefactor was not really a concern.

The quest for deriving SPBs with tight prefactors was put on the map again by Altuğ and Wagner in [14] and [15]. According to [14, Thm. 1] for any DSPC with a positive and symmetric[3] probability transition matrix W and rate R in $(0, C)$, there exists a $K \in \mathbb{R}_+$ such that for any $\epsilon > 0$

$$P_{\mathbf{e}}^{(n)} \geq K n^{-\frac{1-(1+\epsilon)E'_{sp}(R)}{2}} e^{-nE_{sp}(R)} \qquad \forall n \geq n_0 \quad (4)$$

for some n_0 determined by W, R, and ϵ. The same result was established for the constant composition codes on arbitrary DSPCs in [15, Thm. 1]. These results are generalized to

[1] We suppress the dependence of the sphere packing exponent to the channel in our notation and denote it with $E_{sp}(R)$, rather than $E_{sp}(R, W)$, in §I.

[2] These channels are customarily called discrete memoryless channels, i.e. DMCs. We call them DSPCs in order to underline the stationarity of the channel and the absence of any constraints on its input set. Such constraints might exist and stationarity might be absent in a discrete channel which is memoryless.

[3] A W is symmetric if it satisfies the condition given in [8, p. 94]. The binary symmetric channel, the binary erasure channel, and channels considered in [3] are symmetric according this definition.

classical quantum channels in [16, Thms. 8, 14], with a slight improvement, allowing $\epsilon = 0$ for the symmetric channels.

The primary tool for the deviations in [14]–[16], is the Berry-Esseen theorem, albeit through auxiliary results [14, (74)], [15, Proposition 5], [16, Thm. 17] inspired by a theorem of Bahadur and Rao [17]. Our main aim is to demonstrate that a more judicious application of the Berry-Esseen theorem via the concepts of Augustin information and mean leads to a simpler analysis and a stronger result for the constant composition codes, see Theorem 1 and its proof in §V. Under various symmetry hypotheses a similar approach employing Augustin capacity and center leads to analogous results, [18].

[1] and [2] not only established (1) and (2) but also obtained closed form expressions for the upper and lower bounds implicit in (1) and (2). Dobrushin went one step further and calculated the exact asymptotic behavior of the SPB and the random coding bound by analyzing the lattice and non-lattice cases separately in [3, (1.32), (1.33), (1.34)]. Recently, the saddle point approximation is used to derive the SPB with the same asymptotic prefactor [19, Cor. 2], under weaker symmetry hypothesis[4] albeit by assuming a common support for all output distributions of the channel and a non-lattice structure for the random variables involved.[5] The main drawback of the analysis in [19] is the technical conditions that need to be confirmed for applying the saddle point approximation via [20, Proposition 2.3.1].

Let us finish this section with an overview of the paper. In §II, we describe our notation. In §III, we first recall the connection between the hypothesis testing problem and the tilting, and then derive our primary technical tool using the Berry-Esseen theorem. In §IV, we review the concepts of Augustin information, Augustin mean, and the sphere packing exponent. In §V, we state and prove a refined SPB for the constant composition codes using Lemma 2 of §III and the observations recalled in §IV. In §VI, we discuss the generalizations and the weaknesses of our main result.

II. Model and Notation

For any set $\mathcal{X}$, we denote the set of all probability mass functions that are non-zero only on finitely many elements of $\mathcal{X}$ by $\mathcal{P}(\mathcal{X})$. For any measurable space $(\mathcal{Y},\mathcal{Y})$, we denote the set of all probability measures on it by $\mathcal{P}(\mathcal{Y})$. We denote the expected value a measurable function f under the probability measure μ by $\mathbf{E}_\mu[f]$ or $\mathbf{E}_\mu[f(\mathsf{Y})]$. Similarly we denote the variance of f under μ, i.e. $\mathbf{E}_\mu\left[(f-\mathbf{E}_\mu[f])^2\right]$, by $\mathbf{V}_\mu[f]$.

For sets $\mathcal{X}_1,\ldots,\mathcal{X}_n$ we denote their Cartesian product, i.e. $\bigtimes_{t=1}^n \mathcal{X}_t$, by $\mathcal{X}_1^n$ and for σ-algebras $\mathcal{Y}_1,\ldots,\mathcal{Y}_n$ we denote their product, i.e. $\bigotimes_{t=1}^n \mathcal{Y}_t$, by $\mathcal{Y}_1^n$. We use the symbol $\otimes$ to denote the product of measures, as well.

A *channel* W is a function from *the input set* $\mathcal{X}$ to the set of all probability measures on *the output space* $(\mathcal{Y},\mathcal{Y})$:

$$W : \mathcal{X} \to \mathcal{P}(\mathcal{Y}). \tag{5}$$

[4]The binary input Gaussian channel and the binary erasure channel satisfy the symmetry hypothesis of [19], but not that of [3].

[5]Neither of these assumptions was needed while deriving this result in [3].

A channel W is called a *discrete channel* if both $\mathcal{X}$ and $\mathcal{Y}$ are finite sets. The product of $W_t : \mathcal{X}_t \to \mathcal{P}(\mathcal{Y}_t)$ for $t \in \{1,\ldots,n\}$ is a channel of the form $W_{[1,n]} : \mathcal{X}_1^n \to \mathcal{P}(\mathcal{Y}_1^n)$ satisfying

$$W_{[1,n]}(x_1^n) = \bigotimes_{t=1}^n W_t(x_t) \qquad \forall x_1^n \in \mathcal{X}_1^n. \tag{6}$$

Any channel obtained by curtailing the input set of a length n product channel is called a length n *memoryless channel*. A product channel $W_{[1,n]}$ is *stationary* iff $W_t = W$ for all t's for some W. On a stationary channel, we denote the composition (i.e. the empirical distribution, type) of each x_1^n by $\Upsilon(x_1^n)$; thus $\Upsilon(x_1^n) \in \mathcal{P}(\mathcal{X})$.

The pair (Ψ,Θ) is an (M,L) *channel code* on $W_{[1,n]}$ iff

- The *encoding function* Ψ is a function from the message set $\mathcal{M} \triangleq \{1,2,\ldots,M\}$ to the input set $\mathcal{X}_1^n$.
- The *decoding function* Θ is a $\mathcal{Y}_1^n$-measurable function from the output set $\mathcal{Y}_1^n$ to the set $\widehat{\mathcal{M}} \triangleq \{\mathcal{L} : \mathcal{L} \subset \mathcal{M} \text{ and } |\mathcal{L}| \le L\}$.

Given an (M,L) channel code (Ψ,Θ) on $W_{[1,n]}$, *the conditional error probability* $P_{\mathbf{e}}^m$ for $m \in \mathcal{M}$ and *the average error probability* $P_{\mathbf{e}}$ are given by

$$P_{\mathbf{e}}^m \triangleq \mathbf{E}_{W_{[1,n]}(\Psi(m))}\left[\mathbb{1}_{\{m \notin \Theta(\mathsf{Y}_1^n)\}}\right],$$
$$P_{\mathbf{e}} \triangleq \tfrac{1}{M}\sum\nolimits_{m\in\mathcal{M}} P_{\mathbf{e}}^m.$$

A code is called a constant composition code iff all of its codewords have the same composition, i.e. there exists a p in $\mathcal{P}(\mathcal{X})$ satisfying $\Upsilon(\Psi(m)) = p$ for all $m \in \mathcal{M}$.

III. Hypothesis Testing Problem and Berry-Esseen Theorem

The primary aim of this section is to derive an outer bound for the hypothesis testing problem between product measures using the Berry-Esseen theorem. To that end let us first recall the definitions of the Rényi divergence and the tilted probability measure.

Definition 1. For any $\alpha \in \mathbb{R}_+$ and $w, q \in \mathcal{P}(\mathcal{Y})$, *the order* α *Rényi divergence between* w *and* q is

$$D_\alpha(w\|\, q) \triangleq \begin{cases} \frac{1}{\alpha-1}\ln\int(\frac{\mathrm{d}w}{\mathrm{d}\nu})^\alpha(\frac{\mathrm{d}q}{\mathrm{d}\nu})^{1-\alpha}\nu(\mathrm{d}y) & \alpha \neq 1 \\ \int \frac{\mathrm{d}w}{\mathrm{d}\nu}\left[\ln\frac{\mathrm{d}w}{\mathrm{d}\nu} - \ln\frac{\mathrm{d}q}{\mathrm{d}\nu}\right]\nu(\mathrm{d}y) & \alpha = 1 \end{cases} \tag{7}$$

where ν is any measure satisfying $w \prec \nu$ and $q \prec \nu$.

The order one Rényi divergence is the Kullback-Leibler divergence. For other orders, the Rényi divergence can be characterized in terms of the Kullback-Leibler divergence too:

$$(1-\alpha)D_\alpha(w\|\, q) = \inf_{v\in\mathcal{P}(\mathcal{Y})} \alpha D_1(v\|\, w) + (1-\alpha)D_1(v\|\, q) \tag{8}$$

with the convention that $\alpha D_1(v\|\, w) + (1-\alpha)D_1(v\|\, q) = \infty$ if it would be otherwise undefined, see [21, Thm. 30]. The characterization given in (8) is related to another key concept for our analysis: the tilted probability measure.

Definition 2. For any $\alpha \in \mathbb{R}_+$ and $w, q \in \mathcal{P}(\mathcal{Y})$ satisfying $D_\alpha(w\|\, q) < \infty$, *the order* α *tilted probability measure* w_α^q is

$$\tfrac{\mathrm{d}w_\alpha^q}{\mathrm{d}\nu} \triangleq e^{(1-\alpha)D_\alpha(w\|q)}(\tfrac{\mathrm{d}w}{\mathrm{d}\nu})^\alpha(\tfrac{\mathrm{d}q}{\mathrm{d}\nu})^{1-\alpha}. \tag{9}$$

If either α is in $(0,1)$ or $D_1(w_\alpha^q \| w)$ is finite, then the tilted probability measure is the unique probability measure achieving the infimum in (8) by [21, Thm. 30], i.e.

$$(1-\alpha)D_\alpha(w\| q) = \alpha D_1(w_\alpha^q \| w) + (1-\alpha) D_1(w_\alpha^q \| q)\,. \quad (10)$$

Furthermore, under the same hypothesis

$$\ln \tfrac{\mathrm{d}w_\alpha^q}{\mathrm{d}q} - D_1(w_\alpha^q \| q) = \alpha \left(\ln \tfrac{\mathrm{d}w_{ac}}{\mathrm{d}q} - \mathbf{E}_{w_\alpha^q}\left[\ln \tfrac{\mathrm{d}w_{ac}}{\mathrm{d}q} \right] \right) \quad (11)$$

$$\ln \tfrac{\mathrm{d}w_\alpha^q}{\mathrm{d}w} - D_1(w_\alpha^q \| w) = (\alpha-1) \left(\ln \tfrac{\mathrm{d}w_{ac}}{\mathrm{d}q} - \mathbf{E}_{w_\alpha^q}\left[\ln \tfrac{\mathrm{d}w_{ac}}{\mathrm{d}q} \right] \right) \quad (12)$$

where w_{ac} is the component of w that is absolutely continuous in q.

Let us proceed with recalling the Berry-Esseen theorem.

Lemma 1 ([22])**.** *Let $\{\xi_t\}_{t\in\mathbb{Z}_+}$ be independent zero mean random variables satisfying $\sum_{t=1}^n \mathbf{E}\left[{\xi_t}^2\right] < \infty$. Then there exists an absolute constant $\omega \leq 0.5600$ such that*

$$\left| \mathbf{P}\left[\sum\nolimits_{t=1}^n \xi_t < \tau \right] - \Phi\left(\tfrac{\tau}{\sqrt{a_2 n}} \right) \right| \leq \omega \tfrac{a_3}{a_2\sqrt{a_2 n}}$$

where $a_\kappa = \frac{1}{n}\sum_{t=1}^n \mathbf{E}[|\xi_t|^\kappa]$ and $\Phi(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^s e^{-z^2/2}\mathrm{d}z$.

Lemma 2, in the following, is an impossibility result for the hypothesis testing problem with independent samples in the spirit of [5, Thm. 5], which is proved using the Chebyshev's inequality. Lemma 2, on the other hand, is proved using the Berry-Esseen theorem instead.

Lemma 2. *For any $\alpha \in (0,1)$, $n \in \mathbb{Z}_+$, $w_t, q_t \in \mathcal{P}(\mathcal{Y}_t)$, let $w_{t,ac}$ be the component of w_t that is absolutely continuous in q_t and let a_2 and a_3 be*

$$a_2 \triangleq \tfrac{1}{n}\sum\nolimits_{t=1}^n \mathbf{E}_{w_\alpha^q}\left[\left| \ln \tfrac{\mathrm{d}w_{t,ac}}{\mathrm{d}q_t} - \mathbf{E}_{w_\alpha^q}\left[\ln \tfrac{\mathrm{d}w_{t,ac}}{\mathrm{d}q_t} \right] \right|^2 \right] \quad (13)$$

$$a_3 \triangleq \tfrac{1}{n}\sum\nolimits_{t=1}^n \mathbf{E}_{w_\alpha^q}\left[\left| \ln \tfrac{\mathrm{d}w_{t,ac}}{\mathrm{d}q_t} - \mathbf{E}_{w_\alpha^q}\left[\ln \tfrac{\mathrm{d}w_{t,ac}}{\mathrm{d}q_t} \right] \right|^3 \right] \quad (14)$$

where $w = \otimes_{t=1}^n w_t$ and $q = \otimes_{t=1}^n q_t$. Then for any $\mathcal{E} \in \mathcal{Y}_1^n$, and $\gamma \in \mathbb{R}_+$ satisfying $q(\mathcal{E}) = \gamma e^{-D_1(w_\alpha^q \| q)}$ we have

$$w(\mathcal{Y}_1^n \setminus \mathcal{E}) \geq \frac{2-\gamma}{e^{(1-\alpha)2\sqrt{2\pi e}\left(0.56\frac{a_3}{a_2} + \sqrt{a_2}\right)}} n^{-\frac{1}{2\alpha}} e^{-D_1(w_\alpha^q \| w)} \quad (15)$$

provided that $\sqrt{a_2 n} - \frac{\ln n}{2\alpha} \geq 2\sqrt{2\pi e}\left(0.56\frac{a_3}{a_2} + \sqrt{a_2}\right)$.

Lemma 2 states just a lower bound because that suffices for our purposes. For each α in $(0,1)$, however, a matching upper bound can be derived using the Berry-Esseen theorem, as well. That is, the lower bound given in Lemma 2 is tight up to some multiplicative constant. For the stationary case —i.e. assuming $w_t = w_1$, $q_t = q_1$ for all t— with $\gamma = 1$, Csiszár and Longo [23] described how (11) and (12) can be used together with an earlier result by Strassen [24, Thm. 1.1] to obtain the exact value of the constant.[6] The Berry-Esseen theorem, however, is not sufficient for deriving the exact value of the constant; one needs to invoke either finer results on the asymptotic behavior of sums of independent random variables such as the ones in [25, §IV.2,§IV.3], [26, §42,§43] or apply other techniques, such as the saddle point approximation in [20, Prop. 2.3.1], both of which require stronger hypotheses. The situation is similar for other values of α but of no interest for our discussion of the sphere packing bound.

[6] We believe the approach of [23] is sound. Its calculations, however, seem to have some mistakes. Repeating the calculations as described in [23] we recover the second line of [23, (33)] as $\ln \frac{\alpha^\star}{1-\alpha^\star} - \frac{\ln S_1\sqrt{2\pi}}{\alpha^\star} + o(1)$. With this modification [23, Thm. 2] is consistent with the intimately related results about the SPB proved earlier [3, (1.32), (1.33)] and since then [19, (38)].

Proof of Lemma 2. Let the random variables ξ_t and ξ and the event $\mathcal{B}$ be

$$\begin{aligned} \xi_t &\triangleq \ln \tfrac{\mathrm{d}w_{t,ac}}{\mathrm{d}q_t} - \mathbf{E}_{w_\alpha^q}\left[\ln \tfrac{\mathrm{d}w_{t,ac}}{\mathrm{d}q_t} \right], \\ \xi &\triangleq \sum\nolimits_{t=1}^n \xi_t, \\ \mathcal{B} &\triangleq \{ y_1^n : \tau_0 \leq \xi \leq \tau_1 \}. \end{aligned}$$

The definitions of ξ_t, ξ, $\mathcal{B}$ when considered together with (11) and (12) imply that

$$\begin{aligned} \mathcal{B} &= \left\{ y_1^n : \alpha\tau_0 \leq \ln \tfrac{\mathrm{d}w_\alpha^q}{\mathrm{d}q} - D_1(w_\alpha^q \| q) \leq \alpha\tau_1 \right\} \\ &= \left\{ y_1^n : (1-\alpha)\tau_0 \leq D_1(w_\alpha^q \| w) - \ln \tfrac{\mathrm{d}w_\alpha^q}{\mathrm{d}w} \leq (1-\alpha)\tau_1 \right\}. \end{aligned}$$

Thus for any $\mathcal{E} \in \mathcal{Y}_1^n$ we have

$$w_\alpha^q(\mathcal{E} \cap \mathcal{B}) \leq q(\mathcal{E}) e^{D_1(w_\alpha^q \| q) + \alpha\tau_1}, \quad (16)$$

$$w(\mathcal{Y}_1^n \setminus \mathcal{E}) \geq w_\alpha^q(\mathcal{B} \setminus \mathcal{E}) e^{-D_1(w_\alpha^q \| w) + (1-\alpha)\tau_0}. \quad (17)$$

On the other hand, as a result of the Berry-Esseen theorem given in Lemma 1 we have

$$\begin{aligned} w_\alpha^q(\mathcal{B}) &\geq \Phi\left(\tfrac{\tau_1}{\sqrt{a_2 n}} \right) - \Phi\left(\tfrac{\tau_0}{\sqrt{a_2 n}} \right) - 2\tfrac{0.56}{\sqrt{n}} \tfrac{a_3}{a_2\sqrt{a_2}} \\ &= \tfrac{1}{\sqrt{2\pi}} \int_{\frac{\tau_0}{\sqrt{a_2 n}}}^{\frac{\tau_1}{\sqrt{a_2 n}}} e^{-z^2/2}\mathrm{d}z - 2\tfrac{0.56}{\sqrt{n}} \tfrac{a_3}{a_2\sqrt{a_2}} \\ &\geq \tfrac{e^{-\frac{(|\tau_0| \vee |\tau_1|)^2}{2n a_2}}}{\sqrt{2\pi}} \tfrac{\tau_1 - \tau_0}{\sqrt{a_2 n}} - 2\tfrac{0.56}{\sqrt{n}} \tfrac{a_3}{a_2\sqrt{a_2}}. \end{aligned}$$

If we set $\tau_0 = \frac{-\ln n}{2\alpha} - 2\sqrt{2\pi e}\left(0.56\frac{a_3}{a_2} + \sqrt{a_2}\right)$ and $\tau_1 = \frac{-\ln n}{2\alpha}$, then $-\sqrt{a_2 n} \leq \tau_0 \leq \tau_1 \leq 0$ by the hypothesis and

$$w_\alpha^q(\mathcal{B}) \geq \tfrac{2}{\sqrt{n}}.$$

Furthermore, $w_\alpha^q(\mathcal{E} \cap \mathcal{B}) \leq \frac{\gamma}{\sqrt{n}}$ as a result of (16), $\tau_1 = \frac{-\ln n}{2\alpha}$, and the hypothesis $q(\mathcal{E}) = \gamma e^{-D_1(w_\alpha^q \| q)}$. Thus using (17) and $\tau_1 = \frac{-\ln n}{2\alpha}$ we get

$$\begin{aligned} w(\mathcal{Y}_1^n \setminus \mathcal{E}) &\geq \tfrac{2-\gamma}{\sqrt{n}} e^{-D_1(w_\alpha^q \| w) + (1-\alpha)\tau_0} \\ &= \tfrac{2-\gamma}{e^{(1-\alpha)(\tau_1 - \tau_0)}} n^{-\frac{1}{2\alpha}} e^{-D_1(w_\alpha^q \| w)}. \end{aligned}$$

Then (15) follows from $\tau_1 - \tau_0 = 2\sqrt{2\pi e}\left(0.56\frac{a_3}{a_2} + \sqrt{a_2}\right)$, which is an immediate consequence of the definitions of τ_0 and τ_1. □

Remark. While deriving similar bounds, the constants τ_0 and τ_1 are usually assumed to satisfy $\tau_0 = -\tau_1$, see for example [5, Thm. 5] or [16, Thm. 11]. Such a choice, however, does not lead to tight bounds in our case.

IV. Augustin Information, Augustin Mean and The Sphere Packing Exponent

The ultimate aim is this section is to define the sphere packing exponent and review the properties of it that will be useful in our analysis. For that we first recall the definitions of Augustin information and mean and review their elementary properties. Let us define the conditional Rényi divergence first.

Definition 3. For any $\alpha \in \mathbb{R}_+$, $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $q \in \mathcal{P}(\mathcal{Y})$, and $p \in \mathcal{P}(\mathcal{X})$ *the order* α *conditional Rényi divergence for the input distribution* p is

$$D_\alpha(W\|\, q|\, p) \triangleq \sum\nolimits_{x\in\mathcal{X}} p(x) D_\alpha(W(x)\|\, q)\,. \tag{18}$$

Definition 4. For any $\alpha \in \mathbb{R}_+$, $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, and $p \in \mathcal{P}(\mathcal{X})$ *the order* α *Augustin information for the input distribution* p is

$$I_\alpha(p; W) \triangleq \inf\nolimits_{q\in\mathcal{P}(\mathcal{Y})} D_\alpha(W\|\, q|\, p)\,. \tag{19}$$

The infimum in (19) is achieved by a unique probability measure[7] $q_{\alpha,p}$, called *the order* α *Augustin mean for the input distribution* p, by [27, Lemma 13]. Furthermore,

$$D_\alpha(W\|\, q|\, p) - I_\alpha(p; W) \geq D_{1\wedge\alpha}(q_{\alpha,p}\|\, q) \tag{20}$$

for all $q \in \mathcal{P}(\mathcal{Y})$ by [27, Lemma 13], as well.

The Augustin information is continuously differentiable in its order on $\mathbb{R}_+$ and its derivative is given by

$$\tfrac{\partial}{\partial\alpha} I_\alpha(p; W) = \begin{cases} \frac{1}{(\alpha-1)^2} D_1\big(W_\alpha^{q_{\alpha,p}}\big\|\, W\big|\, p\big) & \alpha \neq 1 \\ \sum_x \frac{p(x)}{2} \mathbf{V}_{W(x)}\Big[\ln \frac{\mathrm{d}W(x)}{\mathrm{d}q_{1,p}}\Big] & \alpha = 1 \end{cases} \tag{21}$$

by [27, Lemma 17-(e)], where $W_\alpha^{q_{\alpha,p}}$ is the tilted channel defined as follows.

Definition 5. For any $\alpha \in \mathbb{R}_+$, $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ and $q \in \mathcal{P}(\mathcal{Y})$, *the order* α *tilted channel* W_α^q is a function from $\{x : D_\alpha(W(x)\|\, q) < \infty\}$ to $\mathcal{P}(\mathcal{Y})$ given by

$$\tfrac{\mathrm{d}W_\alpha^q(x)}{\mathrm{d}\nu} \triangleq e^{(1-\alpha)D_\alpha(W(x)\|q)} \big(\tfrac{\mathrm{d}W(x)}{\mathrm{d}\nu}\big)^\alpha \big(\tfrac{\mathrm{d}q}{\mathrm{d}\nu}\big)^{1-\alpha}. \tag{22}$$

The concept of tilted channel can also be used to express $I_\alpha(p; W)$ in terms of the Kullback-Leibler divergences. In particular, (10) and $I_\alpha(p; W) = D_\alpha(W\|\, q_{\alpha,p}|\, p)$ imply that

$$I_\alpha(p; W) = \tfrac{\alpha}{1-\alpha} D_1(W_\alpha^{q_{\alpha,p}}\|\, W|\, p) + D_1(W_\alpha^{q_{\alpha,p}}\|\, q_{\alpha,p}|\, p). \tag{23}$$

Furthermore, the Augustin mean satisfies

$$\sum\nolimits_x p(x)\, W_\alpha^{q_{\alpha,p}}(x) = q_{\alpha,p} \tag{24}$$

and Augustin mean is the only probability measure satisfying both $q_{1,p} \prec q$ and $\sum_x p(x)\, W_\alpha^q(x) = q$ by [27, Lemma 13]. Thus for all $\alpha \in \mathbb{R}_+$ we have

$$D_1(W_\alpha^{q_{\alpha,p}}\|\, q_{\alpha,p}|\, p) = I_1(p; W_\alpha^{q_{\alpha,p}})\,. \tag{25}$$

A more comprehensive discussion of Augustin's information measures can be found in [27].

[7] We refrain from including the channel symbol W in the symbol for the mean because the channel will be clear from the context.

Definition 6. For any $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $p \in \mathcal{P}(\mathcal{X})$, and $R \in \mathbb{R}_+$, the sphere packing exponent (SPE) is

$$E_{sp}(R, W, p) \triangleq \sup\nolimits_{\alpha\in(0,1)} \tfrac{1-\alpha}{\alpha} \left(I_\alpha(p; W) - R\right). \tag{26}$$

Note that since $I_\alpha(p; W)$ is continuously differentiable in α by [27, Lemma 17-(e)], we can apply the derivative test to find the optimal α in (26): Using (21) and (23) we get

$$\tfrac{\partial}{\partial\alpha} \tfrac{1-\alpha}{\alpha} \left(I_\alpha(p; W) - R\right) = \tfrac{1}{\alpha^2}\left(R - D_1(W_\alpha^{q_{\alpha,p}}\|\, q_{\alpha,p}|\, p)\right). \tag{27}$$

On the other hand, either $D_1\big(W_\alpha^{q_{\alpha,p}}\big\|\, q_{\alpha,p}\big|\, p\big) = I_1(p; W)$ for all positive α, or $D_1\big(W_\alpha^{q_{\alpha,p}}\big\|\, q_{\alpha,p}\big|\, p\big)$ is increasing and continuous in α on $\mathbb{R}_+$ by [27, Lemma 17-(f)]. Furthermore, $D_1\big(W_1^{q_{1,p}}\big\|\, q_{1,p}\big|\, p\big)$ is equal to $I_1(p; W)$ by definition and $\lim_{\alpha\downarrow 0} D_1\big(W_\alpha^{q_{\alpha,p}}\big\|\, q_{\alpha,p}\big|\, p\big)$ is equal to $\lim_{\alpha\downarrow 0} I_\alpha(p; W)$ by (25) and [27, Lemma 17-(g)]. Thus for any rate R in $(\lim_{\alpha\downarrow 0} I_\alpha(p; W), I_1(p; W))$, there exists an $\eta \in (0, 1)$ satisfying

$$R = D_1\big(W_\eta^{q_{\eta,p}}\big\|\, q_{\eta,p}\big|\, p\big) \tag{28}$$

by the intermediate value theorem [28, 4.23]. The η satisfying (28) is unique because $D_1\big(W_\alpha^{q_{\alpha,p}}\big\|\, q_{\alpha,p}\big|\, p\big)$ is increasing in α. The monotonicity of $D_1\big(W_\alpha^{q_{\alpha,p}}\big\|\, q_{\alpha,p}\big|\, p\big)$ in α and (27) also implies $E_{sp}(R, W, p) = \frac{1-\eta}{\eta}\left(I_\eta(p; W) - R\right)$. Thus as a result of (23), the unique η satisfying (28) also satisfies

$$E_{sp}(R, W, p) = D_1\big(W_\eta^{q_{\eta,p}}\big\|\, W\big|\, p\big)\,. \tag{29}$$

Since $D_1\big(W_\alpha^{q_{\alpha,p}}\big\|\, q_{\alpha,p}\big|\, p\big)$ is continuous and increasing in α, its inverse is increasing and continuous, as well. Thus the definition of SPE given in (26) and the definition of derivative as a limit imply that for any R in $(\lim_{\alpha\downarrow 0} I_\alpha(p; W)\,, I_1(p; W))$ the unique η satisfying (28) also satisfies

$$\tfrac{\partial}{\partial R} E_{sp}(R, W, p) = \tfrac{\eta-1}{\eta}. \tag{30}$$

V. The Refined Sphere Packing Bound

Theorem 1. *For any* $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $M, L, n \in \mathbb{Z}_+$, $p \in \mathcal{P}(\mathcal{X})$ *satisfying* $\lim_{\alpha\downarrow 0} I_\alpha(p; W) < \frac{1}{n} \ln \frac{M}{L} < I_1(p; W)$ *and* $np(x) \in \mathbb{Z}_{\geq 0}$ *for all* $x \in \mathcal{X}$, *the order* $\alpha \triangleq \frac{1}{1 - E'_{sp}\left(\frac{1}{n} \ln \frac{M}{L}, W, p\right)}$ *satisfies*

$$D_1(W_\alpha^{q_{\alpha,p}}\|\, q_{\alpha,p}|\, p) = \tfrac{1}{n} \ln \tfrac{M}{L}. \tag{31}$$

Furthermore, any (M, L) *channel code of length* n *whose codewords all have the same composition* p *satisfies*

$$P_{\mathbf{e}}^{(n)} \geq \frac{n^{-1/2\alpha}}{e^{(1-\alpha)2\sqrt{2\pi e}[0.56\frac{a_3}{a_2^2} + \sqrt{a_2}]}} e^{-nE_{sp}\left(\frac{1}{n} \ln \frac{M}{L}, W, p\right)} \tag{32}$$

provided that $\sqrt{a_2 n} - \frac{\ln n}{2\alpha} \geq 2\sqrt{2\pi e}[0.56\frac{a_3}{a_2} + \sqrt{a_2}]$ *where*

$$a_2 = \mathbf{E}_{p \circledast W_\alpha^{q_{\alpha,p}}}\left[\left|\ln \frac{\mathrm{d}W}{\mathrm{d}q_{\alpha,p}} - \mathbf{E}_{W_\alpha^{q_{\alpha,p}}}\left[\ln \frac{\mathrm{d}W}{\mathrm{d}q_{\alpha,p}}\right]\right|^2\right], \tag{33}$$

$$a_3 = \mathbf{E}_{p \circledast W_\alpha^{q_{\alpha,p}}}\left[\left|\ln \frac{\mathrm{d}W}{\mathrm{d}q_{\alpha,p}} - \mathbf{E}_{W_\alpha^{q_{\alpha,p}}}\left[\ln \frac{\mathrm{d}W}{\mathrm{d}q_{\alpha,p}}\right]\right|^3\right]. \tag{34}$$

Theorem 1 proves that

$$P_{\mathbf{e}}^{(n)} \geq A n^{\frac{E'_{sp}(R,W,p)-1}{2}} e^{-nE_{sp}(R,W,p)} \qquad \forall n \geq n_0 \tag{35}$$

for constants A and n_0 determined by the rate R, the channel W, and the composition p. Following [14]–[16], we call these

bounds refined SPBs because of their resemblance to the standard SPBs, e.g. [6], establishing

$$P_{\mathbf{e}}^{(n)} \geq e^{-nE_{sp}(R,W,p)-o(n)} \qquad \forall n \geq n_0. \tag{36}$$

Proof of Theorem 1. The existence of a unique order α satisfying (31) was proved and its value was determined in §IV, see (28), (29), and (30).

Let probability measures w_m, q, and v_m in $\mathcal{P}(\mathcal{Y}_1^n)$ be

$$\begin{aligned} w_m &\triangleq \bigotimes_{t=1}^{n} W(\Psi_t(m)), \\ q &\triangleq \bigotimes_{t=1}^{n} q_{\alpha,p}, \\ v_m &\triangleq \bigotimes_{t=1}^{n} W_\alpha^{q_{\alpha,p}}(\Psi_t(m)). \end{aligned}$$

Then v_m is equal to the order α tilted probability measure between w_m and q. Furthermore,

$$\begin{aligned} D_1(v_m \| \, q) &= nD_1(W_\alpha^{q_{\alpha,p}} \| \, q_{\alpha,p} | \, p) \qquad m \in \mathcal{M}, \\ D_1(v_m \| \, w_m) &= nD_1(W_\alpha^{q_{\alpha,p}} \| \, W | \, p) \qquad m \in \mathcal{M}. \end{aligned}$$

Then by applying Lemma 2 for $\mathcal{E} = \{y_1^n : m \in \Theta(y_1^n)\}$ and $\gamma = q(m \in \Theta) e^{nD_1\left(W_\alpha^{q_{\alpha,p}} \| \, q_{\alpha,p} | p\right)}$ we get

$$P_{\mathbf{e}}^{m} \geq \frac{2 - q(m \in \Theta) e^{nD_1\left(W_\alpha^{q_{\alpha,p}} \| \, q_{\alpha,p} | p\right)}}{e^{(1-\alpha)2\sqrt{2\pi e}[0.56\frac{a_3}{a_2^2} + \sqrt{a_2}]}} n^{-\frac{1}{2\alpha}} e^{-nD_1\left(W_\alpha^{q_{\alpha,p}} \| \, W | p\right)}$$

Then (32) follows from the identity $\sum_{m \in \mathcal{M}} q(m \in \Theta) \leq L$, equations (28), (29), (31), and the definition error probability as the average of the conditional error probabilities. □

Remark. Note that the lower bound on $P_{\mathbf{e}}^{m}$ depends on the encoding scheme only through the composition of the code and it is independent of the particular choice of the codeword for the message m.

VI. Discussion

We have confined our analysis to the constant composition codes for the sake of brevity. Nevertheless, similar analyses employing Augustin capacity and center instead of Augustin information and mean, lead to refined SPBs both on channels with certain symmetries and on the additive white Gaussian noise channels with quadratic cost functions. The essential technical challenge in this line of work is the derivation of the refined SPB without any symmetry assumptions.

It is worth noting that the refined SPBs of the form (4) are not always achievable, see for example the binary erasure channels whose optimal error probability decays according to (2). The existence of a general proof of the SPB that can account for the behavior of these channels is not evident to us. We believe the refined SPBs of the form (4) are the best possible bounds for derivations of the SPB relying on the asymptotic behavior of sums of independent random variables.

References

[1] P. Elias, "Coding for two noisy channels," in *Proceedings of Third London Symposium of Information Theory*, (London), pp. 61–74, Butterworth Scientific, 1955.

[2] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *The Bell System Technical Journal*, vol. 38, pp. 611–656, May 1959.

[3] R. Dobrushin, "Asymptotic estimates of the probability of error for transmission of messages over a discrete memoryless communication channel with a symmetric transition probability matrix," *Theory of Probability & Its Applications*, vol. 7, no. 3, pp. 270–300, 1962.

[4] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Transactions on Information Theory*, vol. 11, pp. 3–18, Jan. 1965.

[5] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Information and Control*, vol. 10, no. 1, pp. 65–103, 1967.

[6] E. A. Haroutunian, "Estimates of the error probability exponent for a semicontinuous memoryless channel," *Problems of Information Transmission*, vol. 4, no. 4, pp. 37–48, 1968.

[7] U. Augustin, "Error estimates for low rate codes," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 14, no. 1, pp. 61–88, 1969.

[8] R. G. Gallager, *Information theory and reliable communication*. New York, NY: John Wiley & Sons, Inc., 1968.

[9] U. Augustin, *Noisy Channels*. Habilitation thesis, Universität Erlangen-Nürnberg, 1978. (http://bit.ly/2ID8h7m).

[10] B. Nakiboğlu, "The sphere packing bound via Augustin's method," *IEEE Transactions on Information Theory*, vol. 65, pp. 816–840, Feb 2019. (arXiv:1611.06924 [cs.IT]).

[11] B. Nakiboğlu, "The sphere packing bound for memoryless channels," *arXiv:1804.06372 [cs.IT]*, 2018.

[12] M. Dalai, "Lower bounds on the probability of error for classical and classical-quantum channels," *IEEE Transactions on Information Theory*, vol. 59, pp. 8027–8056, Dec 2013.

[13] M. Dalai and A. Winter, "Constant compositions in the sphere packing bound for classical-quantum channels," *IEEE Transactions on Information Theory*, vol. 63, pp. 5603–5617, Sept 2017.

[14] Y. Altuğ and A. B. Wagner, "Refinement of the sphere packing bound for symmetric channels," in *49th Annual Allerton Conference on Communication, Control, and Computing*, pp. 30–37, Sept 2011.

[15] Y. Altuğ and A. B. Wagner, "Refinement of the sphere-packing bound: Asymmetric channels," *IEEE Transactions on Information Theory*, vol. 60, pp. 1592–1614, March 2014.

[16] H. C. Cheng, M. H. Hsieh, and M. Tomamichel, "Quantum sphere-packing bounds with polynomial prefactors," *IEEE Transactions on Information Theory*, vol. 65, pp. 2872–2898, May 2019. (arXiv:1704.05703 [quant-ph]).

[17] R. R. Bahadur and R. R. Rao, "On deviations of the sample mean," *The Annals of Mathematical Statistics*, vol. 31, pp. 1015–1027, 12 1960.

[18] B. Nakiboğlu, "A simple derivation of the refined SPB under certain symmetry hypotheses," *(in preparation)*, 2019.

[19] G. Vazquez-Vilar, A. G. i. Fabregas, T. Koch, and A. Lancho, "Saddlepoint approximation of the error probability of binary hypothesis testing," in *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 2306–2310, June 2018.

[20] J. L. Jensen, *Saddlepoint approximations*. New York: Oxford University Press, 1995.

[21] T. v. Erven and P. Harremoës, "Rényi divergence and Kullback-Leibler divergence," *IEEE Transactions on Information Theory*, vol. 60, pp. 3797–3820, July 2014.

[22] I. G. Shevtsova, "An improvement of convergence rate estimates in the Lyapunov theorem," *Doklady Mathematics*, vol. 82, no. 3, pp. 862–864, 2010.

[23] I. Csiszár and G. Longo, "On the error exponent for source coding and for testing simple statistical hypotheses," *Studia Scientiarum Mathematicarum Hungarica*, vol. 6, pp. 181–191, 1971.

[24] V. Strassen, "Asymptotische abschätzungen in Shannons Informationstheorie," in *Trans. Third Prague Conf. Inf. Theory*, pp. 689–723, 1962. (http://www.math.cornell.edu/~pmlut/strassen.pdf).

[25] C.-G. Esseen, "Fourier analysis of distribution functions. a mathematical study of the laplace-gaussian law," *Acta Mathematica*, vol. 77, pp. 1–125, 1945.

[26] B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*. Cambridge, MA: Addison-Wesley, 1954.

[27] B. Nakiboğlu, "The Augustin capacity and center," *arXiv:1803.07937 [cs.IT]*, 2018.

[28] W. Rudin, *Principles of Mathematical Analysis*. New York, NY: McGraw-Hill, 1976.