

## A simple derivation of the refined sphere packing bound under certain symmetry hypotheses

Barış NAKİBOĞLU\* 

Department of Electrical and Electronics Engineering, Faculty of Engineering,  
Middle East Technical University, Ankara, Turkey

Received: 28.12.2019

Accepted/Published Online: 05.04.2020

Final Version: 08.05.2020

**Abstract:** A judicious application of the Berry-Esseen theorem via suitable Augustin information measures is demonstrated to be sufficient for deriving the sphere packing bound with a prefactor that is  $\Omega\left(n^{-0.5(1-E'_{sp}(R))}\right)$  for all codes on certain families of channels (including the Gaussian channels and the nonstationary Rényi symmetric channels) and for the constant composition codes on stationary memoryless channels. The resulting nonasymptotic bounds have definite approximation error terms. As a preliminary result that might be of interest on its own, the trade-off between type I and type II error probabilities in the hypothesis testing problem with (possibly non-stationary) independent samples is determined up to some multiplicative constants, assuming that the probabilities of both types of error are decaying exponentially with the number of samples, using the Berry-Esseen theorem.

**Key words:** Hypothesis testing, Berry-Esseen theorem, sphere packing bound, Augustin information measures, constant composition codes, Rényi symmetry, Gaussian channels

### 1. 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 the information theory. For certain channels and for certain values of the rate, sharp bounds were found early on. Elias in [19] for the binary symmetric channel, Shannon \* in [43] for the additive Gaussian noise channel, and Dobrushin in [16] for the strongly symmetric channels (see the original publication in Russian to avoid typos present in the translation) proved that

$$P_e^{(n)} = \Theta\left(n^{-\frac{1-E'_{sp}(R)}{2}} e^{-nE_{sp}(R)}\right) \quad \forall R \in [R_{crit}, C), \quad (1.1)$$

where  $\dagger a_n = \Theta(b_n)$  iff  $0 < \liminf_{n \rightarrow \infty} \left| \frac{a_n}{b_n} \right| \leq \limsup_{n \rightarrow \infty} \left| \frac{a_n}{b_n} \right| < \infty$ ,  $E_{sp}(\cdot)$  is the sphere packing exponent of the channel,  $E'_{sp}(\cdot)$  is its derivative with respect to the rate,  $R_{crit}$  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. On the other

\*Correspondence: bnakib@metu.edu.tr

2010 AMS Mathematics Subject Classification: 94A02

\*The equivalence of the bounds in [43] to (1.1) is proved in Appendix C.

$\dagger$ In this section, we suppress the dependence of the sphere packing exponent to the channel in our notation and denote it by  $E_{sp}(R)$ , rather than  $E_{sp}(R, W)$ .



hand, Elias proved in [19] for the binary erasure channels that

$$P_e^{(n)} = \Theta \left( n^{-\frac{1}{2}} e^{-n E_{sp}(R)} \right) \quad \forall R \in [R_{crit}, C). \tag{1.2}$$

Neither (1.1), nor (1.2), holds for rates below the critical rate. If, however, we replace the equality sign with the greater than or equal to sign, then both (1.1) and (1.2) 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 original derivations.

Derivations of the SPB in [16, 19, 43] relied on the geometric structure of the output space of the channel and parameters that can be defined only for some models. The resulting bounds were expressed in terms of these parameters, as well. Thus, it was not even clear that SPBs in [16, 19, 43] can be interpreted as specific instances of a general bound. The evidence for such an interpretation 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 [23] unified and generalized the upper bounds on the error probability, at least in terms of the exponent, in all the previous studies. It is only with Gallager’s formulation in [23] that one can express the bounds in [16, 19, 43] as (1.1) and (1.2).

The first complete proof of the SPB for arbitrary discrete stationary product channels \* (DSPCs) was presented in [44]. According to [44, Thm. 2]

$$P_e^{(n)} \geq e^{-n [E_{sp}(R - O(n^{-1/2})) + O(n^{-1/2})]} \quad \forall R \in (0, C),$$

where  $a_n = O(b_n)$  iff there exists a  $K \in \mathbb{R}_+$  such that  $|a_n| \leq K b_n$  for all  $n$  large enough. In the following two years, the SPB was proved first for stationary product channels with finite input sets in [26] and then for (possibly) nonstationary product channels in [5]. Since then, the SPB has been proven for various channel models in [4, 6, 9, 13–15, 18, 24, 32–34, 38–41, 50], including some quantum information theoretic ones. However, it is worth noting that a general proof that holds for both Gaussian channels (considered in [18, 41, 43]) and for arbitrary DSPCs (considered in [44]) was absent until recently, see [33] and [34]. These later works on the SPB, i.e., [4–6, 9, 13–15, 18, 24, 26, 32–34, 38–41, 44, 50], were primarily interested in establishing the right exponential decay rate; thus, they were content with prefactors of the form  $e^{-o(n)}$ , where  $a_n = o(b_n)$  iff for all  $\epsilon > 0$  the inequality  $|a_n| \leq \epsilon b_n$  holds for all  $n$  large enough. Some authors did obtain prefactors of the form  $e^{-O(\sqrt{n})}$  or  $e^{-O(\ln n)}$ , but obtaining the best possible –if not tight– prefactor was not an actual concern.

The quest for deriving SPBs with tight prefactors was put on the map again by Altuğ and Wagner in [1] and [2]. According to [1, Thm. 1] for any DSPC with a Gallager symmetric † probability transition matrix  $W$  with positive entries and rate  $R$  in  $(0, C)$ , there exists an  $A \in \mathbb{R}_+$  such that for any  $\epsilon > 0$

$$P_e^{(n)} \geq A n^{-\frac{1-(1+\epsilon)E'_{sp}(R)}{2}} e^{-n E_{sp}(R)} \quad \forall n \geq n_0$$

for some  $n_0$  determined by  $W$ ,  $R$ , and  $\epsilon$ . The corresponding result was established for the constant composition codes on arbitrary DSPCs in [2, Thm. 1]. These results are generalized to classical-quantum channels in [10, Thms. 8 and 14], with a slight improvement, allowing  $\epsilon = 0$  for the symmetric channels.

\*These channels are customarily called discrete memoryless channels, i.e., DMCs. We call them DSPCs in order to underline the stationarity of these channels and the absence of any constraints on their input sets. In principle, such constraints might exist and stationarity might be absent in a discrete channel that is memoryless.

†The condition for Gallager symmetry is described in [24, p. 94]. The binary symmetric channel, the binary erasure channel, and channels considered in [16] are symmetric according to this definition.

The primary tool for the derivations in [1, 2, 10] is the Berry-Esseen theorem, albeit through certain auxiliary results inspired by a theorem of Bahadur and Rao [7], i.e., [1, (74)], [2, Proposition 5], and [10, Thm. 17]. Our main aim in this paper is to demonstrate that the analysis can be simplified and the results can be strengthened and generalized through a more judicious application of the Berry-Esseen theorem via suitable Augustin information measures.

[19] and [43] not only established (1.1) and (1.2) but also obtained closed-form expressions for the upper and lower bounds implicit in (1.1) and (1.2). Dobrushin went one step further and calculated the exact asymptotic behavior of the SPB and the random coding bound by analyzing asymptotic behavior for the lattice and nonlattice cases separately for random variables used to derive the SPB and the random coding bound, see [16, (1.32), (1.33), (1.34)]. Recently, the saddle point approximation is used to derive the SPB with the same asymptotic prefactor [48, Corollary 2], under weaker symmetry hypothesis, \* albeit by assuming a common support for all output distributions of the channel and a nonlattice structure for the random variables involved. † The main drawback of the analysis in [48] is the technical conditions that need to be confirmed for applying the saddle point approximation via [29, Proposition 2.3.1].

**Remark 1.1** *The proof of [48, Corollary 2] holds only for channels whose Augustin center does not change with the order, i.e., for  $W$ 's for which  $\exists q$  such that  $q_{\alpha, W} = q$  for all  $\alpha \in (0, 1)$ . Note that for the channels that violate this additional hypothesis, the  $O(n^{-1/2})$  approximation error terms in [48, (30)] are  $\rho$  dependent because of the implicit  $Q$  dependence of the  $O(n^{-1/2})$  approximation error terms in [48, (12) and (13)]. In order to recover a result similar to [48, Corollary 2] for a channel whose Augustin center changes with the order, one needs a saddle point approximation that holds for a parametric family of i.i.d. sequences of random variables, such as [30, Proposition 1], rather than [48, Lemma 2], which holds for a single i.i.d sequence of random variables.*

Let us finish this section with an overview of the paper. In Section 2, we describe our model and notation. In Section 3, first, we recall the connection between the hypothesis testing problem and the tilting, then we derive our primary technical tool using the Berry-Esseen theorem. In Section 4, we review Augustin's information measures and the sphere packing exponent. In Section 5, we state and prove refined SPBs for various models using Lemma 3.4 and the observations recalled in Section 4. We conclude our presentation with a brief discussion of the results, recent developments, and future work in Section 6.

## 2. Model and notation

For any set  $\mathcal{X}$ , we denote the set of all probability mass functions that are nonzero 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 of a measurable function  $f$  under the probability measure  $\mu$  by  $\mathbf{E}_{\mu}[f]$ . Similarly, we denote the variance of  $f$  under  $\mu$ , i.e.,  $\mathbf{E}_{\mu}[(f - \mathbf{E}_{\mu}[f])^2]$ , by  $\mathbf{V}_{\mu}[f]$ .

For sets  $\mathcal{X}_1, \dots, \mathcal{X}_n$  we denote their Cartesian product by  $\mathcal{X}_1^n$  and for  $\sigma$ -algebras  $\mathcal{Y}_1, \dots, \mathcal{Y}_n$  we denote their product by  $\mathcal{Y}_1^n$ . We use the symbol  $\otimes$  to denote the product of measures.

---

\*The binary input Gaussian channel and the binary erasure channel satisfy the symmetry hypothesis of [48], but not that of [16].

†Neither of these assumptions was needed while deriving this result in [16].

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

$$W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y}).$$

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

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

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, the type) of each  $x_1^n$  by  $\Upsilon(x_1^n)$ ; thus  $\Upsilon(x_1^n) \in \mathcal{P}(\mathcal{X})$ .

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

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

Given an  $(M, L)$  channel code  $(\Psi, \Theta)$  on  $W_{[1,n]}$ , the *conditional error probability*  $P_e^m$  for  $m \in \mathcal{M}$  and the *average error probability*  $P_e$  are defined as

$$P_e^m \triangleq \mathbf{E}_{W_{[1,n]}(\Psi(m))} [\mathbb{1}_{\{m \notin \Theta(\mathcal{Y}_1^n)\}}],$$

$$P_e \triangleq \frac{1}{M} \sum_{m \in \mathcal{M}} P_e^m.$$

An encoding function  $\Psi$  –hence the corresponding code– on a stationary product channel, satisfies an empirical distribution constraint  $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$  iff the composition of all of its codewords are in  $\mathcal{A}$ , i.e., iff  $\Upsilon(\Psi(m)) \in \mathcal{A}$  for all  $m \in \mathcal{M}$ . 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  $m \in \mathcal{M}$ .

### 3. Hypothesis testing problem and Berry-Esseen theorem

Our primary aim in this section is to characterize –up to a multiplicative constant– the asymptotic behavior of type I error probability with the number of samples for a hypothesis testing problem between product measures, when type II error probability is decaying exponentially. We use the Berry-Esseen theorem via the concepts of Rényi divergence and tilted probability measure to do that. First, let us recall the definitions of the Rényi divergence and the tilted probability measure.

**Definition 3.1** For any  $\alpha \in \mathbb{R}_+$  and  $w, q \in \mathcal{P}(\mathcal{Y})$ , the order  $\alpha$  Rényi divergence between  $w$  and  $q$  is

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

where  $\nu$  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:

$$(1 - \alpha)D_\alpha(w \| q) = \inf_{v \in \mathcal{P}(\mathcal{Y})} \alpha D_1(v \| w) + (1 - \alpha)D_1(v \| q) \tag{3.1}$$

with the convention that  $\alpha D_1(v \| w) + (1 - \alpha)D_1(v \| q) = \infty$  if it would be otherwise undefined, see [20, Thm. 30]. The characterization given in (3.1) is related to another key concept of our analysis: the tilted probability measure.

**Definition 3.2** For any  $\alpha \in \mathbb{R}_+$  and  $w, q \in \mathcal{P}(\mathcal{Y})$  satisfying  $D_\alpha(w \| q) < \infty$ , the order  $\alpha$  tilted probability measure  $w_\alpha^q$  is

$$\frac{dw_\alpha^q}{dv} \triangleq e^{(1-\alpha)D_\alpha(w \| q)} \left(\frac{dw}{dv}\right)^\alpha \left(\frac{dq}{dv}\right)^{1-\alpha}.$$

If either  $\alpha$  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 (3.1) by [20, Thm. 30], i.e.,

$$(1 - \alpha)D_\alpha(w \| q) = \alpha D_1(w_\alpha^q \| w) + (1 - \alpha)D_1(w_\alpha^q \| q). \tag{3.2}$$

Furthermore, under the same hypothesis the identities

$$\ln \frac{dw_\alpha^q}{dq} - D_1(w_\alpha^q \| q) = \alpha \left( \ln \frac{dw_{ac}}{dq} - \mathbf{E}_{w_\alpha^q} \left[ \ln \frac{dw_{ac}}{dq} \right] \right), \tag{3.3}$$

$$\ln \frac{dw_\alpha^q}{dw} - D_1(w_\alpha^q \| w) = (\alpha - 1) \left( \ln \frac{dw_{ac}}{dq} - \mathbf{E}_{w_\alpha^q} \left[ \ln \frac{dw_{ac}}{dq} \right] \right), \tag{3.4}$$

hold  $w_\alpha^q$ -a.s., where  $w_{ac}$  is the component of  $w$  that is absolutely continuous in  $q$ .

Let us proceed with recalling the Berry-Esseen theorem.

**Lemma 3.3** ([8, 21, 45]) Let  $\{\xi_t\}_{t \in \mathbb{Z}_+}$  be independent zero mean random variables satisfying  $\sum_{t=1}^n \mathbf{E}[\xi_t^2] < \infty$ . Then there exists an absolute constant  $\omega \leq 0.5600$  satisfying

$$\left| \mathbf{P} \left[ \sum_{t=1}^n \xi_t < \tau \right] - \Phi \left( \frac{\tau}{\sqrt{a_2 n}} \right) \right| \leq \omega \frac{a_3}{a_2 \sqrt{a_2 n}} \quad \forall \tau \in \mathbb{R},$$

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

Lemma 3.4, in the following, characterizes the trade-off between type I and type II error probabilities for a hypothesis testing problem with independent samples, assuming that both error probabilities are decaying at least exponentially with the number of samples. Lemma 3.4, which is derived using the Berry-Esseen theorem, can be interpreted as a refinement of [44, Thm. 5], which is derived using Chebyshev’s inequality.

**Lemma 3.4** 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  $a_2, a_3$ , and  $\Delta$  be

$$a_2 \triangleq \frac{1}{n} \sum_{t=1}^n \mathbf{E}_{w_\alpha^q} \left[ \left( \ln \frac{dw_{t,ac}}{dq_t} - \mathbf{E}_{w_\alpha^q} \left[ \ln \frac{dw_{t,ac}}{dq_t} \right] \right)^2 \right],$$

$$a_3 \triangleq \frac{1}{n} \sum_{t=1}^n \mathbf{E}_{w_\alpha^q} \left[ \left| \ln \frac{dw_{t,ac}}{dq_t} - \mathbf{E}_{w_\alpha^q} \left[ \ln \frac{dw_{t,ac}}{dq_t} \right] \right|^3 \right],$$

$$\Delta \triangleq \exp \left( 2\sqrt{2\pi}e \left( 0.56 \frac{a_3}{a_2} + \sqrt{a_2} \right) \right),$$

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  $\beta \geq n^{-1/2} e^{-\alpha\sqrt{a_2 n}}$ , satisfying  $q(\mathcal{E}) \leq \beta e^{-D_1(w_\alpha^q \| q)}$ , we have

$$w(\mathcal{Y}_1^n \setminus \mathcal{E}) \geq \Delta^{\alpha-1} \beta^{\frac{\alpha-1}{\alpha}} n^{-\frac{1}{2\alpha}} e^{-D_1(w_\alpha^q \| w)} \tag{3.5}$$

provided that  $\beta \leq \Delta^{-\alpha} n^{-1/2} e^{\alpha\sqrt{a_2 n}}$ . Furthermore, for any  $\alpha \in (0, 1)$  and  $\beta \in \mathbb{R}_+$ , there exists an event  $\mathcal{E} \in \mathcal{Y}_1^n$  such that

$$q(\mathcal{E}) \leq \beta e^{-D_1(w_\alpha^q \| q)}, \tag{3.6}$$

$$w(\mathcal{Y}_1^n \setminus \mathcal{E}) \leq \left[ \frac{1\sqrt{8\pi a_2}}{4\pi a_2} \ln \Delta \right]^{\frac{1}{\alpha}} \frac{(\alpha\beta)^{\frac{\alpha-1}{\alpha}}}{1-\alpha} n^{-\frac{1}{2\alpha}} e^{-D_1(w_\alpha^q \| w)}. \tag{3.7}$$

**Proof** [Proof of Lemma 3.4] Let the random variables  $\xi_t$  and  $\xi$  and the event  $\mathcal{B}$  be

$$\xi_t \triangleq \ln \frac{dw_{t,ac}}{dq_t}, \quad \xi \triangleq \sum_{t=1}^n \xi_t, \quad \mathcal{B} \triangleq \{y_1^n : \tau_0 \leq \xi - \mathbf{E}_{w_\alpha^q}[\xi] \leq \tau_1\}.$$

Thus  $\xi = \ln \frac{dw_{ac}}{dq}$  holds  $q$ -a.s. by the definitions of  $\xi_t$  and  $\xi$ . Hence (3.3), (3.4), and the definition of  $\mathcal{B}$  imply that

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

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

$$w_\alpha^q(\mathcal{E} \cap \mathcal{B}) \leq q(\mathcal{E}) e^{D_1(w_\alpha^q \| q) + \alpha\tau_1}, \tag{3.8}$$

$$w(\mathcal{Y}_1^n \setminus \mathcal{E}) \geq w_\alpha^q(\mathcal{B} \setminus \mathcal{E}) e^{-D_1(w_\alpha^q \| w) + (1-\alpha)\tau_0}. \tag{3.9}$$

On the other hand,  $\xi_t$ 's are jointly independent under the tilted probability measure  $w_\alpha^q$ . Thus the Berry-Esseen theorem, given in Lemma 3.3, implies

$$\begin{aligned} w_\alpha^q(\mathcal{B}) &\geq \Phi\left(\frac{\tau_1}{\sqrt{a_2 n}}\right) - \Phi\left(\frac{\tau_0}{\sqrt{a_2 n}}\right) - 2\frac{0.56}{\sqrt{n}} \frac{a_3}{a_2\sqrt{a_2}} \\ &= \frac{1}{\sqrt{2\pi}} \int_{\frac{\tau_0}{\sqrt{a_2 n}}}^{\frac{\tau_1}{\sqrt{a_2 n}}} e^{-z^2/2} dz - 2\frac{0.56}{\sqrt{n}} \frac{a_3}{a_2\sqrt{a_2}} \\ &\geq \frac{1}{\sqrt{2\pi}} e^{-\frac{(\tau_0|\vee|\tau_1|)^2}{2na_2}} \frac{\tau_1 - \tau_0}{\sqrt{a_2 n}} - 2\frac{0.56}{\sqrt{n}} \frac{a_3}{a_2\sqrt{a_2}}. \end{aligned}$$

If we set  $\tau_0 = -\frac{2\ln\beta + \ln n}{2\alpha} - \ln \Delta$  and  $\tau_1 = -\frac{2\ln\beta + \ln n}{2\alpha}$ , then  $-\sqrt{a_2 n} \leq \tau_0 \leq \tau_1 \leq \sqrt{a_2 n}$  by the hypothesis and

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

Furthermore,  $w_\alpha^q(\mathcal{E} \cap \mathcal{B}) \leq \frac{1}{\sqrt{n}}$  as a result of (3.8),  $e^{\alpha\tau_1} = \frac{1}{\beta\sqrt{n}}$ , and the hypothesis  $q(\mathcal{E}) \leq \beta e^{-D_1(w_\alpha^q \| q)}$ . Thus  $w_\alpha^q(\mathcal{B} \setminus \mathcal{E}) \geq \frac{1}{\sqrt{n}}$ . Then using (3.9) and  $e^{(1-\alpha)\tau_0} = \beta^{\frac{\alpha-1}{\alpha}} n^{\frac{\alpha-1}{2\alpha}} \Delta^{\alpha-1}$  we get

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

**Remark 3.5** While deriving bounds similar to (3.5), the constants  $\tau_0$  and  $\tau_1$  are usually assumed to satisfy  $\tau_0 = -\tau_1$ , see for example [44, Thm. 5] or [10, Thm. 11]. Such a choice, however, does not lead to tight bounds in our case.

To establish the existence of an event satisfying both (3.6) and (3.7), let us consider the event  $\mathcal{E}$  given in the following

$$\mathcal{E} \triangleq \{y_1^n : \xi \geq \mathbf{E}_{w_\alpha^q}[\xi] + \gamma\} \cup \left\{y_1^n : \frac{dq}{d\nu} = 0 \ \& \ \frac{dw}{d\nu} > 0\right\}, \tag{3.10}$$

where  $\gamma$  is a real number to be determined later and  $\nu$  is any measure satisfying both  $w \prec \nu$  and  $q \prec \nu$ .

**Remark 3.6** The random variable  $\xi$  is defined only for  $y_1^n$  's with a positive  $\frac{dq}{d\nu}$ . Thus one can define  $\xi$  to be infinite for  $y_1^n$  's satisfying both  $\frac{dq}{d\nu} = 0$  and  $\frac{dw}{d\nu} > 0$ , and define the event  $\mathcal{E}$  to be the event that  $\xi$  is greater than or equal to  $\mathbf{E}_{w_\alpha^q}[\xi] + \gamma$ .

For the event  $\mathcal{E}$  defined in (3.10), as a result of (3.3) we have

$$\begin{aligned} q(\mathcal{E}) &= \mathbf{E}_q \left[ \mathbf{1}_{\{\xi \geq \mathbf{E}_{w_\alpha^q}[\xi] + \gamma\}} \right] \\ &= e^{-D_1(w_\alpha^q \| q)} \mathbf{E}_{w_\alpha^q} \left[ \mathbf{1}_{\{\xi \geq \mathbf{E}_{w_\alpha^q}[\xi] + \gamma\}} e^{-\alpha(\xi - \mathbf{E}_{w_\alpha^q}[\xi])} \right] \\ &\leq e^{-D_1(w_\alpha^q \| q) - \alpha\gamma} \sum_{\kappa=0}^{\infty} w_\alpha^q(\mathcal{E}_\kappa) e^{-\alpha\kappa}, \end{aligned} \tag{3.11}$$

where the event  $\mathcal{E}_\kappa$  is defined for each  $\kappa \in \mathbb{Z}$  to be

$$\mathcal{E}_\kappa \triangleq \{y_1^n : \gamma + \kappa \leq \xi - \mathbf{E}_{w_\alpha^q}[\xi] < \gamma + \kappa + 1\}.$$

On the other hand, we can bound  $w_\alpha^q(\mathcal{E}_\kappa)$  uniformly for all integers  $\kappa$  using the Berry-Esseen theorem, i.e., Lemma 3.3, as follows

$$\begin{aligned} w_\alpha^q(\mathcal{E}_\kappa) &\leq \Phi\left(\frac{\gamma + \kappa + 1}{\sqrt{a_2 n}}\right) - \Phi\left(\frac{\gamma + \kappa}{\sqrt{a_2 n}}\right) + 2\frac{0.56}{\sqrt{n}} \frac{a_3}{a_2 \sqrt{a_2}} \\ &\leq \frac{1}{\sqrt{n}} \left( \frac{1}{\sqrt{2\pi a_2}} + 2\frac{0.56 a_3}{a_2 \sqrt{a_2}} \right). \end{aligned} \tag{3.12}$$

For  $\gamma = \frac{1}{\alpha} \ln \left[ \frac{1}{\sqrt{n}} \left( \frac{1}{\sqrt{2\pi a_2}} + 2\frac{0.56 a_3}{a_2 \sqrt{a_2}} \right) \frac{\beta^{-1}}{1 - e^{-\alpha}} \right]$ , (3.6) follows from (3.11), (3.12), and  $\sum_{\kappa=0}^{\infty} e^{-\alpha\kappa} = \frac{1}{1 - e^{-\alpha}}$ .

$w(\mathcal{Y}_1^n \setminus \mathcal{E})$  is bounded following a similar analysis, by invoking (3.4), instead of (3.3):

$$\begin{aligned} w(\mathcal{Y}_1^n \setminus \mathcal{E}) &= \mathbf{E}_w \left[ \mathbf{1}_{\{\xi < \mathbf{E}_{w_\alpha^q}[\xi] + \gamma\}} \right] \\ &= e^{-D_1(w_\alpha^q \| w)} \mathbf{E}_{w_\alpha^q} \left[ \mathbf{1}_{\{\xi < \mathbf{E}_{w_\alpha^q}[\xi] + \gamma\}} e^{(1-\alpha)(\xi - \mathbf{E}_{w_\alpha^q}[\xi])} \right] \\ &\leq e^{-D_1(w_\alpha^q \| w) + (1-\alpha)\gamma} \sum_{\kappa=-1}^{-\infty} w_\alpha^q(\mathcal{E}_\kappa) e^{(\kappa+1)(1-\alpha)} \\ &\leq e^{-D_1(w_\alpha^q \| w) + (1-\alpha)\gamma} \frac{1}{1 - e^{-\alpha-1}} \sup_{\kappa \in \mathbb{Z}_-} w_\alpha^q(\mathcal{E}_\kappa). \end{aligned}$$

Invoking first (3.12) and then  $\frac{1}{1-e^{\alpha-1}}(1-e^{-\alpha})^{\frac{\alpha-1}{\alpha}} \leq \frac{1}{1-\alpha} \alpha^{\frac{\alpha-1}{\alpha}} e^{\frac{1}{2\alpha}}$  we get,

$$\begin{aligned} w(\mathcal{Y}_1^n \setminus \mathcal{E}) &\leq e^{-D_1(w_\alpha^q \| w)} \beta^{\frac{\alpha-1}{\alpha}} \frac{1}{1-e^{\alpha-1}} (1-e^{-\alpha})^{\frac{\alpha-1}{\alpha}} \left[ \frac{1}{\sqrt{n}} \left( \frac{1}{\sqrt{2\pi a_2}} + 2 \frac{0.56 a_3}{a_2 \sqrt{a_2}} \right) \right]^{\frac{1}{\alpha}} \\ &\leq e^{-D_1(w_\alpha^q \| w)} \frac{1}{1-\alpha} (\beta \alpha)^{\frac{\alpha-1}{\alpha}} \left[ \frac{2\sqrt{e}}{\sqrt{n a_2}} \left( \frac{1}{\sqrt{8\pi}} + \frac{0.56 a_3}{a_2} \right) \right]^{\frac{1}{\alpha}}. \end{aligned}$$

Then (3.7) follows from  $\left( \frac{1}{\sqrt{8\pi}} + \frac{0.56 a_3}{a_2} \right) \leq \frac{1\sqrt{8\pi a_2}}{8\pi\sqrt{a_2 e}} \ln \Delta$ . □

Lemma 3.4 characterizes the asymptotic behavior of the trade-off between the optimal type I and type II error probabilities for a hypothesis testing problem with independent samples:  $P_{e_{II}}^{(n)}$  is  $\Theta\left(n^{-\frac{1}{2\alpha}} e^{-D_1(w_\alpha^q \| w)}\right)$  whenever  $P_{e_I}^{(n)}$  is  $\Theta\left(e^{-D_1(w_\alpha^q \| q)}\right)$ . For the stationary case, i.e., when  $w_t = w_1$ ,  $q_t = q_1$  for all  $t$ , Csiszár and Longo [12] described how (3.3) and (3.4) can be used together with an earlier result by Strassen, [46, Thm. 1.1], to characterize the exact asymptotic behavior of  $P_{e_{II}}^{(n)}$  for the case when  $P_{e_I}^{(n)} = e^{-D_1(w_\alpha^q \| q)}$ , i.e.,  $P_{e_{II}}^{(n)} = (K + o(1)) n^{-\frac{1}{2\alpha}} e^{-D_1(w_\alpha^q \| w)}$ , with some minor inaccuracies discussed in Remark 3.7. One does not need to rely on [46, Thm. 1.1] of Strassen to characterize this exact asymptotic behavior. The Berry-Esseen theorem, however, is not sufficient for determining the value of the constant  $K$ . In order to determine the constant  $K$ , one needs to invoke either finer characterizations of the asymptotic behavior of sums of independent random variables (such as the ones in [22, §IV.2, §IV.3], [25, §42, §43]) or apply other techniques (such as the saddle point approximation described in [29, Prop. 2.3.1]). It is worth noting that both of these approaches require hypotheses stronger than that of the Berry-Esseen theorem. The situation is similar for other values of  $\alpha$ , but of no interest for our discussion of the sphere packing bound.

**Remark 3.7** *We believe the approach of [12] is sound. Its calculations, however, seem to have some mistakes. Repeating the calculations as described in [12], we recover the second line of [12, (33)] as  $\ln \frac{\alpha^*}{1-\alpha^*} - \frac{\ln S_1 \sqrt{2\pi}}{\alpha^*} + o(1)$ . With this modification [12, Thm. 2] is consistent with the intimately related results about the SPB proved earlier [16, (1.32), (1.33)] and since then [48, (38)].*

#### 4. Augustin’s information measure and the sphere packing exponent

The ultimate aim of 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’s information measures and review their elementary properties.

##### 4.1. Augustin’s information measures

Let us start by recalling the definition of the conditional Rényi divergence.

**Definition 4.1** *For any  $\alpha \in \mathbb{R}_+$ ,  $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ ,  $q \in \mathcal{P}(\mathcal{Y})$ , and  $p \in \mathcal{P}(\mathcal{X})$  the order  $\alpha$  conditional Rényi divergence for the input distribution  $p$  is*

$$D_\alpha(W \| q | p) \triangleq \sum_{x \in \mathcal{X}} p(x) D_\alpha(W(x) \| q).$$



**Definition 4.2** For any  $\alpha \in \mathbb{R}_+$ ,  $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ , and  $p \in \mathcal{P}(\mathcal{X})$  the order  $\alpha$  Augustin information for the input distribution  $p$  is

$$I_\alpha(p; W) \triangleq \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(W \| q | p).$$

The infimum is achieved by a unique probability measure  $^* q_{\alpha,p}$ , called the order  $\alpha$  Augustin mean for the input distribution  $p$ , by [35, Lemma 13]. Furthermore,

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

for all  $q \in \mathcal{P}(\mathcal{Y})$  by [35, Lemma 13], as well.

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

$$\frac{\partial}{\partial \alpha} I_\alpha(p; W) = \begin{cases} \frac{1}{(\alpha-1)^2} D_1(W_\alpha^{q_{\alpha,p}} \| W | p) & \alpha \neq 1 \\ \sum_x \frac{p(x)}{2} \mathbf{V}_{W(x)} \left[ \ln \frac{dW(x)}{dq_{1,p}} \right] & \alpha = 1 \end{cases} \quad (4.1)$$

by [35, Lemma 17-(e)], where  $W_\alpha^{q_{\alpha,p}}$  is the tilted channel defined as follows.

**Definition 4.3** For any  $\alpha \in \mathbb{R}_+$ ,  $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$  and  $q \in \mathcal{P}(\mathcal{Y})$ , the order  $\alpha$  tilted channel  $W_\alpha^q$  is a function from  $\{x : D_\alpha(W(x) \| q) < \infty\}$  to  $\mathcal{P}(\mathcal{Y})$  given by

$$\frac{dW_\alpha^q(x)}{d\nu} \triangleq e^{(1-\alpha)D_\alpha(W(x) \| q)} \left( \frac{dW(x)}{d\nu} \right)^\alpha \left( \frac{dq}{d\nu} \right)^{1-\alpha}. \quad (4.2)$$

The tilted channel can be used to express  $I_\alpha(p; W)$  in terms of the Kullback-Leibler divergences using  $^\dagger$  [35, Lemma 13-(e)]:

$$I_\alpha(p; W) = \frac{\alpha}{1-\alpha} D_1(W_\alpha^{q_{\alpha,p}} \| W | p) + D_1(W_\alpha^{q_{\alpha,p}} \| q_{\alpha,p} | p). \quad (4.3)$$

Furthermore, the Augustin mean satisfies

$$\sum_x p(x) W_\alpha^{q_{\alpha,p}}(x) = q_{\alpha,p} \quad (4.4)$$

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 [35, Lemma 13], where  $q_{1,p} = \sum_x p(x) W(x)$ . Thus for all  $\alpha \in \mathbb{R}_+$  we have

$$D_1(W_\alpha^{q_{\alpha,p}} \| q_{\alpha,p} | p) = I_1(p; W_\alpha^{q_{\alpha,p}}). \quad (4.5)$$

**Definition 4.4** For any  $\alpha \in \mathbb{R}_+$ ,  $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ , and  $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$ , the order  $\alpha$  Augustin capacity of  $W$  for the constraint set  $\mathcal{A}$  is

$$C_{\alpha,W,\mathcal{A}} \triangleq \sup_{p \in \mathcal{A}} I_\alpha(p; W). \quad (4.6)$$

When the constraint set  $\mathcal{A}$  is the whole  $\mathcal{P}(\mathcal{X})$ , we denote the order  $\alpha$  Augustin capacity by  $C_{\alpha,W}$ , i.e.,  $C_{\alpha,W} \triangleq C_{\alpha,W,\mathcal{P}(\mathcal{X})}$ .

\*We refrain from including the channel symbol  $W$  in the symbol for the Augustin mean because the channel will be clear from the context.

$^\dagger$ It is worth noting that (4.3) follows from (3.2) and  $I_\alpha(p; W) = D_\alpha(W \| q_{\alpha,p} | p)$  for  $\alpha$  values in  $(0, 1)$ .

Using the definitions of the Augustin information and capacity, we obtain the following expression for the latter

$$C_{\alpha, W, \mathcal{A}} = \sup_{p \in \mathcal{A}} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_{\alpha}(W \| q | p). \tag{4.7}$$

If  $\mathcal{A}$  is convex, then the order of the supremum and the infimum can be changed as a result of [35, Thm. 1]:

$$\sup_{p \in \mathcal{A}} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_{\alpha}(W \| q | p) = \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{p \in \mathcal{A}} D_{\alpha}(W \| q | p). \tag{4.8}$$

If in addition  $C_{\alpha, W, \mathcal{A}}$  is finite, then there exists a unique probability measure  $q_{\alpha, W, \mathcal{A}}$ , called the order  $\alpha$  Augustin center of  $W$  for the constraint set  $\mathcal{A}$ , satisfying

$$C_{\alpha, W, \mathcal{A}} = \sup_{p \in \mathcal{A}} D_{\alpha}(W \| q_{\alpha, W, \mathcal{A}} | p) \tag{4.9}$$

by [35, Thm. 1].

We denote the set of all probability mass functions satisfying a cost constraint  $\varrho$  by  $\mathcal{A}(\varrho)$ , i.e.

$$\mathcal{A}(\varrho) \triangleq \{p \in \mathcal{P}(\mathcal{X}) : \mathbf{E}_p[\rho] \leq \varrho\}. \tag{4.10}$$

With a slight abuse of notation, we denote the cost-constrained Augustin capacity  $C_{\alpha, W, \mathcal{A}(\varrho)}$  by  $C_{\alpha, W, \varrho}$ , as well. A more detailed discussion of Augustin’s information measures can be found in [35].

### 4.2. The sphere packing exponent

**Definition 4.5** For any  $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ ,  $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$ , and  $R \in \mathbb{R}_+$ , the sphere packing exponent (SPE) is

$$E_{sp}(R, W, \mathcal{A}) \triangleq \sup_{\alpha \in (0,1)} \frac{1-\alpha}{\alpha} (C_{\alpha, W, \mathcal{A}} - R). \tag{4.11}$$

When the constraint set  $\mathcal{A}$  is the whole  $\mathcal{P}(\mathcal{X})$ , we denote SPE by  $E_{sp}(R, W)$ , i.e.  $E_{sp}(R, W) \triangleq E_{sp}(R, W, \mathcal{P}(\mathcal{X}))$ .

With a slight abuse of notation, we denote SPE for  $\mathcal{A}(\varrho)$  by  $E_{sp}(R, W, \varrho)$  and SPE for  $\mathcal{A} = \{p\}$  case by  $E_{sp}(R, W, p)$ . Note that as a result of definitions of Augustin capacity and the SPE we have

$$E_{sp}(R, W, \mathcal{A}) = \sup_{p \in \mathcal{A}} E_{sp}(R, W, p). \tag{4.12}$$

Furthermore, since  $I_{\alpha}(p; W)$  is continuously differentiable in  $\alpha$  by [35, Lemma 17-(e)], we can apply the derivative test to find the optimal  $\alpha$  in (4.11) for  $\mathcal{A} = \{p\}$  case: using (4.1) and (4.3) we get

$$\frac{\partial}{\partial \alpha} \frac{1-\alpha}{\alpha} (I_{\alpha}(p; W) - R) = \frac{1}{\alpha^2} (R - D_1(W_{\alpha}^{q_{\alpha,p}} \| q_{\alpha,p} | p)). \tag{4.13}$$

On the other hand, either  $D_1(W_{\alpha}^{q_{\alpha,p}} \| q_{\alpha,p} | p) = I_1(p; W)$  for all positive orders  $\alpha$ , or  $D_1(W_{\alpha}^{q_{\alpha,p}} \| q_{\alpha,p} | p)$  is increasing and continuous in the order  $\alpha$  on  $\mathbb{R}_+$  by [35, Lemma 17-(f)]. Furthermore,  $D_1(W_1^{q_{1,p}} \| q_{1,p} | p)$  is equal to  $I_1(p; W)$  by definition and  $\lim_{\alpha \downarrow 0} D_1(W_{\alpha}^{q_{\alpha,p}} \| q_{\alpha,p} | p)$  is equal to  $\lim_{\alpha \downarrow 0} I_{\alpha}(p; W)$  by (4.5) and [35, Lemma 17-(g)]. Thus for any rate  $R$  in  $(\lim_{\alpha \downarrow 0} I_{\alpha}(p; W), I_1(p; W))$  there exists an order  $\alpha^* \in (0, 1)$  satisfying

$$R = D_1(W_{\alpha^*}^{q_{\alpha^*,p}} \| q_{\alpha^*,p} | p) \tag{4.14}$$

by the intermediate value theorem [42, 4.23]. The order  $\alpha^*$  satisfying (4.14) is unique because  $D_1(W_{\alpha}^{q_{\alpha,p}} \| q_{\alpha,p} | p)$  is increasing in  $\alpha$ . The monotonicity of  $D_1(W_{\alpha}^{q_{\alpha,p}} \| q_{\alpha,p} | p)$  in  $\alpha$  and (4.13) also imply  $E_{sp}(R, W, p) = \frac{1-\alpha^*}{\alpha^*} (I_{\alpha^*}(p; W) - R)$ . Thus as a result of (4.3), the unique  $\alpha^*$  satisfying (4.14) also satisfies

$$E_{sp}(R, W, p) = D_1(W_{\alpha^*}^{q_{\alpha^*,p}} \| W | p). \tag{4.15}$$

Since  $D_1(W_{\alpha}^{q_{\alpha,p}} \parallel q_{\alpha,p} | p)$  is continuous and increasing in  $\alpha$ , its inverse is increasing and continuous, as well. Thus the definition of SPE given in (4.11) 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 order  $\alpha^*$  satisfying (4.14) also satisfies

$$\frac{\partial}{\partial R} E_{sp}(R, W, p) = \frac{\alpha^* - 1}{\alpha^*}, \tag{4.16}$$

as was established in [34, Lemma 2].

### 5. The refined sphere packing bound

In this section, we consider the channel coding problem for various channel models and derive lower bounds to the error probability of the following form

$$P_e^{(n)} \geq A n^{\frac{E'_{sp}(R)-1}{2}} e^{-nE_{sp}(R)} \quad \forall n \geq n_0. \tag{5.1}$$

for constants  $A$  and  $n_0$  determined by the rate, the channel, and the constraints on the codes, if there exist any. Following [1, 2, 10], we call these bounds refined sphere packing bounds (refined SPBs) because of their resemblance to the standard SPBs, e.g. [4–6, 9, 13–15, 18, 24, 26, 32–34, 38–41, 44, 50], establishing \*

$$P_e^{(n)} \geq e^{-nE_{sp}(R)-o(n)} \quad \forall n \geq n_0. \tag{5.2}$$

The refined SPBs that we state and prove in this section are not formally particular cases of a general proposition. Nevertheless, they are all consequences of Lemma 3.4 and the properties of Augustin’s information measures.

We establish a refined SPB for the constant composition codes in Subsection 5.1, for codes on (possibly) nonstationary Rényi symmetric channels in Subsection 5.2, and for codes on additive white Gaussian noise channels with quadratic cost functions in Subsection 5.3.

#### 5.1. Constant composition codes

**Theorem 5.1** *For any  $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ ,  $M, L, n \in \mathbb{Z}_+$ ,  $p \in \mathcal{P}(\mathcal{X})$  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  $\alpha^* \triangleq \frac{1}{1-E'_{sp}(\frac{1}{n} \ln \frac{M}{L}, W, p)}$  satisfies*

$$D_1(W_{\alpha^*}^{q_{\alpha^*,p}} \parallel q_{\alpha^*,p} | p) = \frac{1}{n} \ln \frac{M}{L}. \tag{5.3}$$

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

$$P_e^{(n)} \geq \Delta^{\alpha^* - 1} (4n)^{-\frac{1}{2\alpha^*}} e^{-nE_{sp}(\frac{1}{n} \ln \frac{M}{L}, W, p)} \tag{5.4}$$

provided that  $\sqrt{a_2 n} - \frac{\ln 4n}{2\alpha^*} \geq \ln \Delta$  where

$$\begin{aligned} a_2 &\triangleq \mathbf{E}_{p \otimes W_{\alpha^*}^{q_{\alpha^*,p}}} \left[ \left| \ln \frac{dW}{dq_{\alpha^*,p}} - \mathbf{E}_{W_{\alpha^*}^{q_{\alpha^*,p}}} \left[ \ln \frac{dW}{dq_{\alpha^*,p}} \right] \right|^2 \right], \\ a_3 &\triangleq \mathbf{E}_{p \otimes W_{\alpha^*}^{q_{\alpha^*,p}}} \left[ \left| \ln \frac{dW}{dq_{\alpha^*,p}} - \mathbf{E}_{W_{\alpha^*}^{q_{\alpha^*,p}}} \left[ \ln \frac{dW}{dq_{\alpha^*,p}} \right] \right|^3 \right], \\ \Delta &\triangleq \exp \left( 2\sqrt{2\pi} e \left( 0.56 \frac{a_3}{a_2} + \sqrt{a_2} \right) \right). \end{aligned}$$

---

\*The approximation error terms in standard SPBs are usually  $O(\sqrt{n})$  or  $O(\ln n)$ , rather than just  $o(n)$ .

Although Theorem 5.1 itself is composition-dependent, it implies (via appropriate worst-case assumptions) composition-independent bounds, such as [2, Thm 1]. Similar composition-dependent [10, Proposition 13] and composition independent [10, Thm 8] bounds have, recently, been derived for classical-quantum channels using an approach similar to that of [2]. The primary advantages of Theorem 5.1 over the previous results are the conceptual simplicity and brevity of its proof and its definite approximation error terms.

**Proof** [Proof of Theorem 5.1] The existence of a unique order  $\alpha^*$  satisfying (5.3) was proved, and its value was determined in Section 4, see (4.14), (4.15), and (4.16).

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

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

Then  $v_m$  is equal to the order  $\alpha^*$  tilted probability measure between  $w_m$  and  $q$ . Furthermore, the empirical distribution of the all of the codewords, i.e., all  $\Psi(m)$ 's, are equal to  $p$  by the hypothesis; thus we have

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

On the other hand,  $\sum_{m \in \mathcal{M}} q(m \in \Theta) \leq L$  by the definition of list decoding. Thus, at least half of the messages in  $\mathcal{M}$  (at least  $\lfloor \frac{M+1}{2} \rfloor$  of the messages in  $\mathcal{M}$  to be precise) will satisfy  $q(m \in \Theta) \leq \frac{2L}{M}$  as a result of Markov's inequality. Applying Lemma 3.4 with  $\mathcal{E} = \{y_1^n : m \in \Theta(y_1^n)\}$  and  $\beta = 2$  for the messages satisfying  $q(m \in \Theta) \leq \frac{2L}{M}$ , we get

$$P_e^m \geq \Delta^{\alpha^* - 1} 2^{\frac{\alpha^* - 1}{\alpha^*} n} n^{-\frac{1}{2\alpha^*}} e^{-nD_1(W_{\alpha^*}^{q_{\alpha^*,p}} \| W | p)}$$

as long as  $\sqrt{a_2 n} - \frac{\ln 4n}{2\alpha^*} \geq \ln \Delta$ . Then (5.4) follows from (4.14), (4.15), (5.3), and the definition error probability as the average of the conditional error probabilities of the messages.  $\square$

### 5.2. Codes on Rényi symmetric channels

**Definition 5.2** A channel  $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$  satisfying  $W \prec \nu$  for some  $\nu \in \mathcal{P}(\mathcal{Y})$  is Rényi symmetric iff for each  $\alpha \in \mathbb{R}_+$  with finite  $C_{\alpha,W}$  there exists a function  $G_\alpha^W(\cdot) : \mathbb{R} \rightarrow [0, 1]$  satisfying

$$W \left( \left\{ \frac{dW(x)}{d\nu} \leq e^s \frac{dq_{\alpha,W}}{d\nu} \right\} | x \right) = G_\alpha^W(s) \quad \forall x \in \mathcal{X}, s \in \mathbb{R}. \tag{5.5}$$

**Remark 5.3** If  $W$  is Rényi symmetric, then the identity  $\lim_{s \downarrow -\infty} G_\alpha^W(s) = 0$  holds whenever  $C_{\alpha,W}$  is finite. On the other hand, the identity  $\lim_{s \uparrow \infty} G_\alpha^W(s) = 1$  is violated whenever  $W(x) \not\prec q_{\alpha,W}$ , which can only happen for  $\alpha$ 's in  $(0, 1)$ . Such a Rényi symmetric  $W$  is obtained by removing  $w^{v,1}$  from  $W_i$  described in [36, Example 2] and the resulting  $G_\alpha^W$  is given by  $G_\alpha^W(s) = (1/2)\mathbb{1}_{\{s \geq \ln 1/2\}}$ .

The Rényi symmetry holds for all input symmetric channels described in [49, Definition 3.2] and for all the Gallager symmetric channels described in [24, p. 94], see Appendix A. Recall that the Gallager symmetry holds for all strongly symmetric (Dobrushin symmetric) channels, which is described in [16]. The binary symmetric

channel is strongly symmetric. The binary erasure channel is Gallager symmetric but not strongly symmetric. The binary input Gaussian channel is input symmetric according but not Gallager symmetric. The Rayleigh fading channel with per coherence interval power constraint analyzed in [31, (3)] is Rényi symmetric, by [31, (7) and (10)], but not input symmetric, see [31, (5)].

**Remark 5.4** *The input symmetry described in [49, Definition 3.2] can be generalized by relying on a compact group with the associated Haar measure, rather than a finite additive group with the uniform distribution. The Rayleigh fading channel with per coherence interval power constraint analyzed in [31] is input symmetric for this more general definition. The covariant channels analyzed by Holevo in [28], can be seen as the counterparts of [49, Definition 3.2] and its generalization in the framework of Quantum Information Theory.*

The derivation of the refined SPB for the Rényi symmetric channels is analogous to the derivation of the refined SPB for the constant composition codes. Lemma 5.5, given in the following, is used in lieu of (4.14), (4.15), (4.16) in the latter derivation.

**Lemma 5.5** *For any Rényi symmetric channel  $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$  with finite  $C_{1,W}$  and rate  $R$  in  $(\lim_{\alpha \downarrow 0} C_{\alpha,W}, C_{1,W})$  there exists an order  $\alpha^* \in (0, 1)$  such that*

$$R = D_1(W_{\alpha^*}^{q_{\alpha^*},W}(x) \parallel q_{\alpha^*},W) \quad \forall x \in \mathcal{X}, \tag{5.6}$$

$$E_{sp}(R, W) = D_1(W_{\alpha^*}^{q_{\alpha^*},W}(x) \parallel W(x)) \quad \forall x \in \mathcal{X}. \tag{5.7}$$

Furthermore, if either  $W_{\alpha^*}^{q_{\alpha^*},W} \left( \left\{ \frac{dW(x)}{d\nu} = \gamma \frac{dq_{\alpha^*},W}{d\nu} \right\} \middle| x \right) < 1$  for all  $\gamma \in \mathbb{R}_+$  or  $q_{\alpha,W} = q$  for all  $\alpha \in (0, 1]$ , then

$$\frac{\partial}{\partial R} E_{sp}(R, W) = \frac{\alpha^* - 1}{\alpha^*}. \tag{5.8}$$

Lemma 5.5 is proved in Appendix B. Proving the essential assertions of Lemma 5.5 for input symmetric channels, however, is considerably easier: for any input symmetric channel  $W$  and the uniform probability mass function  $u$  on its input set,  $C_{\alpha,W} = I_{\alpha}(u; W)$  and  $q_{\alpha,W} = q_{\alpha,u}$  for all  $\alpha \in \mathbb{R}_+$ . Consequently, the identities given in (5.6), (5.7), and (5.8) are nothing but the identities given in (4.14), (4.15), and (4.16) for  $p = u$  case because the Kullback-Leibler divergences on the right-hand-sides of (5.6) and (5.7) have the same value for all  $x$  by the symmetry. Hence, (5.8) holds for any input symmetric channels satisfying  $\lim_{\alpha \downarrow 0} C_{\alpha,W} < C_{1,W}$  by (4.16), as well.

**Remark 5.6** *If  $\frac{dW(x)}{dq_{\alpha,u}} = \gamma$  holds  $W(x)$ -a.s. for all  $x$  for a  $(\gamma, \alpha)$  pair for an input symmetric  $W$ , then  $q_{\alpha,u} = \sum_x u(x) W_{\eta}^{q_{\alpha,u}}(x)$  for all  $\eta$ . Thus  $q_{\eta,u} = q_{\alpha,u}$  and  $I_{\eta}(u; W) = \ln \gamma$  for all  $\eta \in \mathbb{R}_+$  by [35, Lemma 13]. Thus such a  $(\gamma, \alpha)$  pair does not exist for input symmetric  $W$ 's satisfying  $\lim_{\alpha \downarrow 0} C_{\alpha,W} \neq C_{1,W}$ .*

**Remark 5.7** *Lemma 5.5 is stated under the finite  $C_{1,W}$  hypothesis, yet it holds under the weaker hypothesis  $\lim_{\alpha \uparrow} \frac{1-\alpha}{\alpha} C_{\alpha,W}$ , as well. However, establishing Lemma 5.5 under this weaker hypothesis would require us to introduce the concepts of power mean, Rényi information, and compactness in the topology of setwise convergence, see [36, Lemma 24-(d)].*

**Theorem 5.8** Let  $W_t : \mathcal{X}_t \rightarrow \mathcal{P}(\mathcal{Y}_t)$  be a Rényi symmetric channel with finite  $C_{1, W_t}$  for all  $t \in \mathbb{Z}_+$  and  $n, M, L \in \mathbb{Z}_+$  satisfy  $\lim_{\alpha \downarrow 0} C_{\alpha, W_{[1, n]}} < \ln \frac{M}{L} < C_{1, W_{[1, n]}}$ . Then there exists an order  $\alpha^* \in (0, 1)$  satisfying

$$D_1(U_{\alpha^*}(x_1^n) \| q_{\alpha^*, W_{[1, n]}}) = \ln \frac{M}{L} \quad \forall x_1^n \in \mathcal{X}_1^n \quad (5.9)$$

where  $U_\alpha \triangleq \{W_{[1, n]}\}_\alpha^{q_{\alpha, W_{[1, n]}}}$  for all  $\alpha \in (0, 1)$ . Furthermore any  $(M, L)$  channel code on  $W_{[1, n]}$  satisfies

$$P_e^{(n)} \geq \Delta^{\alpha^* - 1} (4n)^{-\frac{1}{2\alpha^*}} e^{-E_{sp}(\ln \frac{M}{L}, W_{[1, n]})} \quad (5.10)$$

provided that  $\sqrt{a_2 n} - \frac{\ln 4n}{2\alpha^*} \geq \ln \Delta$  where

$$\begin{aligned} a_2 &\triangleq \frac{1}{n} \sum_{t=1}^n \mathbf{E}_{U_{\alpha^*}(x_1^n)} \left[ \left( \xi_{\alpha^*}^t(x_t) - \mathbf{E}_{U_{\alpha^*}(x_1^n)} [\xi_{\alpha^*}^t(x_t)] \right)^2 \right], \\ a_3 &\triangleq \frac{1}{n} \sum_{t=1}^n \mathbf{E}_{U_{\alpha^*}(x_1^n)} \left[ \left| \xi_{\alpha^*}^t(x_t) - \mathbf{E}_{U_{\alpha^*}(x_1^n)} [\xi_{\alpha^*}^t(x_t)] \right|^3 \right], \\ \Delta &\triangleq \exp \left( 2\sqrt{2\pi e} \left( 0.56 \frac{a_3}{a_2} + \sqrt{a_2} \right) \right), \end{aligned}$$

and  $\xi_\alpha^t(x_t) = \ln \frac{dW_t(x_t)}{d\nu_t} - \ln \frac{dq_{\alpha, W_t}}{d\nu_t}$  for all  $\alpha \in (0, 1)$ .

Furthermore, if  $\{W_t\}_{\alpha^*}^{q_{\alpha^*, W_t}} \left( \left\{ \frac{dW_t(x_t)}{d\nu_t} = \gamma \frac{dq_{\alpha^*, W_t}}{d\nu_t} \right\} \middle| x_t \right) < 1$  for all  $\gamma \in \mathbb{R}_+$  for some  $t$  or  $q_{\alpha, W_{[1, n]}} = q$  for all  $\alpha \in (0, 1]$ , then  $\alpha^* = \frac{1}{1 - E'_{sp}(\ln \frac{M}{L}, W_{[1, n]})}$ .

Note that if any of the component channels, i.e., any of the  $W_t$ 's, is an input symmetric channel satisfying  $\lim_{\alpha \downarrow 0} C_{\alpha, W_t} < C_{1, W_t}$ , then  $\alpha^* = \frac{1}{1 - E'_{sp}(\ln \frac{M}{L}, W_{[1, n]})}$  holds as a result of Remark 5.6.

Theorem 5.8 does not assume the channel to be stationary, i.e., it holds even when  $W_t$ 's are not identical. To the best of our knowledge, refined sphere packing bounds have only been reported for stationary channels before, even in the case of symmetric channels considered in [16, (1.28)], [1, Thm. 1], [10, Thm. 14], [48, Corollary 2], [30, Thm. 4], [31, (36), (37b)], [3, Thm. 1].

For the stationary input symmetric channels Theorem 5.8 is tight both in terms of exponent and prefactor for rates above the critical rate, provided that channel is not singular. For the case of the singular stationary input symmetric channels, Altuğ and Wagner [3] have recently reported a sharper result, which generalizes Elias's result in [19] for the binary erasure channels. In order to obtain such results, however, merely plugging in bounds on binary hypothesis testing is not enough, see Section 6 for a more detailed discussion.

**Remark 5.9** Theorem 5.8 is derived using Lemma 3.4, which is stated for the product measures. Lemma 3.4, however, holds for any  $w$  and  $q$  for which  $\xi_t$ 's are independent random variables under  $w_\alpha^q$ . This condition is satisfied by the output distributions and the Augustin centers on the product channels with feedback, i.e., by  $W_{\overrightarrow{[1, n]}}(x)$  and  $q_{\alpha, W_{\overrightarrow{[1, n]}}}$ , provided that the component channels are Rényi symmetric. Thus Theorem 5.8 holds not just for codes on the product channel  $W_{[1, n]}$  but also for codes on the product channels with feedback  $W_{\overrightarrow{[1, n]}}$ . Similar observations have been used to establish the SPB on product channels with feedback in [3, 6, 17, 27, 39]. The formal definition of the product channels with feedback and the proof of the SPB on these channels without the symmetry assumptions can be found in [6, 38, 39].

**Proof** [Proof of Theorem 5.8] The Rényi symmetry and (4.9) imply

$$C_{\alpha, W_t} = D_{\alpha}(W_t(x_t) \| q_{\alpha, W_t}) \quad \forall x_t \in \mathcal{X}_t, t \in \mathbb{Z}_+. \quad (5.11)$$

On the other hand, [35, Lemma 26] implies

$$C_{\alpha, W_{[1,n]}} = \sum_{t=1}^n C_{\alpha, W_t}, \quad (5.12)$$

$$q_{\alpha, W_{[1,n]}} = \bigotimes_{t=1}^n q_{\alpha, W_t}. \quad (5.13)$$

Then, the product structure  $W_{[1,n]}(x_1^n) = \otimes_{t=1}^n W_t(x_t)$  and the Rényi symmetry of  $W_t$ 's imply the Rényi symmetry of  $W_{[1,n]}$ . In particular

$$g_{\alpha}^{W_{[1,n]}} = g_{\alpha}^{W_1} \circledast g_{\alpha}^{W_2} \circledast \dots \circledast g_{\alpha}^{W_n} \quad (5.14)$$

where  $\circledast$  denotes the convolution and  $g$ 's are the density functions of the corresponding  $G$ 's, i.e.,  $g$ 's and  $G$ 's are uniquely determined by each other via the following relation

$$G_{\alpha}^W(s) = \int_{-\infty}^s g_{\alpha}^W(z) dz.$$

Since  $W_{[1,n]}$  is Rényi symmetric the existence of an order  $\alpha^* \in (0, 1)$  satisfying (5.9) follows from (5.6) of Lemma 5.5.

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

$$w_m \triangleq \bigotimes_{t=1}^n W_t(\Psi_t(m)), \quad q \triangleq \bigotimes_{t=1}^n q_{\alpha^*, W_t}, \quad v_m \triangleq \bigotimes_{t=1}^n \{W_t\}_{\alpha^*}^{q_{\alpha^*, W_t}}(\Psi_t(m)).$$

Note that  $w_m = W_{[1,n]}(\Psi(m))$  by definition,  $q = q_{\alpha^*, W_{[1,n]}}$  by (5.13), and  $v_m$  is equal to the order  $\alpha^*$  tilted probability measure between  $w_m$  and  $q$ , which is equal to  $U_{\alpha^*}(\Psi(m))$ , by construction. Then Lemma 5.5 implies

$$\begin{aligned} D_1(v_m \| q) &= \ln \frac{M}{L} & m \in \mathcal{M}, \\ D_1(v_m \| w_m) &= E_{sp}(\ln \frac{M}{L}, W_{[1,n]}) & m \in \mathcal{M}. \end{aligned}$$

On the other hand,  $\sum_{m \in \mathcal{M}} q(m \in \Theta) \leq L$  by the definition of list decoding. Thus, at least half of the messages in  $\mathcal{M}$  (at least  $\lfloor \frac{M+1}{2} \rfloor$  of the messages in  $\mathcal{M}$  to be precise) will satisfy  $q(m \in \Theta) \leq \frac{2L}{M}$  as a result of Markov's inequality. Applying Lemma 3.4 with  $\mathcal{E} = \{y_1^n : m \in \Theta(y_1^n)\}$  and  $\beta = 2$  for the messages satisfying  $q(m \in \Theta) \leq \frac{2L}{M}$ , we get

$$P_e^m \geq \Delta^{\alpha^* - 1} 2^{\frac{\alpha^* - 1}{\alpha^*}} n^{-\frac{1}{2\alpha^*}} e^{-E_{sp}(\ln \frac{M}{L}, W_{[1,n]})}$$

provided that  $\sqrt{a_2 n} - \frac{\ln 4n}{2\alpha^*} \geq \ln \Delta$ . Then (5.10) follows from the definition  $P_e$  as the average of  $P_e^m$ .

Note that as a result of (5.14),  $g_{\alpha^*}^{W_{[1,n]}}$  is a Dirac delta function iff all  $g_{\alpha^*}^{W_t}$ 's are. This observation together with Lemma 5.5 implies the sufficient condition for  $\alpha^*$  to be  $\frac{1}{1 - E'_{sp}(\ln \frac{M}{L}, W_{[1,n]})}$ .  $\square$

**5.3. Codes on additive white Gaussian noise channels**

The additive white Gaussian noise channel with noise variance  $\sigma^2$  is described via the following transition probability

$$W(\mathcal{E}|x) = \int_{\mathcal{E}} \varphi_{\sigma^2}(y - x) dy \quad \forall \mathcal{E} \in \mathcal{B}(\mathbb{R}) \tag{5.15}$$

where  $\varphi_{\sigma^2}$  is the zero mean Gaussian probability density function of variance  $\sigma^2$ , i.e.,

$$\varphi_{\sigma^2}(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2\sigma^2}} \quad \forall z \in \mathbb{R}.$$

With a slight abuse of notation, we denote the corresponding probability measure, i.e., zero mean Gaussian probability measure of variance  $\sigma^2$ , by  $\varphi_{\sigma^2}$ , as well.

If the cost function is the quadratic one, then the zero mean Gaussian distribution is the maximizer for the Augustin information among all input distributions satisfying the cost constraint (for any positive order) see [35, Example 4], i.e.

$$\begin{aligned} C_{\alpha, W, \varrho} &= I_{\alpha}(\varphi_{\varrho}; W) \\ &= \sup_{p: \mathbf{E}_p[x^2] \leq \varrho} I_{\alpha}(p; W) \end{aligned} \quad \forall \alpha \in \mathbb{R}_+.$$

Furthermore, the order  $\alpha$  Augustin center of this channel is a zero-mean Gaussian probability measure. The closed-form expression for the Augustin capacity and Augustin center are derived in [35, (126), (127), (128)]:

$$C_{\alpha, W, \varrho} = \begin{cases} \frac{\alpha \varrho}{2(\alpha \theta_{\alpha} + (1-\alpha)\sigma^2)} + \frac{1}{\alpha-1} \ln \frac{\theta_{\alpha}^{\alpha/2} \sigma^{1-\alpha}}{\sqrt{\alpha \theta_{\alpha} + (1-\alpha)\sigma^2}} & \alpha \neq 1 \\ \frac{1}{2} \ln \left( 1 + \frac{\varrho}{\sigma^2} \right) & \alpha = 1 \end{cases}, \tag{5.16}$$

$$q_{\alpha, W, \varrho} = \varphi_{\theta_{\alpha}}, \tag{5.17}$$

$$\theta_{\alpha} \triangleq \sigma^2 + \frac{\varrho}{2} - \frac{\sigma^2}{2\alpha} + \sqrt{\left(\frac{\varrho}{2} - \frac{\sigma^2}{2\alpha}\right)^2 + \varrho \sigma^2}. \tag{5.18}$$

One can confirm the following identity by substitution

$$\begin{aligned} \alpha \varrho \theta_{\alpha} &= \alpha(\theta_{\alpha})^2 - (2\alpha - 1)\theta_{\alpha}\sigma^2 + (\alpha - 1)\sigma^4 \\ &= (\theta_{\alpha} - \sigma^2)(\alpha \theta_{\alpha} + (1 - \alpha)\sigma^2). \end{aligned} \tag{5.19}$$

In fact,  $\theta_{\alpha}$  is a root of the equality given above because of a fixed point property similar to the one described in (4.4), see [35, (132), (133)] and the ensuing discussion.

The sphere packing exponent expression resulting from (5.16), (5.17), (5.18), and (5.19), is derived in [34, Example 1]. It is given by the following parametric form in [34, (68), (69)]:

$$R = \frac{1}{2} \ln \frac{\alpha \theta_{\alpha} + (1-\alpha)\sigma^2}{\sigma^2}, \tag{5.20}$$

$$E_{sp}(R, W, \varrho) = \frac{(1-\alpha)\varrho}{2(\alpha \theta_{\alpha} + (1-\alpha)\sigma^2)} + \frac{1}{2} \ln \frac{\alpha \theta_{\alpha} + (1-\alpha)\sigma^2}{\theta_{\alpha}}. \tag{5.21}$$

Using (5.18) and (5.20), one can express the unique  $\alpha^*$  whose rate is  $R$  as a function of  $R$ , as well:

$$\alpha^* \triangleq \frac{e^{2R}-1}{2} \left( \sqrt{1 + \frac{4\sigma^2}{\varrho} \frac{e^{2R}}{e^{2R}-1}} - 1 \right). \tag{5.22}$$



The equivalence of the parametric form given in (5.20) and (5.21) to the expression given by Gallager [24, (7.4.33)] can be confirmed by substitution using (5.22). One can also confirm using (5.18) and (5.19) in (5.20) and (5.21) that

$$\frac{\partial}{\partial R} E_{sp}(R, W, \varrho) = \frac{\alpha^* - 1}{\alpha^*}. \tag{5.23}$$

Furthermore, codes on additive white Gaussian noise channels satisfy both (1.1) and (5.1) as a result of Shannon’s [43, (3)]; this is established in Appendix C, for completeness. Theorem 5.10, which is establishing the refined SPB given in (5.1), is proved using principles and analysis similar to those used in the proofs of Theorems 5.1 and 5.8, which are quite different from the ones employed in [43].

**Theorem 5.10** *Let  $\sigma$  and  $\varrho$  be positive reals,  $n$ ,  $M$ , and  $L$  be positive integers satisfying  $R \in \left(0, \frac{1}{2} \ln \frac{\sigma^2 + \varrho}{\sigma^2}\right)$  for  $R = \frac{1}{n} \ln \frac{M}{L}$ , and  $W_t$  be an additive white Gaussian noise channel with noise variance  $\sigma^2$ , say  $W$ , for all  $t$ . Then any  $(M, L)$  channel code  $(\Psi, \Theta)$  on  $W_{[1, n]}$  satisfying  $\sum_{t=1}^n (\Psi_t(m))^2 = n\varrho$  for all messages  $m$  in the message set  $\mathcal{M}$  satisfies*

$$P_e^{(n)} \geq \Delta^{\alpha^* - 1} (4n)^{-\frac{1}{2\alpha^*}} e^{-nE_{sp}(R, W, \varrho)} \tag{5.24}$$

for  $\alpha^*$  given in (5.22) provided that  $\sqrt{a_2 n} - \frac{\ln n}{2\alpha^*} \geq \Delta$ , where

$$\begin{aligned} a_2 &\triangleq \frac{(\theta_{\alpha^*} - \sigma^2)^2}{2(\alpha^* \theta_{\alpha^*} + (1 - \alpha^*)\sigma^2)^2} + \frac{\sigma^2 \theta_{\alpha^*} \varrho}{(\alpha^* \theta_{\alpha^*} + (1 - \alpha^*)\sigma^2)^3}, \\ a_3 &\triangleq \frac{18(\theta_{\alpha^*} - \sigma^2)^3}{(\alpha^* \theta_{\alpha^*} + (1 - \alpha^*)\sigma^2)^3} + \frac{18(\sqrt{2/\pi})\sigma^3 \theta_{\alpha^*}^{3/2} \varrho^{3/2}}{(\alpha^* \theta_{\alpha^*} + (1 - \alpha^*)\sigma^2)^{4.5}}, \\ \Delta &\triangleq \exp\left(2\sqrt{2\pi}e\left(0.56\frac{a_3}{a_2} + \sqrt{a_2}\right)\right). \end{aligned}$$

Before proving of Theorem 5.10, let us briefly discuss its implications. Theorem 5.10 bounds the performance of codes satisfying an equality cost constraint, but it can also be used to bound the performance of codes satisfying an inequality cost constraint. In particular, Shannon has observed in [43, (83)] that

$$\widetilde{P}_e(n, \varrho) \geq P_e(n, \varrho) \geq \widetilde{P}_e(n + 1, \varrho), \tag{5.25}$$

where  $P_e(n, \varrho)$  is the infimum of error probabilities of  $(M, L)$  channel codes satisfying  $\sum_{t=1}^n (\Psi_t(m))^2 \leq n\varrho$  and  $\widetilde{P}_e(n, \varrho)$  is the analogous quantity for the constraint  $\sum_{t=1}^n (\Psi_t(m))^2 = n\varrho$ . The first inequality of (5.25) holds because any code satisfying the equality constraint also satisfies the inequality constraint. The second inequality of (5.25) is confirmed by considering an extension of codewords by one additional symbol,  $\Psi_{n+1}(m)$ , so as to satisfy  $\sum_{t=1}^{n+1} (\Psi_t(m))^2 = (n + 1)\varrho$ . Recently, Vazquez-Vilar have improved (5.25) in [47, Proposition 1] by observing that the same extension can be constructed for the constraint  $\sum_{t=1}^{n+1} (\Psi_t(m))^2 = n\varrho$ , as well. Thus, we have

$$\widetilde{P}_e(n, \varrho) \geq P_e(n, \varrho) \geq \widetilde{P}_e\left(n + 1, \frac{n}{n+1}\varrho\right). \tag{5.26}$$

One can use Theorem 5.10 together with either (5.25) or (5.26) to determine prefactor for codes satisfying the cost constraint with an inequality. For that let us first note that  $E_{sp}(R, W, \varrho)$  is convex in the rate  $R$  as a

result of (5.22) and (5.23) because  $\alpha^*$  is increasing monotonically with  $R$  on  $[0, C_{1,W,\varrho}]$ . Then  $E_{sp}(R, W, \varrho)$  lies above its tangent at any point and (5.23) implies

$$E_{sp}(R_1, W, \varrho) \geq E_{sp}(R_0, W, \varrho) + \frac{\alpha^*(R_0)-1}{\alpha^*(R_0)}(R_1 - R_0). \tag{5.27}$$

Applying Theorem 5.10 at  $(n+1)$  –rather than  $n$ – together with (5.25) and invoking (5.27) for  $R_0 = \frac{1}{n+1} \ln \frac{M}{L}$  and  $R_1 = R$ , we get the following bound for  $(M, L)$  channel codes  $(\Psi, \Theta)$  satisfying  $\sum_{t=1}^n (\Psi_t(m))^2 \leq n\varrho$

$$\begin{aligned} P_e^{(n)} &\geq \frac{\Delta^{\alpha_0^*-1}}{(4(n+1))^{1/2\alpha_0^*}} e^{-(n+1)E_{sp}(R_0, W, \varrho)} \\ &\geq \frac{\Delta^{\alpha_0^*-1} e^{-E_{sp}(R, W, \varrho)}}{(8n)^{1/2\alpha_0^*}} e^{-nE_{sp}(R, W, \varrho)} e^{\frac{\alpha_0^*-1}{\alpha_0^*} \frac{nR}{n+1}} \\ &= \frac{\Delta^{\alpha_0^*-1} e^{C_{\alpha_0^*, W, \varrho}(1-\alpha_0^*)/\alpha_0^*}}{8^{1/2\alpha_0^*} n^{(\alpha^* - \alpha_0^*)/2\alpha^* \alpha_0^*}} n^{-\frac{1}{2\alpha^*}} e^{-nE_{sp}(R, W, \varrho)} \end{aligned}$$

where  $\alpha_0^* = \alpha^*(R_0)$  and  $a_2, a_3$ , and  $\Delta$  are calculated at  $\alpha_0^*$ , rather than  $\alpha^*$ , provided that  $\sqrt{a_2(n+1)} - \frac{\ln(n+1)}{2\alpha_0^*} \geq \Delta$ . Note that  $|R - R_0| = \frac{R}{n+1}$  and the function  $\alpha^*$  given in (5.22) is analytical in the rate  $R$ ; hence  $|\alpha_0^* - \alpha^*|$  is  $O(n^{-1})$ . Thus

$$\frac{\Delta^{\alpha_0^*-1} e^{C_{\alpha_0^*, W, \varrho}(1-\alpha_0^*)/\alpha_0^*}}{8^{1/2\alpha_0^*} n^{(\alpha^* - \alpha_0^*)/2\alpha^* \alpha_0^*}} \sim A$$

for some constant  $A \in \mathbb{R}_+$  and (5.1), holds not only for codes satisfying  $\sum_{t=1}^n (\Psi_t(m))^2 = n\varrho$  but also for codes satisfying  $\sum_{t=1}^n (\Psi_t(m))^2 \leq n\varrho$ , for the order  $\alpha^*$  given in (5.22), for some constant  $A \in \mathbb{R}_+$  for  $n$  large enough.

**Proof** [Proof of Theorem 5.10] The following expressions for the order one Rényi divergence and the order  $\alpha$  tilted channel for the zero mean Gaussian distribution of variance  $\theta$  can be confirmed by substitution

$$D_1(W(x) \parallel \varphi_\theta) = \frac{\sigma^2 + x^2 - \theta}{2\theta} + \frac{1}{2} \ln \frac{\theta}{\sigma^2} \tag{5.28}$$

$$W_\alpha^{\varphi_\theta}(\mathcal{E}|x) = \int_{\mathcal{E}} \varphi_{\frac{\sigma^2\theta}{\alpha\theta + (1-\alpha)\sigma^2}} \left( y - \frac{\alpha\theta}{\alpha\theta + (1-\alpha)\sigma^2} x \right) dy \tag{5.29}$$

for all  $x \in \mathbb{R}$ ,  $\theta \in \mathbb{R}_+$ , and  $\mathcal{E} \in \mathcal{B}(\mathbb{R})$ . Then as a result of (5.19), (5.20), (5.21), and (5.22), we have

$$D_1\left(W_{\alpha^*}^{\varphi_{\theta_{\alpha^*}}} \parallel \varphi_{\theta_{\alpha^*}} \mid p\right) = \frac{\alpha^{*2}\theta_{\alpha^*}(\mathbf{E}_p[x^2] - \varrho)}{2(\alpha^*\theta_{\alpha^*} + (1-\alpha^*)\sigma^2)^2} + R, \tag{5.30}$$

$$D_1\left(W_{\alpha^*}^{\varphi_{\theta_{\alpha^*}}} \parallel W \mid p\right) = \frac{(1-\alpha^*)^2\sigma^2(\mathbf{E}_p[x^2] - \varrho)}{2(\alpha^*\theta_{\alpha^*} + (1-\alpha^*)\sigma^2)^2} + E_{sp}(R, W, \varrho), \tag{5.31}$$

where  $R = \frac{1}{n} \ln \frac{M}{L}$ .

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

$$w_m \triangleq \bigotimes_{t=1}^n W(\Psi_t(m)), \quad q \triangleq \bigotimes_{t=1}^n \varphi_{\theta_{\alpha^*}}, \quad v_m \triangleq \bigotimes_{t=1}^n W_{\alpha^*}^{\varphi_{\theta_{\alpha^*}}}(\Psi_t(m)).$$

Then  $v_m$  is the order  $\alpha^*$  tilted probability measure between  $w_m$  and  $q$ ; using the hypothesis  $\sum_{t=1}^n (\Psi_t(m))^2 = n\varrho$  together with (5.30) and (5.31), we get

$$D_1(v_m \parallel q) = \ln \frac{M}{L} \quad m \in \mathcal{M}, \tag{5.32}$$

$$D_1(v_m \parallel w_m) = nE_{sp}\left(\frac{1}{n} \ln \frac{M}{L}, W, \varrho\right) \quad m \in \mathcal{M}. \tag{5.33}$$

In order to obtain (5.24), we will apply Lemma 3.4 to the probability measure pairs  $(w_m, q)$  satisfying  $q(m \in \Theta) \leq \frac{2L}{M}$  for  $\beta = 2$  together with a rotation on  $\mathbb{R}^n$  that minimizes the approximation error terms arising from the absolute third order moments.

For any  $x \in \mathbb{R}$ , let the random variable  $\xi_x$  be

$$\xi_x \triangleq \ln \frac{dW(x)}{d\varphi_{\theta_{\alpha^*}}} - \mathbf{E}_{W_{\alpha^*}^{\varphi_{\theta_{\alpha^*}}}(x)} \left[ \ln \frac{dW(x)}{d\varphi_{\theta_{\alpha^*}}} \right].$$

Then using (5.29), we get

$$\begin{aligned} \xi_x &= \frac{1}{2} \frac{y^2}{\theta_{\alpha^*}} - \frac{(y-x)^2}{2\sigma^2} + \frac{1}{2} \ln \frac{\theta_{\alpha^*}}{\sigma^2} - \frac{1}{2} \mathbf{E}_{W_{\alpha^*}^{\varphi_{\theta_{\alpha^*}}}(x)} \left[ \frac{y^2}{\theta_{\alpha^*}} - \frac{(y-x)^2}{\sigma^2} \right] - \frac{1}{2} \ln \frac{\theta_{\alpha^*}}{\sigma^2} \\ &= \frac{\sigma^2 - \theta_{\alpha^*}}{2\sigma^2 \theta_{\alpha^*}} \left[ \left( y - \frac{\alpha^* \theta_{\alpha^*} x}{\alpha^* \theta_{\alpha^*} + (1-\alpha^*)\sigma^2} \right)^2 - \frac{\sigma^2 \theta_{\alpha^*}}{\alpha^* \theta_{\alpha^*} + (1-\alpha^*)\sigma^2} \right] + \frac{x}{\alpha^* \theta_{\alpha^*} + (1-\alpha^*)\sigma^2} \left( y - \frac{\alpha^* \theta_{\alpha^*} x}{\alpha^* \theta_{\alpha^*} + (1-\alpha^*)\sigma^2} \right). \end{aligned} \tag{5.34}$$

On the other hand, moments and absolute moments of a zero mean Gaussian random variable  $Z$  with variance  $\sigma^2$  satisfy the following identities

$$\begin{aligned} \mathbf{E}[Z^\kappa] &= \mathbb{1}_{\{\frac{\kappa}{2} \in \mathbb{Z}_+\}} \sigma^\kappa (\kappa - 1)!!, \\ \mathbf{E}[|Z|^\kappa] &= \left[ \mathbb{1}_{\{\frac{\kappa}{2} \notin \mathbb{Z}_+\}} \sqrt{\frac{2}{\pi}} + \mathbb{1}_{\{\frac{\kappa}{2} \in \mathbb{Z}_+\}} \right] \sigma^\kappa (\kappa - 1)!!, \end{aligned}$$

where  $\kappa!! = \prod_{l=0}^{\lceil \frac{\kappa}{2} \rceil - 1} (\kappa - 2l)$ . Furthermore, for any three random variables  $Z_1, Z_2,$  and  $Z_3$ , we have \*

$$\mathbf{E}[|Z_1 + Z_2 + Z_3|^3] \leq 9 \left( \mathbf{E}[|Z_1|^3] + \mathbf{E}[|Z_2|^3] + \mathbf{E}[|Z_3|^3] \right). \tag{5.35}$$

Then using (5.34), one can confirm by substitution that

$$\mathbf{E}_{W_{\alpha^*}^{\varphi_{\theta_{\alpha^*}}}(x)} [(\xi_x)^2] = \frac{(\theta_{\alpha^*} - \sigma^2)^2}{2(\alpha^* \theta_{\alpha^*} + (1-\alpha^*)\sigma^2)^2} + \frac{\sigma^2 \theta_{\alpha^*} x^2}{(\alpha^* \theta_{\alpha^*} + (1-\alpha^*)\sigma^2)^3}, \tag{5.36}$$

$$\mathbf{E}_{W_{\alpha^*}^{\varphi_{\theta_{\alpha^*}}}(x)} [|\xi_x|^3] \leq \frac{18(\theta_{\alpha^*} - \sigma^2)^3}{(\alpha^* \theta_{\alpha^*} + (1-\alpha^*)\sigma^2)^3} + \frac{18(\sqrt{2/\pi})\sigma^3 \theta_{\alpha^*}^{3/2} x^3}{(\alpha^* \theta_{\alpha^*} + (1-\alpha^*)\sigma^2)^{4.5}}. \tag{5.37}$$

The hypothesis  $\sum_{t=1}^n (\Psi_t(m))^2 = n\varrho$  implies the  $a_2$  of Lemma 3.4 to be equal to the  $a_2$  of Theorem 5.10. One, however, cannot assert the analogous relation for  $a_3$ 's. Nevertheless, there exists a rotation in  $\mathbb{R}^n$ , say  $\mathcal{S}_m$  such that  $\mathcal{S}_m \Psi(m)$  is equal to vector whose all entries are  $\sqrt{\varrho}$ , i.e.

$$\mathcal{S}_m(\Psi_1(m), \dots, \Psi_n(m)) = (\sqrt{\varrho}, \dots, \sqrt{\varrho}).$$

Note that for  $w_* \triangleq \bigotimes_{t=1}^n W(\sqrt{\varrho})$ , we have

$$\begin{aligned} w_m(m \notin \Theta) &= w_*(m \notin \mathcal{S}_m \Theta), \\ q(m \in \Theta) &= q(m \in \mathcal{S}_m \Theta). \end{aligned}$$

Thus one can apply Lemma 3.4 to the pair  $(w_*, q)$  in order to bound  $P_\epsilon^m$ . For the pair  $(w_*, q)$ , however,  $a_3$  of Lemma 3.4 is equal to the  $a_3$  of Theorem 5.10.

---

\*The inequality given in (5.35) follows from the inequality via the observation that the geometric mean is less than the arithmetic mean. A proof is presented in Appendix D for completeness.

As it was the case for the proofs of Theorems 5.1 and 5.8,  $\sum_{m \in \mathcal{M}} q(m \in \Theta) \leq L$  by the definition of list decoding. Thus, at least half of the messages in  $\mathcal{M}$  will satisfy  $q(m \in \Theta) \leq \frac{2L}{M}$  as a result of Markov's inequality. Applying Lemma 3.4 with  $\mathcal{E} = \{y_1^n : m \in \Theta(y_1^n)\}$  and  $\beta = 2$  for the messages satisfying  $q(m \in \Theta) \leq \frac{2L}{M}$  and using (5.32) and (5.33), we get

$$P_e^m \geq \Delta^{\alpha^* - 1} 2^{\frac{\alpha^* - 1}{\alpha^*}} n^{-\frac{1}{2\alpha^*}} e^{-n E_{sp}(\frac{1}{n} \ln \frac{M}{L}, W, \varrho)}$$

as long as  $\sqrt{a_2 n} - \frac{\ln 4n}{2\alpha^*} \geq \ln \Delta$ . Then (5.24) follows from the definition of  $P_e$  as the average of the conditional error probabilities. □

### 6. Discussion

Theorems 5.1, 5.8, 5.10 establish refined sphere packing bounds, i.e., bounds of the form (5.1), for fixed composition codes on stationary memoryless channels, codes on (possibly) nonstationary Rényi symmetric channels, and cost-constrained codes on additive white Gaussian noise channels with the quadratic cost function. Derivations of Theorems 5.1, 5.8, 5.10 rely on the properties of Augustin's information measures and the application of Berry-Esseen theorem to the hypothesis testing problem summarized in Lemma 3.4. For certain cases including the additive white Gaussian channels [43] and the strongly symmetric channels [16], these bounds are known to be tight in the sense that they can be matched by achievability results asserting the existence of codes satisfying

$$P_e^{(n)} \leq \tilde{A} n^{\frac{E'_{sp}(R) - 1}{2}} e^{-n E_{sp}(R)} \quad \forall n \geq n_0 \tag{6.1}$$

for rates between the critical rate and the capacity of the channel. Recently, Altuğ and Wagner [3, Theorem 1] have generalized the results of [19] and [16] and established (5.1) and (6.1) for all nonsingular Gallager symmetric channels.

At least since [19], it is also known that for the binary erasure channel the polynomial prefactor of (5.1) can be improved from  $n^{(E'_{sp}(R) - 1)/2}$  to  $n^{-1/2}$ . Recently, Altuğ and Wagner proved this result for all singular Gallager symmetric channels, [3, Theorem 2]. Both [19] and [3], however, have refrained from relying on bounds on the performance of the binary hypothesis testing problem with independent samples. This is not surprising because Lemma 3.4 characterizes the prefactor for the binary hypothesis testing problem with independent samples, exactly. Thus the refined SPBs of the form (5.1) are the best possible bounds for derivations of the SPB relying on the asymptotic behavior of sums of independent random variables, notwithstanding their suboptimality for singular Gallager symmetric channels.

As pointed out in Section 3, one can improve Lemma 3.4 and determine not only the prefactor but also the asymptotic constant in the tradeoff between the probabilities of type I and type II errors, either by invoking finer characterizations of the asymptotic behavior of sums of independent random variables or by applying a saddle point approximation. Although such results, e.g. [12, 48], require stronger hypothesis and are more nuanced, \* they are important in the context of binary hypothesis testing. From the standpoint of the channel coding problem, however, it is rather hard to justify the extra effort such an analysis requires. First of all, the corresponding achievability results will have different constants, even when the prefactors match, as observed

---

\*Even the statement of these results are more nuanced because they need to distinguish the lattice and nonlattice cases for the random variables involved in the analysis.

in [16, 19, 43]. More importantly, such refined results on binary hypothesis testing will also suffer from the subtlety discussed in the previous paragraph for the case of the singular channels, i.e., their prefactor will be suboptimal for the singular channels, such as the binary erasure channel.

The principal novelty of this manuscript is the use of the Berry-Essen theorem via suitable Augustin information measures to bound the optimal error probability in the channel coding problem. In this manuscript, our primary focus was the rates below the channel capacity; thus, we have derived refined sphere packing bounds. The same idea can be used to strengthen the strong converse bounds under similar symmetry hypothesis, as it has recently been demonstrated in [11]. The essential technical challenge in this line of work is the derivation of the refined SPBs and the refined strong converses without any symmetry assumptions on the channel or on the codes.

**A. Rényi symmetry is implied by input symmetry and Gallager symmetry**

In the following, we will explain briefly why the Rényi symmetry holds both for all input symmetric channels described in [49, Definition 3.2] and for all Gallager symmetric channels described in [24, p. 94]. Let us start with the input symmetric channels. Let  $u$  be the uniform distribution on the input set of the input symmetric  $W$  and let  $q_\alpha$  be

$$q_\alpha \triangleq \frac{\mu_\alpha}{\|\mu_\alpha\|} \quad \text{where} \quad \frac{d\mu_\alpha}{d\nu} \triangleq \left[ \sum_x u(x) \left( \frac{dW(x)}{d\nu} \right)^\alpha \right]^{\frac{1}{\alpha}}. \tag{A.1}$$

Then the using the input symmetry one can confirm that

$$D_\alpha(W(x) \| q_\alpha) = \frac{\alpha}{\alpha-1} \ln \|\mu_\alpha\| \quad \forall x \in \mathcal{X}. \tag{A.2}$$

The definitions of the tilted channel, (A.1), and (A.2) imply  $q_\alpha = \sum_x u(x) W_\alpha^{q_\alpha}(x)$  and  $q_{1,u} \prec q_\alpha$ . Thus  $q_\alpha$  is the order  $\alpha$  Augustin mean for the uniform input distribution and  $I_\alpha(u; W) = D_\alpha(W \| q_\alpha | u)$  by [35, Lemma 13]. Hence

$$I_\alpha(u; W) = \frac{\alpha}{\alpha-1} \ln \|\mu_\alpha\|. \tag{A.3}$$

Since  $D_\alpha(W \| q_\alpha | p) = I_\alpha(u; W)$  for all  $p \in \mathcal{P}(\mathcal{X})$  by (A.2) and (A.3), the probability measure  $q_\alpha$  is not only the order  $\alpha$  Augustin mean for the input distribution  $u$  but also the order  $\alpha$  Augustin center of  $W$  by [35, Thm. 1]. Then the constraint for the Rényi symmetry follows from the definition of  $q_\alpha$  and the input symmetry. Thus every input symmetric channel  $W$  is also a Rényi symmetric channel.

Now let us considers a Gallager symmetric channel  $W$ . Let  $\mathcal{S}_1, \dots, \mathcal{S}_m$  be the partition of the output set  $\mathcal{Y}$  assumed in the definition of Gallager symmetry, e.g., [24, p. 94],  $u$  be the uniform distribution on the input set  $\mathcal{X}$  of  $W$ , and  $\mu_\alpha$  and  $q_\alpha$  be the measures defined in (A.1). Gallager symmetry implies not only that  $\mu_\alpha$  and  $q_\alpha$  are probability mass functions but also that they satisfy the following identities:

$$\mu_\alpha(y) = \sum_x \mathbb{1}_{\{y \in \mathcal{S}_i\}} |\mathcal{S}_i|^{-\frac{1}{\alpha}} \left( \sum_{s \in \mathcal{S}_i} [W(s|x)]^\alpha \right)^{\frac{1}{\alpha}} \quad \forall x \in \mathcal{X}, y \in \mathcal{Y}. \tag{A.4}$$

Note that  $\mu_\alpha(y) = \mu_\alpha(z)$ , and hence  $q_\alpha(y) = q_\alpha(z)$ , whenever  $y$  and  $z$  are in the same  $\mathcal{S}_i$  as a result of (A.4). Using this fact together with the Gallager symmetry, one can confirm both (A.2) and  $q_\alpha = \sum_x u(x) W_\alpha^{q_\alpha}(x)$ . On the other hand,  $q_{1,u} \prec q_\alpha$ . Thus  $q_\alpha$  is the order  $\alpha$  Augustin mean for the uniform input distribution and  $I_\alpha(u; W) = D_\alpha(W \| q_\alpha | u)$  by [35, Lemma 13]. Hence (A.3) holds for Gallager symmetric  $W$ , as well. Thus

$D_\alpha(W \| q_\alpha | p) = I_\alpha(u; W)$  for all  $p \in \mathcal{P}(\mathcal{X})$  and  $q_\alpha$  is not only the order  $\alpha$  Augustin mean for the input distribution  $u$  but also the order  $\alpha$  Augustin center of  $W$  by [35, Thm. 1]. Then the constraint for the Rényi symmetry follows from the definition of  $q_\alpha$  given in (A.1), the Gallager symmetry and (A.4). Thus every Gallager symmetric channel  $W$  is also a Rényi symmetric channel.

**B. Proof of Lemma 5.5**

We first prove the existence of an order  $\alpha^*$  in  $(0, 1)$  satisfying (5.6) using the intermediate value theorem. The Rényi symmetry of  $W$  implies

$$D_\alpha(W(x) \| q_{\eta, W}) = \sup_{z \in \mathcal{X}} D_\alpha(W(z) \| q_{\eta, W}).$$

for all  $x \in \mathcal{X}$ ,  $\eta \in (0, 1]$ , and  $\alpha \in \mathbb{R}_+$ . Then (4.7), (4.8), and (4.9) imply

$$C_{\alpha, W} = \inf_{\eta \in (0, 1]} D_\alpha(W(x) \| q_{\eta, W}) \tag{B.1}$$

$$= D_\alpha(W(x) \| q_{\alpha, W}) \tag{B.2}$$

for all  $x \in \mathcal{X}$  and  $\alpha \in (0, 1]$ . Thus the nonnegativity of the Rényi divergence and (3.2) imply

$$D_1(W_\alpha^{q_{\alpha, W}}(x) \| q_{\alpha, W}) \leq C_{\alpha, W} \tag{B.3}$$

$$D_1(W_\alpha^{q_{\alpha, W}}(x) \| W(x)) \leq \frac{1-\alpha}{\alpha} C_{\alpha, W} \tag{B.4}$$

for all  $x \in \mathcal{X}$  and  $\alpha \in (0, 1)$ . On the other hand the Pinsker's inequality imply for all  $x \in \mathcal{X}$  and  $\alpha \in (0, 1)$

$$\| W_\alpha^{q_{\alpha, W}}(x) - W(x) \| \leq \sqrt{2D_1(W_\alpha^{q_{\alpha, W}}(x) \| W(x))}.$$

Thus  $\lim_{\alpha \uparrow 1} \| W_\alpha^{q_{\alpha, W}}(x) - W(x) \| = 0$  for all  $x \in \mathcal{X}$  as a result of (B.4). On the other hand, the Augustin center  $q_{\alpha, W}$  is continuous in  $\alpha$  on  $(0, 1]$  for the total variation topology on  $\mathcal{P}(\mathcal{Y})$  by [35, Lemmas 23-(d) and 24]. Then the lower semicontinuity of the Rényi divergence in its arguments for the topology of setwise convergence, i.e., [20, Thm. 15], implies

$$\liminf_{\alpha \uparrow 1} D_1(W_\alpha^{q_{\alpha, W}}(x) \| q_{\alpha, W}) \geq D_1(W(x) \| q_{1, W}).$$

Then using (B.2) we get

$$\liminf_{\alpha \uparrow 1} D_1(W_\alpha^{q_{\alpha, W}}(x) \| q_{\alpha, W}) \geq C_{1, W}. \tag{B.5}$$

Furthermore,  $D_1(W_\alpha^{q_{\alpha, W}}(x) \| q_{\alpha, W})$  is continuous in  $\alpha$  on  $(0, 1)$  by [39, Lemma 16] because  $q_{\alpha, W}$  is continuous in  $\alpha$  on  $(0, 1]$  for the total variation topology. Then (B.3) implies

$$\limsup_{\alpha \downarrow 0} D_1(W_\alpha^{q_{\alpha, W}}(x) \| q_{\alpha, W}) \leq \lim_{\alpha \downarrow 0} C_{\alpha, W}. \tag{B.6}$$

The existence of an  $\alpha^* \in (0, 1)$  satisfying (5.6) follows from (B.5), (B.6), and the continuity of  $D_1(W_\alpha^{q_{\alpha, W}}(x) \| q_{\alpha, W})$  in  $\alpha$  on  $(0, 1)$  by the intermediate value theorem [42, 4.23].

We proceed with showing that any order  $\alpha^*$  satisfying (5.6) also satisfies (5.7). The definition of the SPE given in (4.11) and the consequence of the Rényi symmetry given in (B.1), imply that

$$\begin{aligned} E_{sp}(R, W) &= \sup_{\alpha \in (0, 1)} \inf_{\eta \in (0, 1]} \frac{1-\alpha}{\alpha} [D_\alpha(W(x) \| q_{\eta, W}) - R] \\ &\leq \sup_{\alpha \in (0, 1)} f(\alpha, R), \end{aligned} \tag{B.7}$$

where  $f(\alpha, \tau) \triangleq \frac{(1-\alpha)}{\alpha} [D_\alpha(W(x) \| q_{\alpha^*, W}) - \tau]$ . We show in the following that

$$\sup_{\alpha \in (0,1)} f(\alpha, R) = D_1(W_{\alpha^*}^{q_{\alpha^*, W}}(x) \| W(x)). \tag{B.8}$$

Then for any order  $\alpha^*$  satisfying (5.6), equations (3.2), (B.2) and the definition of SPE given in (4.11) imply

$$\begin{aligned} \sup_{\alpha \in (0,1)} f(\alpha, R) &= \frac{(1-\alpha^*)}{\alpha^*} [C_{\alpha^*, W} - R] \\ &\leq E_{sp}(R, W). \end{aligned} \tag{B.9}$$

Thus the inequalities given in (B.7) and (B.9) hold as equalities and  $\alpha^*$  satisfying (5.6) also satisfies (5.7) by (B.8).

Now we prove the identity given in (B.8), which we have assumed in the preceding. The Rényi divergence is nondecreasing in its order by [20, Thm. 3] and  $D_\alpha(W(x) \| q_{\alpha^*, W}) = \frac{\alpha}{1-\alpha} D_{1-\alpha}(q_{\alpha^*, W} \| W(x))$  for all  $\alpha$  in  $(0, 1)$  by definition. Then  $D_\alpha(W(x) \| q_{\alpha^*, W})$  is finite for all  $\alpha$  in  $(0, 1)$ . Thus both  $D_\alpha(W(x) \| q_{\alpha^*, W})$  and  $D_1(W_{\alpha^*}^{q_{\alpha^*, W}}(x) \| q_{\alpha^*, W})$  are continuously differentiable in  $\alpha$  on  $(0, 1)$  by [35, Lemma 11] and their derivatives on  $(0, 1)$  are

$$\frac{d}{d\alpha} D_\alpha(W(x) \| q_{\alpha^*, W}) = \frac{1}{(\alpha-1)^2} D_1(W_{\alpha^*}^{q_{\alpha^*, W}}(x) \| W(x)), \tag{B.10}$$

$$\frac{d}{d\alpha} D_1(W_{\alpha^*}^{q_{\alpha^*, W}}(x) \| q_{\alpha^*, W}) = \alpha \mathbf{V}_{W_{\alpha^*}^{q_{\alpha^*, W}}(x)}[\xi_x], \tag{B.11}$$

where  $\xi_x = \ln \frac{dW(x)}{d\nu} - \ln \frac{dq_{\alpha^*, W}}{d\nu}$ . Then

$$\frac{\partial}{\partial \alpha} f(\alpha, \tau) = \frac{1}{\alpha^2} (\tau - D_1(W_{\alpha^*}^{q_{\alpha^*, W}}(x) \| q_{\alpha^*, W})). \tag{B.12}$$

Consequently, (5.6) implies (B.8) by the derivative test provided that  $\xi_x = \gamma$  does not hold  $W_{\alpha^*}^{q_{\alpha^*, W}}(x)$ -a.s. for any  $\gamma \in \mathbb{R}_+$  and  $\alpha \in (0, 1)$ . If  $W_{\alpha^*}^{q_{\alpha^*, W}}(\{\xi_x = \gamma\} | x) = 1$  for some  $\gamma \in \mathbb{R}_+$  and  $\alpha \in (0, 1)$ , on the other hand, then the identities (B.13), (B.14), and (B.15) hold for all  $\alpha \in (0, 1)$  and one can confirm (B.8) by substitution.

$$D_\alpha(W(x) \| q_{\alpha^*, W}) = \gamma + \frac{1}{1-\alpha} \ln \frac{1}{W(\{\xi_x = \ln \gamma\} | x)}, \tag{B.13}$$

$$D_1(W_{\alpha^*}^{q_{\alpha^*, W}}(x) \| q_{\alpha^*, W}) = \gamma + \ln \frac{1}{W(\{\xi_x = \ln \gamma\} | x)}, \tag{B.14}$$

$$D_1(W_{\alpha^*}^{q_{\alpha^*, W}}(x) \| W(x)) = \ln \frac{1}{W(\{\xi_x = \ln \gamma\} | x)}. \tag{B.15}$$

Now we are left with establishing (5.8) with either of the additional hypotheses. Let us first assume that there does not exist a  $\gamma$  satisfying  $W_{\alpha^*}^{q_{\alpha^*, W}}(\{\frac{dW(x)}{d\nu} = \gamma \frac{dq_{\alpha^*, W}}{d\nu}\} | x) = 1$ . Then  $\mathbf{V}_{W_{\alpha^*}^{q_{\alpha^*, W}}(x)}[\xi_x] > 0$  for all  $\alpha$  in  $(0, 1)$  and thus  $D_1(W_{\alpha^*}^{q_{\alpha^*, W}}(x) \| q_{\alpha^*, W})$  is increasing in  $\alpha$  on  $(0, 1)$ . Since  $D_1(W_{\alpha^*}^{q_{\alpha^*, W}}(x) \| q_{\alpha^*, W})$  is also continuous in  $\alpha$  on  $(0, 1)$ , it has a continuous and increasing inverse function. Then as a result of (B.12) there exists an  $\epsilon > 0$  and an increasing continuous function  $h : (R - \epsilon, R + \epsilon) \rightarrow (0, 1)$  satisfying

$$\sup_{\alpha \in (0,1)} f(\alpha, \tau) = f(h(\tau), \tau) \quad \forall \tau \in (R - \epsilon, R + \epsilon).$$

Then (B.7) implies

$$E_{sp}(\tau, W) \leq f(h(\tau), \tau) \quad \forall \tau \in (R - \epsilon, R + \epsilon). \tag{B.16}$$

On the other hand, (4.11) and (B.2) imply

$$E_{sp}(\tau, W) \geq f(\alpha^*, \tau) \quad \forall \tau \in \mathbb{R}_+. \tag{B.17}$$

Furthermore, (B.9) implies

$$E_{sp}(R, W) \geq f(\alpha, R) \quad \forall \alpha \in (0, 1). \tag{B.18}$$

Then (5.8) follows from (B.16), (B.17), (B.18), the identity  $h(R) = \alpha^*$ , the definition of the derivative as a limit, and the continuity of  $h(\cdot)$ , which is defined as the inverse function of  $D_1(W_{\alpha^*}^{q_{\alpha^*, W}}(x) \parallel q_{\alpha^*, W})$  as a function of  $\alpha$ .

Now we establish (5.8) assuming the existence of a  $q$  satisfying  $q_{\alpha, W} = q$  for all  $\alpha$  in  $(0, 1)$ . If there does not exist a  $\gamma$  satisfying  $W_{\alpha^*}^q \left( \left\{ \frac{dW(x)}{d\nu} = \gamma \frac{dq}{d\nu} \right\} \middle| x \right) = 1$ , then (5.8) holds by the preceding discussion. If there exists such a  $\gamma$ , then as a result of (B.2) and (B.13) we have

$$C_{\alpha, W} = \gamma + \frac{1}{1-\alpha} \ln \frac{1}{W(\{\xi_x = \ln \gamma\} | x)} \quad \forall \alpha \in (0, 1) \tag{B.19}$$

Then such a  $\gamma$  does not exist by the hypotheses of the lemma because (B.19) and  $C_{1, W} = \lim_{\alpha \uparrow 1} C_{\alpha, W}$  imply  $C_{1, W} = \infty$  for  $W(\{\xi_x = \ln \gamma\} | x) < 1$  case and  $\lim_{\alpha \downarrow 0} C_{\alpha, W} = C_{1, W}$  for  $W(\{\xi_x = \ln \gamma\} | x) = 1$  case.

### C. Shannon’s bounds for AWGN channels and the sphere packing exponent

Shannon, [43, (3)], bounded the error probability of length  $n$  block codes described in Theorem 5.10 as

$$Q(\theta) \leq P_e^{(n)} \leq Q(\theta) - \int_0^\theta \frac{\Omega(\xi)}{\Omega(\theta)} dQ(\xi) \tag{C.1}$$

where  $\Omega(\cdot) : [0, \pi] \rightarrow [0, \frac{2\pi^{n/2}}{\Gamma(n/2)}]$  is the function mapping the cone angle to the corresponding solid angle in  $\mathbb{R}^n$ ,  $\theta$  is the cone angle satisfying  $\Omega(\theta) = \frac{\Omega(\pi)}{M}$ , and  $Q(\xi)$  is the probability that a point  $X$  in  $\mathbb{R}^n$  at a distance  $\sqrt{n\rho}$  from the origin  $O$  being moved outside a circular cone of half-angle  $\xi$  with the vertex at the origin  $O$  and the axis at  $OX$  by a Gaussian noise of variance  $\sigma^2$ .

Shannon, [43, (4) and (5)], derived the exact asymptotic expressions for both the upper bound and the lower bound given in (C.1) in terms of functions  $f(\cdot)$  and  $g(\cdot)$  that do not depend on the block length  $n$ :

$$Q(\theta) \sim \frac{f(\theta)}{\sqrt{n}} e^{-nE_L(\theta)} \quad \forall \theta > \theta_c, \tag{C.2}$$

$$Q(\theta) - \int_0^\theta \frac{\Omega(\xi)}{\Omega(\theta)} dQ(\xi) \sim \frac{g(\theta)}{\sqrt{n}} e^{-nE_L(\theta)} \quad \forall \theta \in (\theta_c, \theta_{cr}), \tag{C.3}$$

where  $\theta_c$  and  $\theta_{cr}$ , the cone angles corresponding to the channel capacity and the critical rate, are given by

$$\theta_c \triangleq \arcsin \left[ \left( 1 + \frac{\rho}{\sigma^2} \right)^{-1/2} \right], \tag{C.4}$$

$$\theta_{cr} \triangleq \arcsin \left[ \left( \frac{1}{2} + \frac{\rho}{4\sigma^2} + \sqrt{\frac{1}{4} + \left( \frac{\rho}{4\sigma^2} \right)^2} \right)^{-1/2} \right], \tag{C.5}$$



and  $E_L(\cdot)$  (the fixed cone angle exponent) is defined via the function  $G(\cdot)$  as follows

$$E_L(\theta) \triangleq \frac{\rho}{2\sigma^2} - \frac{\sqrt{\rho}}{2\sigma} G(\theta) \cos \theta - \ln(G(\theta) \sin \theta), \tag{C.6}$$

$$G(\theta) \triangleq \frac{1}{2} \left( \frac{\sqrt{\rho}}{\sigma} \cos \theta + \sqrt{\frac{\rho}{\sigma^2} \cos^2 \theta + 4} \right). \tag{C.7}$$

**Remark C.1** Shannon’s notation in [43] is slightly different from ours; Shannon works with the signal to noise “amplitude” ratio  $A \triangleq \frac{\sqrt{\rho}}{\sigma}$ , rather than cost constraint  $\rho$  and the noise power  $\sigma^2$ . Furthermore, Shannon specifies the critical cone angle  $\theta_{cr}$  as the solution of the equation given in (C.8). Nevertheless, one obtains the closed-form expression given in (C.5), by plugging in the definition of  $G(\cdot)$  –given in (C.7)– in (C.8) and solving the resulting quadratic equation for  $\sin^2 \theta_{cr}$ .

$$2 \cos \theta_{cr} - \frac{\sqrt{\rho}}{\sigma} G(\theta_{cr}) \sin^2 \theta_{cr} = 0. \tag{C.8}$$

Shannon presented the exact asymptotic expression for the rate  $R$  in terms of the cone angle  $\theta$  in [43, (11)], as well:

$$e^{nR} \sim \frac{\sqrt{2\pi n} \cos \theta}{\sin^{n-1} \theta}. \tag{C.9}$$

We obtain the fixed-rate asymptotic expression corresponding to the fixed cone angle asymptotic expressions (C.2), (C.3), and (C.6), by first deriving the asymptotic expression for the cone angle  $\theta$  for a fixed-rate  $R$  using (C.9). If  $|\delta_n| \ll 1$  and

$$\theta_n = \xi + \delta_n, \tag{C.10}$$

then using the small angle approximation for the trigonometric functions we get

$$\frac{\sqrt{2\pi n} \cos \theta_n}{\sin^{n-1} \theta_n} = \frac{\sqrt{2\pi n} \cos \xi}{\sin^{n-1} \xi} \frac{1 - \delta_n \tan \xi + O(\delta_n^2)}{(1 + \delta_n \cot \xi + O(\delta_n^2))^{n-1}}.$$

Invoking  $\ln(1 + \epsilon) = \epsilon + O(\epsilon^2)$ , we get

$$\ln \frac{\sqrt{2\pi n} \cos \theta_n}{\sin^{n-1} \theta_n} = n \ln \frac{1}{\sin \xi} + \ln \left( \sqrt{\frac{\pi n}{2}} \sin 2\xi \right) - [(n - 1) \cot \xi + \tan \xi] \delta_n + nO(\delta_n^2)$$

Consequently, if

$$\xi = \arcsin(e^{-R}), \tag{C.11}$$

$$\delta_n = \frac{1}{n \cot \xi} \ln \left( \sqrt{\frac{\pi n}{2}} \sin 2\xi \right), \tag{C.12}$$

then the rate corresponding to the cone angle  $\theta_n$  at the block length  $n$  is  $R + O\left(\frac{\ln^2 n}{n^2}\right)$ . In other words, we get a fixed-rate by changing the cone angle by an additive factor proportional to  $\frac{\ln n}{n}$ . In order to obtain the exact asymptotic expressions for the upper and lower bounds to the error probability given in (C.1) via (C.2)

and (C.3) at a fixed-rate  $R$ , we will apply Taylor’s expansion. To that end, we first calculate the derivatives of  $G$  and  $E_L$ . As a result of (C.7), we have

$$\begin{aligned} \frac{d}{d\theta} G &= \frac{1}{2} \left( -\frac{\sqrt{\varrho}}{\sigma} \sin \theta - \frac{(e/\sigma^2) \sin \theta \cos \theta}{\sqrt{(e/\sigma^2) \cos^2 \theta + 4}} \right) \\ &= -\frac{(\sqrt{\varrho}/\sigma) \sin \theta}{\sqrt{(e/\sigma^2) \cos^2 \theta + 4}} G. \end{aligned}$$

Then using (C.6), we get

$$\begin{aligned} \frac{d}{d\theta} E_L &= \frac{\sqrt{\varrho}}{2\sigma} \left[ -\cos \theta \frac{d}{d\theta} G + \sin \theta G \right] - \frac{1}{G} \frac{d}{d\theta} G - \cot \theta \\ &= \frac{(\sqrt{\varrho}/\sigma) \sin \theta}{\sqrt{(e/\sigma^2) \cos^2 \theta + 4}} G^2 + \frac{(\sqrt{\varrho}/\sigma) \sin \theta}{\sqrt{(e/\sigma^2) \cos^2 \theta + 4}} - \cot \theta \\ &= \frac{\sqrt{\varrho}}{\sigma} \sin \theta G - \cot \theta. \end{aligned}$$

Thus (C.10) and the Taylor’s expansion imply

$$E_L(\theta_n) = E_L(\xi) + \left( \frac{\sqrt{\varrho}}{\sigma} \sin \xi G(\xi) - \cot \xi \right) \delta_n + O(\delta_n^2).$$

Invoking (C.12), we get the following asymptotic expression

$$\frac{e^{-nE_L(\theta_n)}}{\sqrt{n}} \sim \frac{e^{-nE_L(\xi)}}{\sqrt{n}} \left( \sqrt{\frac{\pi}{2}} n \sin 2\xi \right)^{1 - \frac{\sqrt{\varrho}}{\sigma} \frac{\sin \xi G(\xi)}{\cot \xi}}. \tag{C.13}$$

On the other hand, invoking (C.7) in (C.6), we get

$$\begin{aligned} E_L(\xi) &= \frac{\varrho}{4\sigma^2} \left[ 2 - \cos^2 \xi + \cos^2 \xi \sqrt{1 + \frac{4\sigma^2}{\varrho} \frac{1}{\cos^2 \xi}} \right] - \ln \sin \xi \frac{\sqrt{\varrho}}{2\sigma} \left( \cos \xi + \sqrt{\cos^2 \xi + \frac{4\sigma^2}{\varrho}} \right) \\ &= \frac{\varrho}{4\sigma^2} \left[ 1 + \sin^2 \xi + (1 - \sin^2 \xi) \sqrt{1 + \frac{4\sigma^2}{\varrho} \frac{1}{1 - \sin^2 \xi}} \right] + \ln \frac{1}{\sin \xi} \frac{\sqrt{\varrho}}{2\sigma} \left( \sqrt{\cos^2 \xi + \frac{4\sigma^2}{\varrho}} - \cos \xi \right) \\ &= \frac{\varrho}{4\sigma^2} \left[ 1 + \sin^2 \xi + (1 - \sin^2 \xi) \sqrt{1 + \frac{4\sigma^2}{\varrho} \frac{1}{1 - \sin^2 \xi}} \right] + \frac{1}{2} \ln \frac{1}{\sin^2 \xi} \left( 1 + \frac{\varrho \cos^2 \xi}{2\sigma^2} - \frac{\varrho \cos^2 \xi}{2\sigma^2} \sqrt{1 + \frac{4\sigma^2}{\varrho \cos^2 \xi}} \right). \end{aligned}$$

Then using (C.11), we get the closed-form expression for the sphere packing exponent given in [34, (74)]:

$$\begin{aligned} E_L(\xi) &= \frac{\varrho}{4\sigma^2} \left[ 1 + \frac{1}{e^{2R}} + \left( 1 - \frac{1}{e^{2R}} \right) \sqrt{1 + \frac{4\sigma^2}{\varrho} \frac{e^{2R}}{e^{2R} - 1}} \right] + \frac{1}{2} \ln \left[ e^{2R} - \frac{\varrho(e^{2R} - 1)}{2\sigma^2} \left( \sqrt{1 + \frac{4\sigma^2 e^{2R}}{\varrho(e^{2R} - 1)}} - 1 \right) \right] \\ &= E_{sp}(R, W, \varrho). \end{aligned} \tag{C.14}$$

On the other hand, (C.7) implies

$$\begin{aligned} \frac{\sqrt{\varrho}}{\sigma} \frac{\sin \xi G(\xi)}{\cot \xi} &= \frac{\varrho \sin^2 \xi}{2\sigma^2} \left( 1 + \sqrt{1 + \frac{4\sigma^2}{\varrho} \frac{1}{1 - \sin^2 \xi}} \right) \\ &= \frac{2 \sin^2 \xi}{1 - \sin^2 \xi} \left( \sqrt{1 + \frac{4\sigma^2}{\varrho} \frac{1}{1 - \sin^2 \xi}} - 1 \right)^{-1}. \end{aligned}$$

Invoking first (C.11), then (5.22), and finally (5.23), we get

$$\begin{aligned} \frac{\sqrt{\varrho} \sin \xi G(\xi)}{\sigma \cot \xi} &= \frac{2}{e^{2R}-1} \left( \sqrt{1 + \frac{4\sigma^2}{\varrho} \frac{e^{2R}}{e^{2R}-1}} - 1 \right)^{-1} \\ &= \frac{1}{\alpha^*} \\ &= 1 - \frac{\partial}{\partial s} E_{sp}(s, W, \varrho) \Big|_{s=R}. \end{aligned}$$

Then (C.13) and (C.14) imply

$$\frac{e^{-nE_L(\theta_n)}}{\sqrt{n}} \sim \left( \sqrt{\frac{\pi}{2}} \sin 2\xi \right)^{E'_{sp}(R)} n^{\frac{E'_{sp}(R)-1}{2}} e^{-nE_{sp}(R, W, \varrho)}$$

where  $E'_{sp}(R) = \frac{\partial}{\partial s} E_{sp}(s, W, \varrho) \Big|_{s=R}$ . Then [43, (3)], i.e., (C.1), imply both (1.1) and (5.1), because (C.1) is nothing but (1.1) and (5.1) for certain multiplicative constants.

#### D. Proof of (5.35)

Note that for any three random variables  $Z_1, Z_2$ , and  $Z_3$ ,

$$\mathbf{E} \left[ \left| \sum_{i=1}^3 Z_i \right|^3 \right] \leq \sum_{i=1}^3 \mathbf{E} \left[ |Z_i|^3 \right] + 6\mathbf{E} [|Z_1 Z_2 Z_3|] + 3 \sum_{i,j:i \neq j} \mathbf{E} [|Z_i Z_j^2|].$$

On the other hand, as a result of Hölder's inequality we have

$$\begin{aligned} \mathbf{E} [|Z_1 Z_2 Z_3|] &\leq \mathbf{E} \left[ |Z_1|^3 \right]^{\frac{1}{3}} \mathbf{E} \left[ |Z_2|^3 \right]^{\frac{1}{3}} \mathbf{E} \left[ |Z_3|^3 \right]^{\frac{1}{3}}, \\ \mathbf{E} [|Z_i Z_j^2|] &\leq \mathbf{E} \left[ |Z_i|^3 \right]^{\frac{1}{3}} \mathbf{E} \left[ |Z_j|^3 \right]^{\frac{2}{3}}. \end{aligned}$$

Furthermore, since the geometric mean is upper bounded by the arithmetic mean, we also have

$$\begin{aligned} \mathbf{E} \left[ |Z_1|^3 \right]^{\frac{1}{3}} \mathbf{E} \left[ |Z_2|^3 \right]^{\frac{1}{3}} \mathbf{E} \left[ |Z_3|^3 \right]^{\frac{1}{3}} &\leq \frac{1}{3} \sum_{i=1}^3 \mathbf{E} \left[ |Z_i|^3 \right], \\ \mathbf{E} \left[ |Z_i|^3 \right]^{\frac{1}{3}} \mathbf{E} \left[ |Z_j|^3 \right]^{\frac{2}{3}} &\leq \frac{1}{3} \mathbf{E} \left[ |Z_i|^3 \right] + \frac{2}{3} \mathbf{E} \left[ |Z_j|^3 \right]. \end{aligned}$$

Thus

$$\begin{aligned} \mathbf{E} \left[ \left| \sum_{i=1}^3 Z_i \right|^3 \right] &\leq \sum_{i=1}^3 \mathbf{E} \left[ |Z_i|^3 \right] + 2 \sum_{i=1}^3 \mathbf{E} \left[ |Z_i|^3 \right] + \sum_{i,j:i \neq j} \left( \mathbf{E} \left[ |Z_i|^3 \right] + 2\mathbf{E} \left[ |Z_j|^3 \right] \right) \\ &\leq 9 \sum_{i=1}^3 \mathbf{E} \left[ |Z_i|^3 \right]. \end{aligned}$$

#### Acknowledgments

The author would like to thank Hao-Chung Cheng (Hao-Chung Cheng is a postdoctoral research associate at the Department of Applied Mathematics and Theoretical Physics of the Faculty of Mathematics, University of Cambridge, Cambridge, UK) both for numerous inspiring discussions on the sphere packing bound and its

refinements, which helped the author to simplify and improve the statement of Lemma 3.4, and for his feedback on the manuscript. The author would also like to thank the reviewer for pointing out [28] and for his feedback on the manuscript.

This work is supported by The Science Academy, Turkey, under The Science Academy's Young Scientist Awards Program (BAGEP) and by The Scientific and Technological Research Council of Turkey (TÜBİTAK) under Grant 119E053.

## Disclaimers

This paper was presented in part at the 2019 IEEE International Symposium on Information Theory [37].

## References

- [1] Altuğ Y, Wagner AB. Refinement of the sphere packing bound for symmetric channels. In: 49th Annual Allerton Conference on Communication, Control, and Computing; Monticello, IL, USA; 2011. pp. 30-37. doi: 10.1109/Allerton.2011.6120146
- [2] Altuğ Y, Wagner AB. Refinement of the sphere-packing bound: asymmetric channels. IEEE Transactions on Information Theory 2014; 60 (3):1592-1614. doi: 10.1109/TIT.2014.2299275
- [3] Altuğ Y, Wagner AB. On exact asymptotics of the error probability in channel coding: symmetric channels 2019. [arXiv:1908.11419](https://arxiv.org/abs/1908.11419) [cs.IT].
- [4] Arikan E. On the reliability exponent of the exponential timing channel. IEEE Transactions on Information Theory 2002; 48 (6): 1681-1689. doi:10.1109/TIT.2002.1003846
- [5] Augustin U. Error estimates for low rate codes. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 1969; 14 (1): 61-88. doi:10.1007/BF00534118
- [6] Augustin U. [Noisy Channels](#). Habilitation thesis, Universität Erlangen-Nürnberg, Erlangen, Germany, 1978.
- [7] Bahadur RR, Ranga Rao R. On deviations of the sample mean. The Annals of Mathematical Statistics 1960; 31 (4): 1015-1027. doi: 10.1214/aoms/1177705674
- [8] Berry AC. The accuracy of the Gaussian approximation to the sum of independent variates. Transactions of the American Mathematical Society 1941; 49 (1): 122-136. doi: 10.2307/1990053
- [9] Burnashev MV, Kutoyants YuA. On the sphere-packing bound, capacity, and similar results for Poisson channels. Problems of Information Transmission 1999; 35 (2): 95-111.
- [10] Cheng HC, Hsieh MH, Tomamichel M. Quantum sphere-packing bounds with polynomial prefactors. IEEE Transactions on Information Theory 2019; 65 (5): 2872-2898. doi: 10.1109/TIT.2019.2891347
- [11] Cheng HC, Nakiboğlu B. Refined strong converse for the constant composition codes 2020. [arXiv:2002.11414](https://arxiv.org/abs/2002.11414) [cs.IT].
- [12] Csiszár I, Longo G. On the error exponent for source coding and for testing simple statistical hypotheses. Studia Scientiarum Mathematicarum Hungarica 1971; 6:181-191.
- [13] Csiszár I, Körner J. Information theory: coding theorems for discrete memoryless systems. Cambridge University Press, Cambridge, UK, 2011.
- [14] Dalai M. Lower bounds on the probability of error for classical and classical-quantum channels. IEEE Transactions on Information Theory 2013; 59 (12): 8027-8056. doi:10.1109/TIT.2013.2283794
- [15] Dalai M, Winter A. Constant compositions in the sphere packing bound for classical-quantum channels. IEEE Transactions on Information Theory 2017; 63 (9): 5603-5617. doi:10.1109/TIT.2017.2726555

- [16] Dobrushin RL. 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 and Its Applications* 1962; 7 (3): 270-300. doi: 10.1137/1107027
- [17] Dobrushin RL. An asymptotic bound for the probability error of information transmission through a channel without memory using the feedback. *Problemy Kibernetiki* 1962; (8): 161-168.
- [18] Ebert PM. Error Bounds For Parallel Communication Channels, [Technical Report 448](#), Research Laboratory of Electronics at Massachusetts Institute of Technology. Cambridge, MA, USA: Massachusetts Institute of Technology, 1966.
- [19] Elias P. Coding for two noisy channels. In: *Proceedings of Third London Symposium of Information Theory*; London, UK; Butterworth Scientific, 1955. pp. 61-74.
- [20] Van Erven T, Harremoës P. Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory* 2014; 60 (7): 3797-3820. doi: 10.1109/TIT.2014.2320500
- [21] Esseen CG. On the Liapunoff limit of error in the theory of probability. *Arkiv för Matematik, Astronomi Och Fysik* 1942; A28: 1-19.
- [22] Esseen CG. Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. *Acta Mathematica* 1945; 77: 1-125. doi:10.1007/BF02392223
- [23] Gallager RG. A simple derivation of the coding theorem and some applications. *IEEE Transactions on Information Theory* 1965; 11 (1): 3-18. doi: 10.1109/TIT.1965.1053730
- [24] Gallager RG. *Information theory and reliable communication*. New York, NY, USA: John Wiley & Sons, Inc., 1968.
- [25] Gnedenko BV, Kolmogorov AN. *Limit Distributions for Sums of Independent Random Variables*. Cambridge, MA, USA: Addison-Wesley, 1954.
- [26] Haroutunian EA. Bounds for the exponent of the probability of error for a semicontinuous memoryless channel. *Problems of Information Transmission* 1968; 4 (4).
- [27] Haroutunian EA. Lower bound for error probability in channels with feedback. *Problems of Information Transmission* 1977; 13 (2):107-114.
- [28] Holevo AS. Remarks on the classical capacity of quantum channel. 2002. [arXiv:quant-ph/0212025](#).
- [29] Jensen JL. *Saddlepoint approximations*. New York, NY, USA: Oxford University Press, 1995.
- [30] Lanco A, Östman J, Durisi G, Koch T, and Vazquez-Vilar G. Saddlepoint approximations for noncoherent single-antenna rayleigh block-fading channels. In: *2019 International Symposium on Information Theory*; Paris, France; 2019. pp.612-616. doi:10.1109/ISIT.2019.8849659
- [31] Lanco A, Östman J, Durisi G, Koch T, and Vazquez-Vilar G. Saddlepoint approximations for short-packet wireless communications. 2019. [arXiv:1904.10442](#) [cs.IT].
- [32] Lapidot A. On the reliability function of the ideal Poisson channel with noiseless feedback. *IEEE Transactions on Information Theory* 1993; 39 (2): 491-503. doi: 10.1109/18.212279
- [33] Nakiboğlu B. The Augustin center and the sphere packing bound for memoryless channels. In: *2017 International Symposium on Information Theory*; Aachen, Germany; 2017. pp.1401-1405. doi: 10.1109/ISIT.2017.8006759
- [34] Nakiboğlu B. The sphere packing bound for memoryless channels. 2018. [arXiv:1804.06372](#) [cs.IT].
- [35] Nakiboğlu B. The Augustin capacity and center. *Problems of Information Transmission* 2019; 55 (4): 299-342. doi: 10.1134/S003294601904001 ([arXiv:1803.07937](#) [cs.IT])
- [36] Nakiboğlu B. The Rényi capacity and center. *IEEE Transactions on Information Theory* 2019; 65 (2): 841-860. doi: 10.1109/TIT.2018.2861002
- [37] Nakiboğlu B. A simple derivation of the refined spb for the constant composition codes. In: *2019 International Symposium on Information Theory*; Paris, France; 2019. pp. 2659-2663. doi: 10.1109/ISIT.2019.8849819

- [38] Nakiboğlu B. The sphere packing bound for DSPCs with feedback à la Augustin. *IEEE Transactions on Communications* 2019; 67 (11): 7456-7467. doi: 10.1109/TCOMM.2019.2931302
- [39] Nakiboğlu B. The sphere packing bound via Augustin's method. *IEEE Transactions on Information Theory* 2019; 65 (2): 816-840. doi: 10.1109/TIT.2018.2882547
- [40] Omura JK. A lower bounding method for channel and source coding probabilities. *Information and Control* 1975; 27 (2): 148-177. doi: 10.1016/S0019-9958(75)90120-5
- [41] Richters JS. Communication over fading dispersive channels, [Technical report 464](#), Research Laboratory of Electronics at Massachusetts Institute of Technology. Cambridge, MA, USA: Massachusetts Institute of Technology, 1967.
- [42] Rudin W. *Principles of Mathematical Analysis*. New York, NY, USA: McGraw-Hill, 1976.
- [43] Shannon CE. Probability of error for optimal codes in a Gaussian channel. *Bell System Technical Journal* 1959; 38 (3): 611-656. doi: 10.1002/j.1538-7305.1959.tb03905.x
- [44] Shannon CE, Gallager RG, Berlekamp ER. Lower bounds to error probability for coding on discrete memoryless channels. I. *Information and Control* 1967; 10 (1): 65-103. doi: 10.1016/S0019-9958(67)90052-6
- [45] Shevtsova IG. An improvement of convergence rate estimates in the Lyapunov theorem. *Doklady Mathematics* 2010; 82 (3): 862-864. doi: 10.1134/S1064562410060062
- [46] Strassen V. Asymptotische abschätzungen in Shannons Informationstheorie (in German). In: *Transactions of the Third Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*. Prague, Czech Republic: Czechoslovak Academy of Sciences, 1962. pp. 689-723.
- [47] Vazquez-Vilar G. Error probability bounds for gaussian channels under maximal and average power constraints. 2019. [arXiv:1907.03163](#) [cs.IT].
- [48] Vazquez-Vilar G, i Fabregas AG, Koch T, Lancho A. Saddlepoint approximation of the error probability of binary hypothesis testing. In: *2018 International Symposium on Information Theory; Colorado, USA; 2018*. pp. 2306-2310. doi: 10.1109/ISIT.2018.8437503
- [49] Wiechman G, Sason I. An improved sphere-packing bound for finite-length codes over symmetric memoryless channels. *IEEE Transactions on Information Theory* 2008; 54 (5): 1962-1990. doi: 10.1109/TIT.2008.920216
- [50] Wyner AD. Capacity and error exponent for the direct detection photon channel. II. *IEEE Transactions on Information Theory* 1988; 34 (6): 1462-1471. doi: 10.1109/18.21285