==== **INFORMATION THEORY** ====

The Sphere Packing Bound
for Memoryless Channels

B. Nakiboğlu

Middle East Technical University, Ankara, Turkey
e-mail: bnakib@metu.edu.tr

Abstract—Sphere packing bounds (SPBs)—with prefactors that are polynomial in the block length—are derived for codes on two families of memoryless channels using Augustin's method: (possibly nonstationary) memoryless channels with (possibly multiple) additive cost constraints and stationary memoryless channels with convex constraints on the composition (i.e., empirical distribution, type) of the input codewords. A variant of Gallager's bound is derived in order to show that these sphere packing bounds are tight in terms of the exponential decay rate of the error probability with the block length under mild hypotheses.

Key words: Augustin's method, sphere packing exponent, error exponent, reliability function, memoryless channels, Gaussian channels, Poisson channels.

DOI: 10.1134/S0032946020030011

1. INTRODUCTION

Most proofs of the sphere packing bound (SPB) have been either for the stationary channels with finite input sets [2–14] or for the stationary channels with a specific noise structure, e.g., Poisson or Gaussian [15–22]. Proofs of the SPB based on Augustin's method are exceptions to this observation: [23–25] do not assume either the finiteness of the input set or a specific noise structure, nor do they assume the stationarity of the channel. However, [23; 24, Section 31; 25] establish the SPB for the product channels rather than the memoryless channels; hence, proofs of the SPB for the composition constrained codes[1] on the stationary channels [9–20] —which include the important special case of the cost constrained ones [15–20]—are not subsumed by [23; 24, Section 31; 25]. In [24, Section 36], Augustin proved the SPB for the cost constrained (possibly nonstationary) memoryless channels assuming a bounded cost function. The framework of [24, Theorem 36.6] subsumes all previously considered models [2–22] except the Gaussian ones [17–20].

Theorem 2, presented in Section 3, establishes the SPB for a framework that subsumes all of the models considered in [2–25] by employing [26], which analyzes Augustin's information measures. Our use of [26] and Augustin's information measures is similar to the use of [27] and Rényi's information measures in [25]. For the product channels, [25, Theorem 2] improved the previous results by Augustin in [23; 24, Section 31] by establishing the SPB with a prefactor that is polynomial in the block length n for the hypothesis that the order $1/2$ Rényi capacity of the component channels are $O(\ln n)$. For the cost constrained memoryless channels, Theorem 2 enhances the prefactor of [24, Theorem 36.6] in an analogous way, from $e^{-O(\sqrt{n})}$ to $e^{-O(\ln n)}$. The prefactor of Theorem 2, however, is inferior to the prefactors reported in [3–6] for various symmetric channels, in [13] for the constant composition codes on discrete stationary product channels, in [17] for the stationary

---

[1] According to [12, p. 183], the SPB for the constant composition codes appears in [9] with an incomplete proof. The first complete proof of the SPB for the constant composition codes is provided in [10].

Gaussian channel, and in [21, 22] for certain noncoherent fading channels. Determination of the optimal prefactor, in the spirit of [3–5], remains an open problem for the general case.[2] Similarly to [24, Theorem 36.6], Theorem 2 holds for nonstationary channels as well. Unlike [24, Theorem 36.6], Theorem 2 does not assume the cost functions to be bounded.

The stationarity is assumed in most of the previous derivations of the SPB [2–22]. Given a stationary product channel, one can obtain a stationary memoryless channel by imposing composition—i.e., type, empirical distribution—constraints on the input codewords. The cost constraints can be interpreted as a particular convex case of this more general composition constraints. This interpretation, considered together with the composition based expurgations, is one of the main motivating factors behind the study of constant composition codes. The composition-based expurgations, however, are useful only when the input set of the channel is finite. Nevertheless, if the constraint set for the composition of the codewords is convex, then one can derive an SPB with a polynomial prefactor using Augustin's information measures; see Theorem 1 in Section 3. The derivation of Theorem 1 relies on the Augustin center of the constraint set rather than the Augustin mean of the most populous composition of the code. Note that the most populous composition of the code might not even have more than one codeword when the input set is infinite. The framework of Theorem 1 is general enough to subsume the frameworks of all previous proofs of the SPB for the memoryless channels that we are aware of, except the frameworks of the proofs based on Augustin's method [23–25]. Theorems 1 and 2 are asymptotic SPBs; but they are proved using nonasymptotic SPBs presented in Lemmas 9 and 10.

The SPB implies that exponential decay rate of the optimal error probability with the block length—i.e., the reliability function, the error exponent—is bounded from above by the sphere packing exponent (SPE). For the memoryless channels in consideration, Augustin's variant of Gallager's bound implies that the SPE bounds the reliability function from below as well, provided that the list decoding is allowed. Augustin's variant of Gallager's bound is presented in Section 2.4. One can use standard results such as [29, 30] with minor modifications in order to establish the SPE as a lower bound to the reliability function for the list decoding as well. Thus, Augustin's variant is of interest to us not because of what it implies about the reliability function but because of how it implies it. What is unique about Augustin's variant is that it establishes an achievability result in terms of the Augustin information rather than the Rényi information used in the standard form of the Gallager's bound [29]. Augustin's variant relies on the fixed point property of the Augustin mean described in (7) to do that. It is worth mentioning that [30] implicitly employs the same fixed point property but in a different way.

Before starting our discussion in earnest, let us point out a subtlety about the derivations of the SPB that is usually overlooked. [31] claimed to prove the SPB for arbitrary stationary product channels, without using any constant composition arguments.[3] The derivation of [31, Theorem 19], however, establishes an upper bound on the reliability function that is strictly greater than the SPE in many channels. This has been demonstrated numerically in [13, p. 1594 and Appendix A]. An analytic confirmation of this observation is presented in Appendix A. The problematic step in [31] is the application of Lagrange multiplier techniques; see [13, footnote 8]. The proof of [31, Theorem 19] invokes [31, Theorem 16], which is valid for the Lagrange multiplier s associated with the input distribution p satisfying $E_{\mathrm{sp}}(R, W) = E_{\mathrm{sp}}(R, W, p)$. For an arbitrary input distribution p, however, the associated Lagrange multiplier may or may not be equal to the one for the optimal input distribution p. This is the reason why the upper bound to the reliability function established

---

[2] In [28], we have derived refined SPBs (which are optimal in terms of the prefactor for nonsingular cases) for all of the cases considered in [3–6, 13, 17, 21, 22] using Augustin information measures via [26].

[3] [31, p. 413] reads "An important feature of the lower bound, which will be derived, is that no assumption of constant-composition codewords is made, not even as an intermediate step."

in [31, Theorem 19] is not equal to the SPE in general, contrary to the claim repeated in [32, Lemma 1] and [33, Theorem 10.1.4]. In a nutshell, the proof of [31, Theorem 19] tacitly asserts a minimax equality that does not hold in general. For stationary memoryless channels with finite input alphabets, one can avoid this issue using the constant composition arguments. However, in that case, the proof presented in [31] becomes a mere reproduction of the one in [10]. More recently, [34] proposed a derivation of the SPB for stationary channels with a single cost constraint using the approach presented in [31]. Similar to [31], however, the proof in [34] asserts a minimax equality that does not hold in general. In particular, it is claimed that Q^n does not depend on $\boldsymbol{x}_m$ in [34, equation (26)]. To assert that, one has to include an additional supremum over $\boldsymbol{x}_m$ as the innermost optimization in both [34, equations (25) and (26)]. With the additional supremum, the explanation provided on [34, p. 931] is no longer valid. Considering Appendix A, we do not believe that the proof in [34] can be salvaged without introducing major new ideas, such as composition based expurgations similar to [10] or codeword cost based expurgations similar to [18]. In short, neither [31] nor [34] successfully proved the SPB for stationary memoryless channels even for the finite input set case.

In the rest of this section, we introduce our notation and channel model and define the channel codes with list decoding. In Section 2, we first present a brief review of the Rényi divergence, Augustin information measures, and the SPE; then, we derive Augustin's variant of Gallager's bound. In Section 3, we first state our main asymptotic results—i.e., SPBs given in Theorem 1 and Theorem 2—and then derive the nonasymptotic SPBs implying them. In Section 4, we derive the SPE for particular Gaussian and Poisson channels and confirm the equivalence of the definition invoked in Section 2 to the ones derived for these channels previously. In Section 5, we discuss why Augustin's method works briefly and compare our results with Augustin's in [24] and discuss applications of Augustin's method and use of Augustin's information measures in related problems.

1.1. Notational Conventions

For any two vectors μ and q in $\mathbb{R}^\ell$ their inner product, denoted by $\mu \cdot q$, is $\sum_{i=1}^{\ell} \mu^i q^i$. For any $\ell \in \mathbb{Z}_+$, the ℓ-dimensional vector whose all entries are one is denoted by $\mathbb{1}$, the dimension ℓ will be clear from the context. We denote the closure, interior, and convex hull of a set $\mathcal{S}$ by $\mathrm{cl}\,\mathcal{S}$, $\mathrm{int}\,\mathcal{S}$, and $\mathrm{ch}\,\mathcal{S}$, respectively; the relevant topology or vector space structure will be evident from the context.

For any set $\mathcal{Y}$, we denote the set of all probability mass functions that are nonzero only on finitely many members of $\mathcal{Y}$ by $\mathcal{P}(\mathcal{Y})$. For any $p \in \mathcal{P}(\mathcal{Y})$, we call the set of all y in $\mathcal{Y}$ for which $p(y) > 0$ the support of p and denote it by $\mathrm{supp}\,p$. For any measurable space $(\mathcal{Y}, \mathscr{Y})$, we denote the set of all probability measures on it by $\mathcal{P}(\mathscr{Y})$ and set of all finite measures by $\mathcal{M}^+(\mathscr{Y})$. We denote the integral of a measurable function f with respect to the measure μ by $\int f \, \mu(dy)$ or $\int f(y) \, \mu(dy)$. If the integral is on the real line and if it is with respect to the Lebesgue measure, we denote it by $\int f \, dy$ or $\int f(y) \, dy$ as well. If μ is a probability measure, then we also call the integral of f with respect μ the expectation of f or the expected value of f and denote it by $E_\mu[f]$ or $\mathbf{E}_\mu[f(Y)]$.

Our notation will be overloaded for certain symbols; however, the relations represented by these symbols will be clear from the context. We denote the Cartesian product of sets [35, p. 38] by $\times$. We use $|\cdot|$ to denote the absolute value of real numbers and the size of sets. The sign $\leq$ stands for the usual less than or equal to relation for real numbers and the corresponding point-wise inequity for functions and vectors. For two measures μ and q on the measurable space $(\mathcal{Y}, \mathscr{Y})$, $\mu \leq q$ if and only if $\mu(\mathcal{E}) \leq q(\mathcal{E})$ for all $\mathcal{E} \in \mathscr{Y}$. We denote the product of topologies [35, p. 38], σ-algebras [35, p. 118], and measures [35, Theorem 4.4.4] by $\otimes$. We use the shorthand $\mathcal{X}_1^n$ for the Cartesian product of sets $\mathcal{X}_1, \ldots, \mathcal{X}_n$ and $\mathscr{Y}_1^n$ for the product of the σ-algebras $\mathscr{Y}_1, \ldots, \mathscr{Y}_n$.

1.2. Channel Model

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

$$W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y}). \tag{1}$$

$\mathcal{Y}$ is called *the output set*, and $\mathcal{Y}$ is called *the σ-algebra of the output events*. We denote the set of all channels from the input set $\mathcal{X}$ to the output space $(\mathcal{Y}, \mathcal{Y})$ by $\mathcal{P}(\mathcal{Y} \,|\, \mathcal{X})$. For any $p \in \mathcal{P}(\mathcal{X})$ and $W \in \mathcal{P}(\mathcal{Y} \,|\, \mathcal{X})$, $p \circledast W$ is the probability measure whose marginal on $\mathcal{X}$ is p and whose conditional distribution given x is $W(x)$. The structure described in (1) is not sufficient on its own to ensure the existence of a unique $p \circledast W$ with the desired properties for all $p \in \mathcal{P}(\mathcal{X})$ in general. The existence of a unique $p \circledast W$ is guaranteed for all $p \in \mathcal{P}(\mathcal{X})$ if W is a transition probability from $(\mathcal{X}, \mathcal{X})$ to $(\mathcal{Y}, \mathcal{Y})$, i.e., a member of $\mathcal{P}(\mathcal{Y} \,|\, \mathcal{X})$ rather than $\mathcal{P}(\mathcal{Y} \,|\, \mathcal{X})$.

A channel W is called a *discrete channel* if both $\mathcal{X}$ and $\mathcal{Y}$ are finite sets. For any $n \in \mathbb{Z}_+$ and channels $W_t \colon \mathcal{X}_t \to \mathcal{P}(\mathcal{Y}_t)$ for $t \in \{1, \ldots, n\}$, the *length n product channel* $W_{[1,n]} \colon \mathcal{X}_1^n \to \mathcal{P}(\mathcal{Y}_1^n)$ is defined via the following relation:

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

A channel $U \colon \mathcal{Z} \to \mathcal{P}(\mathcal{Y}_1^n)$ is called a *length n memoryless channel* if and only if there exists a product channel $W_{[1,n]}$ satisfying $U(z) = W_{[1,n]}(z)$ for all $z \in \mathcal{Z}$ and $\mathcal{Z} \subset \mathcal{X}_1^n$. A product channel is *stationary* if and only if $W_t = W$ for all $t \in \{1, \ldots, n\}$ for some $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$. For such a channel, we denote the composition (i.e., the empirical distribution, type) of each $x_1^n \in \mathcal{X}_1^n$ by $\Upsilon(x)$, where $\Upsilon(x) \in \mathcal{P}(\mathcal{X})$.

For any $\ell \in \mathbb{Z}_+$, an ℓ-dimensional *cost function* ρ is a function from the input set to $\mathbb{R}^\ell$ that is bounded from below, i.e., that is of the form $\rho \colon \mathcal{X} \to \mathbb{R}^\ell_{\geq z}$ for some $z \in \mathbb{R}$. We assume without loss of generality that[4]

$$\inf_{x \in \mathcal{X}} \rho^i(x) \geq 0, \quad \forall i \in \{1, \ldots, \ell\}.$$

We denote the set of all cost constraints that can be satisfied by some member of $\mathcal{X}$ by $\Gamma_\rho^{\mathrm{ex}}$ and the set of all cost constraints that can be satisfied by some member of $\mathcal{P}(\mathcal{X})$ by Γ_ρ:

$$\Gamma_\rho^{\mathrm{ex}} \triangleq \{\varrho \in \mathbb{R}^\ell_{\geq 0} : \exists x \in \mathcal{X} \text{ such that } \rho(x) \leq \varrho\},$$
$$\Gamma_\rho \triangleq \{\varrho \in \mathbb{R}^\ell_{\geq 0} : \exists p \in \mathcal{P}(\mathcal{X}) \text{ such that } \mathbf{E}_p[\rho] \leq \varrho\}.$$

Then both $\Gamma_\rho^{\mathrm{ex}}$ and Γ_ρ have nonempty interiors and Γ_ρ is the convex hull of $\Gamma_\rho^{\mathrm{ex}}$, i.e., $\Gamma_\rho = \mathrm{ch}\,\Gamma_\rho^{\mathrm{ex}}$.

A cost function on a product channel is said to be additive if and only if it can be written as the sum of cost functions defined on the component channels. Given $W_t \colon \mathcal{X}_t \to \mathcal{P}(\mathcal{Y}_t)$ and $\rho_t \colon \mathcal{X}_t \to \mathbb{R}^\ell_{\geq 0}$ for $t \in \{1, \ldots, n\}$, we denote the resulting additive cost function on $\mathcal{X}_1^n$ for the channel $W_{[1,n]}$ by $\rho_{[1,n]}$, i.e.,

$$\rho_{[1,n]}(x_1^n) = \sum_{t=1}^{n} \rho_t(x_t), \quad \forall x_1^n \in \mathcal{X}_1^n.$$

1.3. Codes With List Decoding

The pair (Ψ, Θ) is an (M, L) *channel code* on $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ if and only if

---

[4] Augustin [24, Section 33] has an additional hypothesis, $\bigvee_{x \in \mathcal{X}} \rho(x) \leq \mathbb{1}$, which excludes certain important cases such as the Gaussian channels.

- The *encoding function* Ψ is a function from the message set $\mathcal{M} \triangleq \{1, 2, \ldots, M\}$ to the input set $\mathcal{X}$;
- The *decoding function* Θ is a measurable function from the output space $(\mathcal{Y}, \mathcal{Y})$ 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 (Ψ, Θ) on $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, *the conditional error probability* P_{e}^m for $m \in \mathcal{M}$ and *the average error probability* P_{e} are defined as

$$P_{\mathrm{e}}^m \triangleq \mathbf{E}_{W(\Psi(m))}[\mathbb{1}_{\{m \notin \Theta(y)\}}],$$

$$P_{\mathrm{e}} \triangleq \frac{1}{M} \sum_{m \in \mathcal{M}} P_{\mathrm{e}}^m.$$

An encoding function Ψ, hence the corresponding code, is said to satisfy the cost constraint ϱ if and only if $\bigvee_{m \in \mathcal{M}} \rho(\Psi(m)) \leq \varrho$. An encoding function Ψ, hence the corresponding code, on a stationary product channel is said to satisfy an empirical distribution constraint $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$ if and only if the composition of all of the codewords are in $\mathcal{A}$, i.e., if and only if $\Upsilon(\Psi(m)) \in \mathcal{A}$ for all $m \in \mathcal{M}$.

2. PRELIMINARIES

The Rényi divergence, tilting, and Augustin's information measures are central to the analysis that we present in the following sections. We introduce these concepts in Sections 2.1 and 2.2; a more detailed discussion can be found in [26, 36]. In Section 2.3 we define the SPE and derive widely known properties of it for our general channel model. In Section 2.4 we derive Augustin's variant of Gallager's bound.

2.1. The Rényi Divergence and Tilting

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

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

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

For properties of the Rényi divergence, throughout the manuscript, we will refer to the comprehensive study provided by van Erven and Harremoës [36]. Note that 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 as well; see [36, Theorem 30]. That characterization is related to another key concept for our analysis: the tilted probability measure.

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

$$\frac{dw_\alpha^q}{d\nu} \triangleq e^{(1-\alpha)D_\alpha(w\|q)} \left(\frac{dw}{d\nu}\right)^\alpha \left(\frac{dq}{d\nu}\right)^{1-\alpha}. \tag{2}$$

The conditional Rényi divergence and the tilted channel are straightforward generalizations of the Rényi divergence and the tilted probability measure that will allow us to express certain relations succinctly throughout our analysis.

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

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

If $\exists q \in \mathcal{P}(\mathcal{Y})$ such that $Q(x) = q$ for all $x \in \mathcal{X}$, then we denote $D_\alpha(W \| Q | p)$ by $D_\alpha(W \| q | p)$.

Definition 4. For any $\alpha \in \mathbb{R}_+$, $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ and $Q \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, *the order α tilted channel* W_α^Q *is a function from* $\{x : D_\alpha(W(x) \| Q(x)) < \infty\}$ *to* $\mathcal{P}(\mathcal{Y})$ *given by*

$$\frac{dW_\alpha^Q(x)}{d\nu} \triangleq e^{(1-\alpha)D_\alpha(W(x)\|Q(x))} \left(\frac{dW(x)}{d\nu} \right)^\alpha \left(\frac{dQ(x)}{d\nu} \right)^{1-\alpha}. \tag{3}$$

If $\exists q \in \mathcal{P}(\mathcal{Y})$ such that $Q(x) = q$ for all $x \in \mathcal{X}$, then we denote W_α^Q by W_α^q.

The following operator $\mathrm{T}_{\alpha,p}(\cdot)$ was considered implicitly by Fano [9, ch. 9], Haroutunian [10], and Polytrev [30] and explicitly by Augustin [24, Section 34], but only for orders less than one in all four manuscripts.

Definition 5. For any $\alpha \in \mathbb{R}_+$, $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, and $p \in \mathcal{P}(\mathcal{X})$, the order α Augustin operator for the input distribution p, i.e., $\mathrm{T}_{\alpha,p}(\cdot) \colon \mathcal{Q}_{\alpha,p} \to \mathcal{P}(\mathcal{Y})$, is given by

$$\mathrm{T}_{\alpha,p}(q) \triangleq \sum_x p(x) W_\alpha^q(x), \quad \forall q \in \mathcal{Q}_{\alpha,p}, \tag{4}$$

where $\mathcal{Q}_{\alpha,p} \triangleq \{q \in \mathcal{P}(\mathcal{Y}) : D_\alpha(W \| q | p) < \infty\}$ and the tilted channel W_α^q is defined in (3).

2.2. Augustin's Information Measures

Definition 6. For any $\alpha \in \mathbb{R}_+$, $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, and $p \in \mathcal{P}(\mathcal{X})$ *the order α Augustin information for the input distribution p is*

$$I_\alpha(p; W) \triangleq \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(W \| q | p). \tag{5}$$

The infimum in (5) is achieved by a unique probability measure denoted by $q_{\alpha,p}$ and called *the order α Augustin mean for the input distribution p*. Furthermore, the order α Augustin mean satisfies the following identities:

$$D_{1 \vee \alpha}(q_{\alpha,p} \| q) \geq D_\alpha(W \| q | p) - I_\alpha(p; W) \geq D_{1 \wedge \alpha}(q_{\alpha,p} \| q), \qquad \forall q \in \mathcal{P}(\mathcal{Y}),\ \alpha \in \mathbb{R}_+, \tag{6}$$

$$\mathrm{T}_{\alpha,p}(q_{\alpha,p}) = q_{\alpha,p}, \qquad \forall \alpha \in \mathbb{R}_+. \tag{7}$$

These observations are established in [26, Lemma 13(b)–(d)]; previously they were reported by Augustin [24, Lemma 34.2] for orders less than one. Throughout the manuscript, we refer to [26] for propositions about Augustin's information measures. A more detailed account of the previous work on Augustin's information measures can be found in [26] as well.

Definition 7. For any $\alpha \in \mathbb{R}_+$, $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, and $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$, *the order α Augustin capacity of W for the constraint set $\mathcal{A}$ is*

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

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

Using the definitions of the Augustin information and capacity, we get the following expression for $C_{\alpha,W,\mathcal{A}}$:

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

If $\mathcal{A}$ is convex, then the order of the supremum and the infimum can be changed as a result of [26, Theorem 1]:

$$\sup_{p \in \mathcal{A}} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(W \,\|\, q \,|\, p) = \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{p \in \mathcal{A}} D_\alpha(W \,\|\, q \,|\, p). \tag{8}$$

If in addition $C_{\alpha,W,\mathcal{A}}$ is finite, then [26, Theorem 1] implies that there exists a unique probability measure $q_{\alpha,W,\mathcal{A}}$, *called the order α Augustin center of W for the constraint set $\mathcal{A}$,* satisfying

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

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

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

For the constraint sets defined through cost constraints, we use the symbol $C_{\alpha,W,\varrho}$ rather than $C_{\alpha,W,\mathcal{A}(\varrho)}$ with a slight abuse of notation. In order to be able to apply convex conjugation techniques without any significant modifications, we extend the definition of the Augustin capacity to the infeasible cost constraints, i.e., to ϱ outside Γ_ρ, as follows:

$$C_{\alpha,W,\varrho} \triangleq \begin{cases} \sup_{p \in \mathcal{A}(\varrho)} I_\alpha(p; W) & \text{if } \varrho \in \Gamma_\rho, \\ -\infty & \text{if } \varrho \in \mathbb{R}^\ell_{\geq 0} \setminus \Gamma_\rho, \end{cases} \quad \forall \alpha \in \mathbb{R}_+.$$

In order to characterize $C_{\alpha,W,\varrho}$ through convex conjugation techniques, we first define Augustin–Legendre (A-L) information and capacity. These concepts are first introduced in [1, Section III-A] and [26, Section 5.2] as an extension of the analogous concepts in [12, ch. 8].

Definition 8. *For any $\alpha \in \mathbb{R}_+$, channel W of the form $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ with a cost function $\rho \colon \mathcal{X} \to \mathbb{R}^\ell_{\geq 0}$, $p \in \mathcal{P}(\mathcal{X})$, and $\lambda \in \mathbb{R}^\ell_{\geq 0}$, the order α Augustin–Legendre information for the input distribution p and the Lagrange multiplier λ is*

$$I_\alpha^\lambda(p; W) \triangleq I_\alpha(p; W) - \lambda \cdot \mathbf{E}_p[\rho].$$

Definition 9. *For any $\alpha \in \mathbb{R}_+$, channel W of the form $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ with a cost function $\rho \colon \mathcal{X} \to \mathbb{R}^\ell_{\geq 0}$, and $\lambda \in \mathbb{R}^\ell_{\geq 0}$ the order α Augustin–Legendre (A-L) capacity for the Lagrange multiplier λ is*

$$C_{\alpha,W}^\lambda \triangleq \sup_{p \in \mathcal{P}(\mathcal{X})} I_\alpha^\lambda(p; W).$$

Except for certain sign changes, $C_{\alpha,W}^\lambda$ is the convex conjugate of $C_{\alpha,W,\varrho}$ because of an analogous relation between $I_\alpha^\lambda(p; W)$ and $I_\alpha(p; W)$, see [26, equations (72)–(74) and (76)]:

$$C_{\alpha,W}^\lambda = \sup_{\varrho \geq 0} C_{\alpha,W,\varrho} - \lambda \cdot \varrho, \quad \forall \lambda \in \mathbb{R}^\ell_{\geq 0}.$$

Then $C_{\alpha,W,\varrho}$ can be expressed in terms of $C_{\alpha,W}^\lambda$ at least for the interior points of Γ_ρ:

$$C_{\alpha,W,\varrho} = \inf_{\lambda \geq 0} C_{\alpha,W}^\lambda + \lambda \cdot \varrho.$$

Furthermore, there exists a nonempty convex compact set of $\lambda_{\alpha,W,\varrho}$ satisfying $C_{\alpha,W,\varrho} = C_{\alpha,W}^{\lambda_{\alpha,W,\varrho}} + \lambda_{\alpha,W,\varrho} \cdot \varrho$ provided that $C_{\alpha,W,\varrho}$ is finite, by [26, Lemma 29].

On the other hand, using the definitions of $I_\alpha(p;W)$, $I_\alpha^\lambda(p;W)$, and $C_{\alpha,W}^\lambda$ we get the following expression for $C_{\alpha,W}^\lambda$:

$$C_{\alpha,W}^\lambda = \sup_{p \in \mathcal{P}(\mathcal{X})} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(W \,\|\, q \,|\, p) - \lambda \cdot \mathbf{E}_p[\rho].$$

$C_{\alpha,W}^\lambda$ satisfies a minimax relation similar to the one given in (8); see [26, Theorem 2]. That minimax relation, however, is better understood via the concept of Augustin–Legendre radius defined as follows.

Definition 10. For any $\alpha \in \mathbb{R}_+$, channel $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ with a cost function $\rho \colon \mathcal{X} \to \mathbb{R}_{\geq 0}^\ell$, and $\lambda \in \mathbb{R}_{\geq 0}^\ell$, the order α Augustin–Legendre radius of W for the Lagrange multiplier λ is

$$S_{\alpha,W}^\lambda \triangleq \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{x \in \mathcal{X}} D_\alpha(W(x) \,\|\, q) - \lambda \cdot \rho(x).$$

Then, as a result of [26, Theorem 2], for any $\alpha \in \mathbb{R}_+$, $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ with $\rho \colon \mathcal{X} \to \mathbb{R}_{\geq 0}^\ell$, and $\lambda \in \mathbb{R}_{\geq 0}^\ell$ we have

$$C_{\alpha,W}^\lambda = S_{\alpha,W}^\lambda. \tag{9}$$

If in addition $C_{\alpha,W}^\lambda$ is finite, then there exists a unique $q_{\alpha,W}^\lambda \in \mathcal{P}(\mathcal{Y})$, called *the order α Augustin–Legendre center of W for the Lagrange multiplier λ*, satisfying

$$C_{\alpha,W}^\lambda = \sup_{x \in \mathcal{X}} D_\alpha(W(x) \,\|\, q_{\alpha,W}^\lambda) - \lambda \cdot \rho(x).$$

The A-L information measures are defined through a standard application of the convex conjugation techniques. However, starting with [29, Theorems 8 and 10]—i.e., the cost constrained variants of Gallager's bound—the Rényi–Gallager (R-G) information measures rather than the A-L information measures have been the customary tools for applying convex conjugation techniques in the error exponent calculations; see, for example, [18–20]. A brief discussion of the R-G information measures can be found in Appendix B; for a more detailed discussion, see [26].

2.3. The Sphere Packing Exponent

Definition 11. For any $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$, and $R \in \mathbb{R}_{\geq 0}$, the SPE is

$$E_{\mathrm{sp}}(R, W, \mathcal{A}) \triangleq \sup_{\alpha \in (0,1)} \frac{1 - \alpha}{\alpha} \left(C_{\alpha,W,\mathcal{A}} - R \right). \tag{10}$$

We denote the $\mathcal{A} = \mathcal{P}(\mathcal{X})$ case by $E_{\mathrm{sp}}(R, W)$. Furthermore, with a slight abuse of notation, we denote the $\mathcal{A} = \{p\}$ case by $E_{\mathrm{sp}}(R, W, p)$, and the $\mathcal{A} = \{p \colon \mathbf{E}_p[\rho] \leq \varrho\}$ case by $E_{\mathrm{sp}}(R, W, \varrho)$.

Lemma 1. *For any $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ and $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$, $E_{\mathrm{sp}}(R, W, \mathcal{A})$ is nonincreasing and convex in R on $\mathbb{R}_{\geq 0}$, finite on $(C_{0^+,W,\mathcal{A}}, \infty)$, and continuous on $[C_{0^+,W,\mathcal{A}}, \infty)$, where $C_{0^+,W,\mathcal{A}} = \lim_{\alpha \downarrow 0} C_{\alpha,W,\mathcal{A}}$. In particular,*

$$E_{\mathrm{sp}}(R, W, \mathcal{A}) = \begin{cases} \infty, & R < C_{0^+,W,\mathcal{A}}, \\ \displaystyle\sup_{\alpha \in (0,1)} \frac{1-\alpha}{\alpha} \left(C_{\alpha,W,\mathcal{A}} - R \right), & R = C_{0^+,W,\mathcal{A}}, \\ \displaystyle\sup_{\alpha \in [\phi,1)} \frac{1-\alpha}{\alpha} \left(C_{\alpha,W,\mathcal{A}} - R \right), & R = C_{\phi,W,\mathcal{A}} \text{ for some } \phi \in (0,1), \\ 0, & R \geq C_{1,W,\mathcal{A}}. \end{cases} \tag{11}$$

Lemma 1 follows from the continuity and the monotonicity properties of $C_{\alpha,W,\mathcal{A}}$ established in [26, Lemma 23]; a proof can be found in Appendix C. The proof of Lemma 1 is analogous to that of [25, Lemma 13], which relies on [25, Lemma 8] instead of [26, Lemma 23].

One can express $E_{\mathrm{sp}}(R, W, \mathcal{A})$ in terms of $E_{\mathrm{sp}}(R, W, p)$ using the definitions of $C_{\alpha,W,\mathcal{A}}$, $E_{\mathrm{sp}}(R, W, \mathcal{A})$, and $E_{\mathrm{sp}}(R, W, p)$:

$$
\begin{aligned}
E_{\mathrm{sp}}(R, W, \mathcal{A}) &= \sup_{\alpha \in (0,1)} \sup_{p \in \mathcal{A}} \frac{1-\alpha}{\alpha} \left(I_\alpha(p; W) - R \right) \\
&= \sup_{p \in \mathcal{A}} \sup_{\alpha \in (0,1)} \frac{1-\alpha}{\alpha} \left(I_\alpha(p; W) - R \right) \\
&= \sup_{p \in \mathcal{A}} E_{\mathrm{sp}}(R, W, p).
\end{aligned}
\tag{12}
$$

Lemma 1 holds for $E_{\mathrm{sp}}(R, W, p)$ by definition, but it can be strengthened significantly for R in $(\lim_{\alpha \downarrow 0} I_\alpha(p; W), I_1(p; W)]$ using the elementary properties of the Augustin information.

Lemma 2. *Let* $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ *and* $p \in \mathcal{P}(\mathcal{X})$ *be such that* $I_{0^+}(p; W) \neq I_1(p; W)$, *where* $I_{0^+}(p; W) \triangleq \lim_{\alpha \downarrow 0} I_\alpha(p; W)$. *Then for any rate* $R \in (I_{0^+}(p; W), I_1(p; W)]$ *there exists a unique order* $\alpha^* \in (0, 1]$ *satisfying*

$$
R = I_1(p; W_{\alpha^*}^{q_{\alpha^*},p}).
\tag{13}
$$

The orders α^* *determined by* (13) *form an increasing continuous bijective function of rate* R, *from* $(I_{0^+}(p; W), I_1(p; W)]$ *to* $(0, 1]$ *satisfying*

$$
E_{\mathrm{sp}}(R, W, p) = D_1(W_{\alpha^*}^{q_{\alpha^*},p} \,\|\, W \,|\, p),
\tag{14}
$$

$$
\frac{\partial}{\partial R} E_{\mathrm{sp}}(R, W, p) = \frac{\alpha^* - 1}{\alpha^*}.
\tag{15}
$$

Thus, $E_{\mathrm{sp}}(R, W, p)$ *is finite, convex, continuously differentiable, and decreasing in* R *on the interval* $(I_{0^+}(p; W), I_1(p; W))$, *and it satisfies*

$$
E_{\mathrm{sp}}(I_{0^+}(p; W), W, p) = \lim_{\alpha \downarrow 0} D_1(W_\alpha^{q_{\alpha},p} \,\|\, W \,|\, p).
\tag{16}
$$

Furthermore, if $E_{\mathrm{sp}}(I_{0^+}(p; W), W, p)$ *is finite, then there exists a* $V \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ *satisfying both* $I_1(p; V) = I_{0^+}(p; W)$ *and* $D_1(V \,\|\, W \,|\, p) = E_{\mathrm{sp}}(I_{0^+}(p; W), W, p)$.

Proof. Note that $I_1(p; W_\alpha^{q_{\alpha},p})$ is an increasing and continuous function of the order α by [26, Lemma 17(a),(f)], because $I_{0^+}(p; W) \neq I_1(p; W)$ by the hypothesis. In addition, $\lim_{\alpha \downarrow 0} I_1(p; W_\alpha^{q_{\alpha},p}) = I_{0^+}(p; W)$ by [26, Lemma 17(g)]. Then there exists a unique α^* satisfying (13) by the intermediate value theorem [37, Theorem 4.23]. The function defined by (13) is an increasing continuous bijective function of the rate R from $(I_{0^+}(p; W), I_1(p; W)]$ to $(0, 1]$, because it is the inverse of an increasing continuous bijective function from $(0, 1]$ to $(I_{0^+}(p; W), I_1(p; W)]$, i.e., $\alpha \rightsquigarrow I_1(p; W_\alpha^{q_{\alpha},p})$.

On the other hand, $I_\alpha(p; W)$ is continuously differentiable in α by [26, Lemma 17(e)]; then [26, equations (35) and (46)] imply

$$
\frac{\partial}{\partial \alpha} \frac{1-\alpha}{\alpha} (I_\alpha(p; W) - R) = \frac{1}{\alpha^2} (R - I_1(p; W_\alpha^{q_{\alpha},p})).
\tag{17}
$$

Hence, for any $R \in (I_{0^+}(p; W), I_1(p; W)]$, the supremum in the definition of $E_{\mathrm{sp}}(R, W, p)$ is achieved at the order α^* satisfying (13). Then (14) follows from [26, equation (35)]. Furthermore, for any

$R \in (I_{0^+}(p;W), I_1(p;W))$ and $\overline{R} \geq 0$ we have

$$E_{\mathrm{sp}}(\overline{R}, W, p) \geq \frac{1 - \alpha^*(R)}{\alpha^*(R)}(I_{\alpha^*}(p;W) - \overline{R})$$

$$= E_{\mathrm{sp}}(R, W, p) + \frac{1 - \alpha^*(R)}{\alpha^*(R)}(R - \overline{R}). \qquad (18)$$

For any $\overline{R} \in (I_{0^+}(p;W), I_1(p;W))$ and $R \geq 0$, following a similar analysis and reversing the roles of R and $\overline{R}$, we obtain

$$E_{\mathrm{sp}}(R, W, p) \geq E_{\mathrm{sp}}(\overline{R}, W, p) + \frac{1 - \alpha^*(\overline{R})}{\alpha^*(\overline{R})}(\overline{R} - R). \qquad (19)$$

Since α^* is increasing and continuous in the rate, (18) and (19) imply (15) for all R in the interval $(I_{0^+}(p;W), I_1(p;W)]$.

For the $R = I_{0^+}(p;W)$ case, note that $\dfrac{1-\alpha}{\alpha}(I_\alpha(p;W) - R)$ is decreasing in α on $(0,1)$ by (17) and [26, Lemma 17(f),(g)]. Thus,

$$E_{\mathrm{sp}}(I_{0^+}(p;W), W, p) = \lim_{\alpha \downarrow 0} \frac{1-\alpha}{\alpha}(I_\alpha(p;W) - I_{0^+}(p;W)).$$

Then (16) follows from the mean value theorem [37, Theorem 5.10] and [26, equation (46)]. Furthermore, if $E_{\mathrm{sp}}(I_{0^+}(p;W), W, p) = \gamma$ for a $\gamma \in \mathbb{R}_+$, then $D_1(W_\alpha^{q_\alpha,p}(x) \| W(x)) \leq \dfrac{\gamma}{p(x)}$ as a result of nonnegativity of the Rényi divergence. Hence,

$$\int G\left(\frac{dW_\alpha^{q_\alpha,p}(x)}{dW(x)}\right) W(dy \mid x) \leq \frac{\gamma}{p(x)} + \frac{1}{e} + 1$$

for $G(\tau) = \tau \mathbb{1}_{\{0 \leq \tau < e\}} + \tau \ln \tau \mathbb{1}_{\{\tau \geq e\}}$, because $\tau \ln \tau \geq -1/e$. Then $\left\{\dfrac{dW_\alpha^{q_\alpha,p}(x)}{dW(x)}\right\}_{\alpha \in (0,1)}$ are uniformly $W(x)$-integrable by [38, Theorem 4.5.9], i.e., by the necessary and sufficient condition for the uniform integrability determined by de la Vallée Poussin. Thus, any sequence of members of $\left\{W_\alpha^{q_\alpha,p}(x)\right\}_{\alpha \in (0,1)}$ has a convergent subsequence for the topology of setwise convergence by [38, Theorem 4.7.25]. For each $x \in \operatorname{supp} p$, let $V(x)$ be the limit point for the aforementioned subsequence for the sequence $\left\{W_{1/\kappa}^{q_{1/\kappa},p}(x)\right\}_{\kappa \in \mathbb{Z}_+}$. Then (13), (14), and the lower semicontinuity of the Rényi divergence in its arguments for the topology of setwise convergence, i.e., [36, Theorem 15], imply $I_1(p;V) \leq I_{0^+}(p;W)$ and $D_1(V \| W \mid p) \leq \gamma$. On the other hand, as a result of the definition of $E_{\mathrm{sp}}(R, W, p)$ and [26, Lemma 13(e)], we have

$$E_{\mathrm{sp}}(R, W, p) = \sup_{\alpha \in (0,1)} \inf_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} D_1(V \| W \mid p) + \frac{1-\alpha}{\alpha}(I_1(p;V) - R)$$

$$\leq \sup_{\alpha \in (0,1)} D_1(V \| W \mid p) + \frac{1-\alpha}{\alpha}(I_1(p;V) - R)$$

$$= \begin{cases} D_1(V \| W \mid p), & R \geq I_1(p;V), \\ \infty, & R < I_1(p;V). \end{cases}$$

Thus, $I_1(p;V)$ cannot be less than $I_{0^+}(p;W)$, because $E_{\mathrm{sp}}(R, W, p)$ is infinite for all $R < I_{0^+}(p;W)$. Hence, $I_1(p;V) = I_{0^+}(p;W)$. Consequently, $D_1(V \| W \mid p)$ cannot be less than γ, because $E_{\mathrm{sp}}(I_{0^+}(p;W), W, p) = \gamma$. Hence, $D_1(V \| W \mid p) = \gamma$. $\triangle$

Lemma 2 provides a simple confirmation of the alternative expression for $E_{\mathrm{sp}}(R, W, p)$, which is commonly known as Haroutunian's form [10].

Lemma 3. *For any* $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $p \in \mathcal{P}(\mathcal{X})$, *and* $R \in \mathbb{R}_+$,

$$E_{\mathrm{sp}}(R, W, p) = \inf_{V \,:\, I_1(p;V) \leq R} D_1(V \,\|\, W \,|\, p). \tag{20}$$

Proof. If $R \in [I_1(p;W), \infty)$, then (20) holds, because the expression on the right-hand side of (20) is zero as a result of the substitution $V = W$ and the nonnegativity of the Rényi divergence.

On the other hand, as a result of the definition of $E_{\mathrm{sp}}(R, W, p)$, [26, Lemma 13(e)], and the max-min inequality we have

$$\begin{aligned}
E_{\mathrm{sp}}(R, W, p) &= \sup_{\alpha \in (0,1)} \; \inf_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} D_1(V \,\|\, W \,|\, p) + \frac{1-\alpha}{\alpha}(I_1(p;V) - R) \\
&\leq \inf_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} \; \sup_{\alpha \in (0,1)} D_1(V \,\|\, W \,|\, p) + \frac{1-\alpha}{\alpha}(I_1(p;V) - R) \\
&= \inf_{V \,:\, I_1(p;V) \leq R} D_1(V \,\|\, W \,|\, p).
\end{aligned}$$

Then (20) holds whenever $E_{\mathrm{sp}}(R, W, p)$ is infinite, i.e., for all $R \in [0, I_{0^+}(p;W))$ and possibly for $R = I_{0^+}(p;W)$, trivially and whenever $E_{\mathrm{sp}}(R, W, p)$ is finite as a result of Lemma 2. $\triangle$

Haroutunian's form implies the following sufficient condition for the optimality of an order α in the definition of the SPE given in (10).

Lemma 4. *For any* $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$, *and* $R \in (C_{0^+,W,\mathcal{A}}, C_{1,W,\mathcal{A}})$ *if there exists an* $\alpha \in (0,1)$ *and a function* V_p *of* p *from* $\mathcal{A}$ *to* $\mathcal{P}(\mathcal{Y}|\mathcal{X})$ *satisfying the two inequalities*

$$D_1(V_p \,\|\, q_{\alpha,W,\mathcal{A}} \,|\, p) \leq R, \qquad\qquad \forall p \in \mathcal{A}, \tag{21}$$

$$D_1(V_p \,\|\, W \,|\, p) \leq \frac{1-\alpha}{\alpha}(C_{\alpha,W,\mathcal{A}} - R), \qquad \forall p \in \mathcal{A}, \tag{22}$$

then $E_{\mathrm{sp}}(R, W, \mathcal{A}) = \dfrac{1-\alpha}{\alpha}(C_{\alpha,W,\mathcal{A}} - R)$.

For some channels, $V_p = W_\alpha^{q_{\alpha,W,\mathcal{A}}}$ satisfies both (21) and (22); for these channels, the value of SPE can be determined using Lemma 4. However, for an arbitrary channel, rate, and the corresponding optimal order α in (10), a V_p satisfying both (21) and (22) might not exist, e.g., for the Z-channel discussed in Appendix A. Had the sufficient condition for the optimality of the order α given in (21) and (22) been also necessary, Blahut's proof in [31] would have been correct; this, however, is not the case in general as we demonstrate in Appendix A. It is worth mentioning that for the channels satisfying the necessary conditions given (21) and (22), one can derive the SPB using the approach presented in [31].

Proof of Lemma 4. Note that as a result of (6) we have

$$D_1(V_p \,\|\, q_{\alpha,W,\mathcal{A}} \,|\, p) = I_1(p;V_p) + D_1\left(\sum_x p(x)V_p(x) \,\|\, q_{\alpha,W,\mathcal{A}}\right).$$

Thus, $I_1(p;V_p) \leq R$ for all $p \in \mathcal{A}$, because the Rényi divergence is nonnegative. Then $E_{\mathrm{sp}}(R, W, p) \leq \dfrac{1-\alpha}{\alpha}(C_{\alpha,W,\mathcal{A}} - R)$ for all $p \in \mathcal{A}$ by Lemma 3. Then $E_{\mathrm{sp}}(R, W, \mathcal{A}) \leq \dfrac{1-\alpha}{\alpha}(C_{\alpha,W,\mathcal{A}} - R)$ by (12). On the other hand, $E_{\mathrm{sp}}(R, W, \mathcal{A}) \geq \dfrac{1-\alpha}{\alpha}(C_{\alpha,W,\mathcal{A}} - R)$ by the definition. Thus, $E_{\mathrm{sp}}(R, W, \mathcal{A}) = \dfrac{1-\alpha}{\alpha}(C_{\alpha,W,\mathcal{A}} - R)$. $\triangle$

2.4. Augustin's Variant of Gallager's Bound

The SPE is an upper bound on the exponential decay rate of the optimal error probability with block length, i.e., on the reliability function, for memoryless channels satisfying rather mild

hypotheses, with or without the list decoding, as a result of the SPBs given in Theorems 1 and 2 of Section 3. For the list decoding, the SPE is also a lower bound on the exponential decay rate of the optimal error probability with block length [8, Exercise 5.20; 12, Exercise 10.28; 39]. The latter observation can be confirmed using standard results such as [29, 30], albeit with minor modifications, as well. In the following, we confirm this observation using a variant of Gallager's bound in terms of the Augustin information. Recall that Gallager's bound is derived, customarily, for the Rényi information rather than the Augustin information. The fixed point property described in (7) plays a critical role in the proof. We name this variant of Gallager's bound after Augustin, because [24, Lemma 36.1] of Augustin is the first achievability result making use of the fixed point property described in (7), to the best of our knowledge.

Lemma 5. *For any $M, L \in \mathbb{Z}_+$ such that $L < M$, $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $p \in \mathcal{P}(\mathcal{X})$, $\mathcal{B} \subset \mathcal{X}$, and $\alpha \in \left[\frac{1}{1+L}, 1\right)$ there exists an (M, L) channel code with an encoding function of the form $\Psi \colon \mathcal{M} \to \mathcal{B}$ satisfying*

$$\ln P_{\mathrm{e}} \le \frac{\alpha - 1}{\alpha} \left[\tau - \ln \frac{(M-1)e}{L}\right] - \frac{\ln p(\mathcal{B})}{\alpha}$$

where $\tau = \inf_{x \in \mathcal{B}} D_\alpha(W(x) \| q_{\alpha,p})$.

We do not present a separate proof for Lemma 5, because Lemma 5 is a zero Lagrange multiplier special case of Lemma 6 presented in the following. Before stating Lemma 6, let us point out an immediate consequence of Lemma 5 for the constant composition codes on the stationary memoryless channels. Recall that for any composition p for the block length n, the probability of the set of all composition p sequences (i.e., the probability of $\mathcal{T}_{p,n}$) for i.i.d. samples (i.e., $p_t = p$ for all t) satisfies the following identity for some $\xi \in [0, 1]$ by [12, p. 26]:

$$\left(\bigotimes_{t=1}^{n} p_t\right) (\mathcal{T}_{p,n}) = e^{-\xi \frac{|\mathrm{supp}\, p|}{12 \ln 2}} (2\pi n)^{-\frac{|\mathrm{supp}\, p|-1}{2}} \sqrt{\prod_{x \colon p(x)>0} \frac{1}{p(x)}}.$$

Then using $q_{\alpha, \bigotimes_{t=1}^{n} p_t} = \bigotimes_{t=1}^{n} q_{\alpha,p_t}$, established in [26, Lemma 14], we get the following corollary by setting $\mathcal{B}$ to $\mathcal{T}_{p,n}$.

Corollary 1. *For any $n \in \mathbb{Z}_+$, $\phi \in (0, 1)$, $\{W_t\}_{t \in \mathbb{Z}_+}$ satisfying $W_t = W$ for some $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $p \in \mathcal{P}(\mathcal{X})$ satisfying $np(x) \in \mathbb{Z}_{\ge 0}$ for all $x \in \mathcal{X}$, and integers M, L satisfying both $\frac{1}{L+1} < \eta$ and $\frac{1}{n} \ln \frac{M}{L} = I_\eta(p; W)$ for some $\eta \in [\phi, 1)$, there exists an (M, L) channel code on $\bigotimes_{t=1}^{n} W_t$ with an encoding function of the form $\Psi \colon \mathcal{M} \to \mathcal{T}_{p,n}$ satisfying*

$$\ln P_{\mathrm{e}} \le -n E_{\mathrm{sp}}\left(\frac{1}{n} \ln \frac{M}{L}, W, p\right)$$
$$+ \frac{1}{\eta}\left[1 - \eta + \frac{|\mathrm{supp}\, p|}{12 \ln 2} + \frac{|\mathrm{supp}\, p| - 1}{2} \ln(2\pi n) + \frac{1}{2} \sum_{x \colon p(x)>0} \ln p(x)\right].$$

Lemma 6. *For any $\ell, M, L \in \mathbb{Z}_+$ such that $L < M$, $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $\rho \colon \mathcal{X} \to \mathbb{R}_{\ge 0}^\ell$, $p \in \mathcal{P}(\mathcal{X})$, $\mathcal{B} \subset \mathcal{X}$, and $\alpha \in \left[\frac{1}{1+L}, 1\right)$ there exists an (M, L) channel code with an encoding function of the form $\Psi \colon \mathcal{M} \to \mathcal{B}$ such that*

$$\ln P_{\mathrm{e}} \le \frac{\alpha - 1}{\alpha}\left[\tau + \left(\inf_{x \in \mathcal{B}} \lambda \cdot \rho(x)\right) - \ln \frac{(M-1)e}{L}\right] - \frac{\ln p(\mathcal{B})}{\alpha}, \qquad (23)$$

where $\tau = \inf_{x \in \mathcal{B}} D_\alpha(W(x) \| q_{\alpha,p}) - \lambda \cdot \rho(x)$.

It is worth mentioning that one can obtain a similar bound in terms of the R-G information; see Lemma 12 in Appendix B. Before presenting the proof of Lemma 6, let us discuss its consequences briefly. For the cost constrained memoryless channels with additive cost functions one can apply Lemma 6 for $\lambda = 0$ and for the product input distributions satisfying the cost constraint $\varrho - \epsilon$. Then using the weak law of large numbers together with [26, Lemmas 14, 27, and 28] one can conclude that the reliability function is bounded from below by the SPE, under rather mild hypotheses. This observation can be strengthened for $\ell = 1$ case by invoking the Berry–Esseen theorem via [25, Lemma 19], under appropriate hypotheses. The following corollary can be generalized to the $\ell > 1$ case in a straightforward way once a result analogous to [25, Lemma 19] for sums of random vectors is obtained.

Corollary 2. *Let $n \in \mathbb{Z}_+$ such that $n \geq 21$, $\phi \in (0,1)$, $W_{[1,n]} \colon \mathfrak{X}_1^n \to \mathcal{P}(\mathcal{Y}_1^n)$ be a length n product channel with an additive cost function $\rho_{[1,n]} \colon \mathfrak{X}_1^n \to \mathbb{R}_{\geq 0}$ satisfying $\inf\limits_{x_t \in \mathfrak{X}_t} \rho_t(x_t) = 0$ for all $t \in [1,n]$ and*

$$C_{\alpha, \widetilde{W}_{[1,n]}, n\varrho} \geq C_{\alpha, W_{[1,n]}, n\varrho} - \epsilon, \quad \forall \alpha \in [\phi, 1), \ \varrho \in \left[\widetilde{\varrho}, \widetilde{\varrho} + \frac{9\varsigma}{n} \right], \tag{24}$$

for some $\widetilde{\varrho} \in \mathbb{R}_+$, $\varsigma \in \mathbb{R}_{\geq 0}$, $\epsilon \in \mathbb{R}_{\geq 0}$, and product channel $\widetilde{W}_{[1,n]} \colon \widetilde{\mathfrak{X}}_1^n \to \mathcal{P}(\mathcal{Y}_1^n)$ with the input sets $\widetilde{\mathfrak{X}}_t = \{ x_t \in \mathfrak{X}_t \colon \rho_t(x_t) \leq \varsigma \}$ satisfying $\widetilde{W}_{[1,n]}(x_1^n) = W_{[1,n]}(x_1^n)$ for all $x_1^n \in \widetilde{\mathfrak{X}}_1^n$. Then for any $\delta \in (0, \widetilde{\varrho})$, $\varrho \in \left(\widetilde{\varrho} + \frac{3e\varsigma}{n}, \widetilde{\varrho} + \frac{9\varsigma}{n} \right)$, and integers M, L satisfying both $\frac{1}{L+1} < \eta$ and $\ln \frac{M}{L} = C_{\eta, W_{[1,n]}, n\varrho}$ for some $\eta \in [\phi, 1)$, there exists an (M, L) channel code with an encoding function of the form $\Psi \colon \mathcal{M} \to \left\{ x_1^n \in \widetilde{\mathfrak{X}}_1^n \colon \rho_{[1,n]}(x_1^n) \leq n\varrho \right\}$ such that

$$\ln P_{\mathrm{e}} \leq -E_{\mathrm{sp}}\left(\ln \frac{M}{L}, W_{[1,n]}, n\varrho \right) + \frac{1 - \eta}{\eta} \left(\frac{C_{\eta, W_{[1,n]}, n\varrho}}{n} \frac{6\varsigma e}{\delta} + 2\epsilon + 1 \right) + \frac{\ln 4n}{2\eta} + \frac{1}{n}. \tag{25}$$

Note that if the region for ϱ had not depended on ς in (24) and the interval for α had been a compact subset of $(0,1)$, then one could have found a ς for any positive ϵ by invoking Dini's theorem [35, Theorem 2.4.10] through a construction similar to the one used in the proof of [27, Lemma 15(g)], because $C_{\alpha, W, \varrho}$ is continuous in (α, ϱ) on $(0,1) \times \mathbb{R}_+$ by [26, Lemma 27(c)]. This, however, is not the case, and there might be product channels for which (24) is not satisfied for some $\widetilde{\varrho}$ values by small enough ε values for any ς. Furthermore, even when the hypotheses of Corollary 2 are satisfied, the tightness of the bound given in (25) depends on the values of the constants satisfying the hypotheses. Nevertheless, it is easy to see that if either the cost functions of the component channels are all bounded or $C_{\phi, W_{[1,n]}, n\varrho}$ is $\Theta(n)$—i.e., scaling linearly with n—and (24) is satisfied for ϵ and ς that are $O(\ln n)$—i.e., scaling no faster than linearly with $\ln n$—then (25) has a prefactor that is $e^{-O(\ln n)}$, i.e., vanishing polynomially with n. The proof of Corollary 2 is presented in Appendix C.

Proof of Lemma 6. We establish the existence of the code with the desired properties through a random coding argument. For any $p \in \mathcal{P}(\mathfrak{X})$ and $\mathcal{B} \subset \mathfrak{X}$ satisfying $p(\mathcal{B}) > 0$, let $p_{\mathcal{B}} \in \mathcal{P}(\mathfrak{X})$ be

$$p_{\mathcal{B}}(x) \triangleq \frac{\mathbb{1}_{\{x \in \mathcal{B}\}} p(x)}{p(\mathcal{B})}, \quad \forall x \in \mathfrak{X}. \tag{26}$$

Let us consider an ensemble of codes in which the assignments of the messages to the elements of $\mathcal{B}$ are jointly independent and $\Psi(m) = x$ with probability $p_{\mathcal{B}}(x)$ for all m in the message set. The decoder chooses L messages with the greatest $\frac{f_{\Psi(m)}}{h_{\Psi(m)}}$ for f_x and h_x defined as

$$f_x(y) \triangleq \frac{dW(x)}{dq}(y), \qquad \forall x \in \mathfrak{X}, \ y \in \mathcal{Y}, \tag{27}$$

$$h_x \triangleq e^{\frac{\alpha-1}{\alpha} \lambda \cdot \rho(x)}, \qquad \forall x \in \mathfrak{X}, \tag{28}$$

where $q \in \mathcal{P}(\mathcal{Y})$ to be chosen later satisfies $q_{\alpha,p} \prec q$. When there is a tie, the decoder prefers the message or messages with the lower index among the tied messages. In order to bound the expected value of the error probability of the code over the ensemble, let us consider the expected value of the conditional error probability of the message with the greatest index. An error occurs only when $\frac{f}{h}$ associated with L or more other messages are at least as large as $\frac{f}{h}$ associated with the transmitted message. We can bound this probability using a y-dependent auxiliary threshold $\gamma(y)$ and obtain the following bound on the expected value of P_e over the ensemble[5]:

$$\mathbf{E}[P_e] \leq \sum_x p_{\mathcal{B}}(x) \, \mathbf{E}_q\Big[\mathbb{1}_{\{\frac{f_x}{h_x}\leq\gamma\}} f_x\Big]$$
$$+ \binom{M-1}{L} \sum_x p_{\mathcal{B}}(x) \, \mathbf{E}_q\bigg[\mathbb{1}_{\{\frac{f_x}{h_x}>\gamma\}}\Big[\sum_z p_{\mathcal{B}}(z)\mathbb{1}_{\{\frac{f_z}{h_z}\geq\frac{f_x}{h_x}\}}\Big]^L f_x\bigg]. \quad (29)$$

We invoke $\alpha \leq 1$ to bound the first term in the above sum, and $\alpha \geq \frac{1}{1+L}$ to bound the second term:

$$\mathbf{E}_q\Big[\mathbb{1}_{\{\frac{f_x}{h_x}\leq\gamma\}} f_x\Big] \leq h_x \, \mathbf{E}_q\bigg[\mathbb{1}_{\{\frac{f_x}{h_x}\leq\gamma\}} \Big(\frac{f_x}{h_x}\Big)^\alpha \gamma^{1-\alpha}\bigg],$$

$$\mathbf{E}_q\bigg[\mathbb{1}_{\{\frac{f_x}{h_x}>\gamma\}}\Big[\sum_z p_{\mathcal{B}}(z)\mathbb{1}_{\{\frac{f_z}{h_z}\geq\frac{f_x}{h_x}\}}\Big]^L f_x\bigg] \leq h_x \, \mathbf{E}_q\bigg[\mathbb{1}_{\{\frac{f_x}{h_x}>\gamma\}}\Big[\sum_z p_{\mathcal{B}}(z)\Big(\frac{f_z}{h_z}\Big)^\alpha\Big]^L \Big(\frac{f_x}{h_x}\Big)^{1-L\alpha}\bigg]$$

$$\leq h_x \, \mathbf{E}_q\bigg[\mathbb{1}_{\{\frac{f_x}{h_x}>\gamma\}}\Big[\sum_z p_{\mathcal{B}}(z)\Big(\frac{f_z}{h_z}\Big)^\alpha\Big]^L \Big(\frac{f_x}{h_x}\Big)^\alpha \gamma^{1-\alpha-L\alpha}\bigg].$$

If we set $\gamma = \Big[\sum_x p_{\mathcal{B}}(x)\Big(\frac{f_x}{h_x}\Big)^\alpha\Big]^{\frac{1}{\alpha}} \Big[\binom{M-1}{L}\Big]^{\frac{1}{L\alpha}}$, we get

$$\mathbf{E}[P_e] \leq \sum_x p_{\mathcal{B}}(x) h_x \, \mathbf{E}_q\bigg[\Big(\frac{f_x}{h_x}\Big)^\alpha \gamma^{1-\alpha}\bigg]$$

$$= \mathbf{E}_q\bigg[\Big[\sum_x p_{\mathcal{B}}(x)\Big(\frac{f_x}{h_x}\Big)^\alpha\Big]^{\frac{1}{\alpha}}\bigg] \Big(\sup_{x\in\mathcal{B}} h_x\Big) \Big[\binom{M-1}{L}\Big]^{\frac{1-\alpha}{L\alpha}}.$$

On the other hand, Stirling's approximation for the factorials, i.e., $\sqrt{2\pi n}(n/e)^n \leq n! \leq e\sqrt{n}(n/e)^n$, and the identity $\ln z \leq z - 1$ imply that

$$\frac{1}{L}\ln\binom{M-1}{L} \leq \frac{1}{L}\ln\frac{e\sqrt{M-1}}{2\pi\sqrt{L(M-1-L)}} + \ln\frac{M-1}{L} + \frac{M-1-L}{L}\ln\Big(1+\frac{L}{M-1-L}\Big)$$

$$\leq \ln\frac{M-1}{L} + 1.$$

Then we get the following bound on the expected value of P_e over the ensemble:

$$\ln\mathbf{E}[P_e] \leq \ln\mathbf{E}_q\Bigg[\Big(\sum_x p_{\mathcal{B}}(x) e^{(1-\alpha)\lambda\cdot\rho(x)}\Big(\frac{dW(x)}{dq}\Big)^\alpha\Big)^{\frac{1}{\alpha}}\Bigg]$$

$$+ \frac{\alpha-1}{\alpha}\Big[\Big(\inf_{x\in\mathcal{B}}\lambda\cdot\rho(x)\Big) - \ln\frac{(M-1)e}{L}\Big]. \quad (30)$$

---

[5] Note that $\sum_{t=L}^{M}\binom{M}{t}s^t(1-s)^{M-t} = \binom{M}{L}s^L\sum_{t=0}^{M-L}\frac{L!(M-L)!}{(L+t)!(M-L-t)!}s^t(1-s)^{M-L-t} \leq \binom{M}{L}s^L$ for all $s \in [0,1]$.

If we set $q = q_{\alpha,p}$, using first the definition of the tilted channel given in (3) and then the definition of τ, we get

$$
\begin{aligned}
\ln \mathbf{E}[P_{\mathrm{e}}] &\leq \ln \mathbf{E}_{q_{\alpha,p}} \left[\left(\sum_x p_{\mathcal{B}}(x) e^{(\alpha-1)[D_\alpha(W(x)\|q_{\alpha,p})-\lambda\cdot\rho(x)]} \frac{dW_\alpha^{q_{\alpha,p}}(x)}{dq_{\alpha,p}} \right)^{\frac{1}{\alpha}} \right] \\
&\quad + \frac{\alpha-1}{\alpha} \left[\left(\inf_{x\in\mathcal{B}} \lambda\cdot\rho(x) \right) - \ln \frac{(M-1)e}{L} \right] \\
&\leq \ln \mathbf{E}_{q_{\alpha,p}} \left[\left(\sum_x p_{\mathcal{B}}(x) \frac{dW_\alpha^{q_{\alpha,p}}(x)}{dq_{\alpha,p}} \right)^{\frac{1}{\alpha}} \right] + \frac{\alpha-1}{\alpha} \left[\tau + \left(\inf_{x\in\mathcal{B}} \lambda\cdot\rho(x) \right) - \ln \frac{(M-1)e}{L} \right].
\end{aligned}
$$

Since there exists a code with P_{e} less than or equal to $\mathbf{E}[P_{\mathrm{e}}]$, the existence of a code satisfying (23) with an encoding function of the form $\Psi\colon \mathcal{M} \to \mathcal{B}$ follows from (4), (7), and (26). △

3. SPHERE PACKING BOUNDS FOR MEMORYLESS CHANNELS

Assumption 1. *The maximum $C_{1/2,U_{\mathcal{A}}^{(t)}}$ for integers t less than or equal to n is $O(\ln n)$, i.e.,*

$$
\exists n_0 \in \mathbb{Z}_+, \ K \in \mathbb{R}_+ \quad such \ that \quad \max_{t\,:\,t\leq n} C_{1/2,U_{\mathcal{A}}^{(t)}} \leq K\ln(n), \quad \forall n \geq n_0,
$$

where $U_{\mathcal{A}}^{(t)}\colon \mathcal{X}_{\mathcal{A}}^{(t)} \to \mathcal{P}(\mathcal{Y})$ satisfies $U_{\mathcal{A}}^{(t)}(x) = W(x)$ for all $x \in \mathcal{X}_{\mathcal{A}}^{(t)}$ and $\mathcal{X}_{\mathcal{A}}^{(t)}$ is defined for any $t \in \mathbb{Z}_+$ and $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$ as

$$
\mathcal{X}_{\mathcal{A}}^{(t)} \triangleq \{x \in \mathcal{X} : \exists p \in \mathcal{A} \ such \ that \ p(z)t \in \mathbb{Z}, \ \forall z \in \mathcal{X} \ for \ which \ p(x) > 0\}.
$$

Theorem 1. *Let $\{W_t\}_{t\in\mathbb{Z}_+}$ be a stationary sequence of channels satisfying $W_t = W$ for all $t \in \mathbb{Z}_+$, $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$ be a convex constraint set satisfying Assumption 1, and ε, α_0, α_1 be positive parameters satisfying $0 < \alpha_0 < \alpha_1 < 1$. Then for any sequence of codes $\{(\Psi^{(n)},\Theta^{(n)})\}_{n\in\mathbb{Z}_+}$ on the product channels $\{W_{[1,n]}\}_{n\in\mathbb{Z}_+}$ satisfying*

$$
C_{\alpha_1,W,\mathcal{A}} \geq \frac{1}{n}\ln\frac{M_n}{L_n} \geq C_{\alpha_0,W,\mathcal{A}} + \frac{\ln n}{n}[\ln(\ln n)]^\varepsilon, \quad \forall n \geq n_0, \tag{31}
$$

and $\Upsilon(\Psi^{(n)}(m)) \in \mathcal{A}$ for all $m \in \mathcal{M}^{(n)}$, there exists a $\tau \in \mathbb{R}_+$ and an $n_1 \geq n_0$ such that

$$
P_{\mathrm{e}}^{(n)} \geq n^{-\tau} e^{-nE_{\mathrm{sp}}\left(\frac{1}{n}\ln\frac{M_n}{L_n},W,\mathcal{A}\right)}, \quad \forall n \geq n_1. \tag{32}
$$

It is worth mentioning that the hypotheses of Theorem 1 are satisfied by the Gaussian and Poisson models considered in [15–22]; thus, Theorem 1 implies an asymptotic SPB for these channels. The nonasymptotic counterpart of this implication is presented in Section 3.3, between Lemma 9 and Corollary 3.

For codes on the cost constrained memoryless channels, one can remove the stationarity hypothesis using the convex conjugation techniques [24, Theorem 36.6]. Theorem 2 strengthens [24, Theorem 36.6] by removing the bounded cost function hypothesis and by establishing the SPB with a prefactor that is polynomial in the block length n rather than a prefactor of the form $e^{-O(\sqrt{n})}$.

Assumption 2. *The maximum $C_{1/2,W_t,n\varrho}$ for integers t less than or equal to n is $O(\ln n)$.*

$$
\exists n_0 \in \mathbb{Z}_+, \ K \in \mathbb{R}_+ \quad such \ that \quad \max_{t\,:\,t\leq n} C_{1/2,W_t,n\varrho} \leq K\ln(n), \quad \forall n \geq n_0.
$$

Theorem 2. *Let $\{(W_t, \rho_t)\}_{t \in \mathbb{Z}_+}$ be a sequence of channels with associated cost functions satisfying Assumption 2, ε, α_0, α_1 be positive parameters satisfying $0 < \alpha_0 < \alpha_1 < 1$. Then on the product channels $\left\{W_{[1,n]}\right\}_{n \in \mathbb{Z}_+}$ with additive cost functions $\left\{\rho_{[1,n]}\right\}_{n \in \mathbb{Z}_+}$ of the form $\rho_{[1,n]}(x_1^n) = \sum\limits_{t=1}^{n} \rho_t(x_t)$, for any sequence of codes $\left\{(\Psi^{(n)}, \Theta^{(n)})\right\}_{n \in \mathbb{Z}_+}$ satisfying*

$$C_{\alpha_1, W_{[1,n]}, n\varrho} \geq \ln \frac{M_n}{L_n} \geq C_{\alpha_0, W_{[1,n]}, n\varrho} + \varepsilon(\ln n)^2, \quad \forall n \geq n_0, \tag{33}$$

and $\rho_{[1,n]}(\Psi^{(n)}(m)) \leq n\varrho$ for all $m \in \mathcal{M}^{(n)}$ for a per channel use cost constraint ϱ satisfying $n\varrho \in \operatorname{int} \Gamma_{\rho_{[1,n]}}$, there exists a $\tau \in \mathbb{R}_+$ and an $n_1 \geq n_0$ such that

$$P_{\mathrm{e}}^{(n)} \geq n^{-\tau} e^{-E_{\mathrm{sp}}\left(\ln \frac{M_n}{L_n}, W_{[1,n]}, n\varrho\right)}, \quad \forall n \geq n_1. \tag{34}$$

In order to establish the asymptotic SPB with a polynomial prefactor in Theorem 2, we have assumed that the order $1/2$ cost constrained Augustin capacity is scaling no faster than logarithmically with the cost. Although this hypothesis is true for most cases of interest, including various Gaussian and Poisson channels considered in [15–22] and many of their nonstationary variants, there do exist channels violating it; see [26, Example 1]. Thus, one might want to remove the logarithmic growth with the cost constraint hypothesis. One can do so by using Lemma 11 instead of Lemma 10 in the proof of Theorem 2.

Our ultimate aim in this section is to prove the two asymptotic SPBs given in Theorems 1 and 2, which constitute the main contribution of this article. To that end, we first provide an impossibility result for the hypothesis testing problem with independent samples using the Berry–Esseen theorem in Section 3.1. Then we introduce the concepts of averaged Augustin capacity and averaged SPE in Section 3.2. We derive nonasymptotic—but parametric—SPBs in terms of these averaged quantities for the composition constrained codes on stationary memoryless channels in Section 3.3 and for the cost constrained codes on (possibly nonstationary) memoryless channels in Section 3.4. The derivations of the asymptotic SPBs using the nonasymptotic ones and the proof of Lemma 8 of Section 3.2, which are rather straightforward, are presented in Appendix C.

3.1. An Impossibility Result For Hypothesis Testing

Lemma 7. *For any $\kappa \geq 3$, $\alpha \in (0, 1)$, $n \in \mathbb{Z}_+$, and product measures $w, q \in \mathcal{P}(\mathcal{Y}_1^n)$ of the form $w = \bigotimes\limits_{t=1}^{n} w_t$ and $q = \bigotimes\limits_{t=1}^{n} q_t$, let $\xi_t \triangleq \ln \frac{dw_{t,\mathrm{ac}}}{dq_t} - \mathbf{E}_{w_\alpha^q}\left[\ln \frac{dw_{t,\mathrm{ac}}}{dq_t}\right]$, where $w_{t,\mathrm{ac}}$ is the component of w_t that is absolutely continuous in q_t and*

$$g_\kappa \triangleq \left(\sum_{t=1}^{n} \mathbf{E}_{w_\alpha^q}[|\xi_t|^\kappa]\right)^{1/\kappa}.$$

Then any $\mathcal{E} \in \mathcal{Y}$ satisfying $q(\mathcal{E}) \leq \dfrac{1}{4\sqrt{n}} e^{-D_1(w_\alpha^q \| q) - \alpha 3 g_\kappa}$ also satisfies

$$w(\mathcal{Y}_1^n \setminus \mathcal{E}) \geq \frac{1}{4\sqrt{n}} e^{-D_1(w_\alpha^q \| w) - (1-\alpha) 3 g_\kappa}. \tag{35}$$

Lemma 7 provides an impossibility result for the hypothesis testing problem with independent samples in the spirit of [7, Theorem 5]. It, however, relies on the Berry–Esseen theorem via [25, Lemma 19] rather than the Chebyshev inequality.

Proof of Lemma 7. Let $\mathcal{E}_0 = \left\{ y : \left| \sum\limits_{t=1}^{n} \xi_t \right| \leq 3g_\kappa \right\}$; then

$$\left| \ln \frac{dw_\alpha^q}{dq}(y_1^n) - D_1(w_\alpha^q \| q) \right| \leq \alpha 3g_\kappa, \qquad \forall y_1^n \in \mathcal{E}_0,$$

$$\left| \ln \frac{dw_\alpha^q}{dw}(y_1^n) - D_1(w_\alpha^q \| w) \right| \leq (1-\alpha)3g_\kappa, \qquad \forall y_1^n \in \mathcal{E}_0.$$

Hence,

$$w_\alpha^q(\mathcal{E} \cap \mathcal{E}_0) \leq q(\mathcal{E} \cap \mathcal{E}_0)e^{D_1(w_\alpha^q \| q) + \alpha 3g_\kappa}, \tag{36}$$

$$w_\alpha^q(\mathcal{E} \cap \mathcal{E}_0) \leq w(\mathcal{E} \cap \mathcal{E}_0)e^{D_1(w_\alpha^q \| w) + (1-\alpha)3g_\kappa}. \tag{37}$$

On the other hand, as a result of the definition of $\mathcal{E}_0$ and [25, Lemma 19] we have

$$\mathbf{P}_{w_\alpha^q}[\mathcal{E}_0] \geq \frac{1}{2\sqrt{n}}. \tag{38}$$

Now (35) follows from (36)–(38). $\triangle$

3.2. Augustin's Averaging

Our nonasymptotic SPBs are expressed in terms of the averaged Augustin capacity and averaged SPE defined in the following for all $\epsilon \in (0,1)$ and $R \in \mathbb{R}_{\geq 0}$:

$$\widetilde{C}_{\alpha,W,\mathcal{A}}^\epsilon \triangleq \frac{1}{\epsilon} \int\limits_{\alpha - \epsilon\alpha}^{\alpha + \epsilon(1-\alpha)} \left[1 \vee \left(\frac{\alpha}{1-\alpha} \frac{1-\eta}{\eta} \right) \right] C_{\eta,W,\mathcal{A}} \, d\eta, \tag{39}$$

$$\widetilde{E}_{\mathrm{sp}}^\epsilon(R, W, \mathcal{A}) \triangleq \sup_{\alpha \in (0,1)} \frac{1-\alpha}{\alpha} (\widetilde{C}_{\alpha,W,\mathcal{A}}^\epsilon - R). \tag{40}$$

Note that $\lim\limits_{\epsilon \downarrow 0} \widetilde{C}_{\alpha,W,\mathcal{A}}^\epsilon = C_{\alpha,W,\mathcal{A}}$ for all $\alpha \in (0,1)$, because $C_{\alpha,W,\mathcal{A}}$ is continuous in α on $(0,1)$ by [26, Lemma 23(d)]. Furthermore, one can show that the convergence is uniform on compact subsets of $(0,1)$ using the monotonicity of $C_{\alpha,W,\mathcal{A}}$ and $\frac{1-\alpha}{\alpha} C_{\alpha,W,\mathcal{A}}$ in α, i.e., [26, Lemma 23(a),(b)]. Nevertheless, $C_{\alpha,W,\mathcal{A}}$ cannot be approximated by $\widetilde{C}_{\alpha,W,\mathcal{A}}^\epsilon$ uniformly on $(0,1)$ itself, because $\lim\limits_{\alpha \uparrow 1} \widetilde{C}_{\alpha,W,\mathcal{A}}^\epsilon = \infty$ for all positive ϵ whenever $C_{1/2,W,\mathcal{A}}$ is positive, even if $C_{1,W,\mathcal{A}}$ is finite. The latter observation follows from the monotonicity of $C_{\alpha,W,\mathcal{A}}$ and $\frac{1-\alpha}{\alpha} C_{\alpha,W,\mathcal{A}}$ in α:

$$\widetilde{C}_{\alpha,W,\mathcal{A}}^\epsilon \geq \frac{1}{\epsilon} \int\limits_{\alpha - \epsilon\alpha}^{\alpha - \frac{\epsilon}{2}\alpha} \frac{\alpha}{1-\alpha} \frac{1-\eta}{\eta} C_{\eta,W,\mathcal{A}} \, d\eta$$

$$\geq \frac{\alpha}{2-\epsilon} \left(1 + \frac{\alpha\epsilon}{2(1-\alpha)} \right) C_{\alpha-\epsilon\alpha,W,\mathcal{A}}.$$

As a function of the rate, the averaged SPE converges uniformly to the SPE on any compact set of rates less than the order one Augustin capacity, as is demonstrated by the following lemma.

Lemma 8. *For any* $W \colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$ *satisfying* $C_{1/2,W,\mathcal{A}} \in \mathbb{R}_+$, $\phi \in (0,1)$, $R \in [C_{\phi,W,\mathcal{A}}, \infty)$, *and* $\epsilon \in (0,\phi)$,

$$0 \leq \widetilde{E}_{\mathrm{sp}}^\epsilon(R, W, \mathcal{A}) - E_{\mathrm{sp}}(R, W, \mathcal{A}) \leq \frac{\epsilon}{1-\epsilon} \frac{R \vee E_{\mathrm{sp}}(R, W, \mathcal{A})}{\phi} \tag{41}$$

$$\leq \frac{\epsilon}{1-\epsilon} \frac{R}{\phi^2}. \tag{42}$$

Lemma 8 follows from the monotonicity of $C_{\alpha,W,\mathcal{A}}$ and $\frac{1-\alpha}{\alpha}C_{\alpha,W,\mathcal{A}}$ in α established in [26, Lemma 23(a),(b)]; a proof can be found in Appendix C. The proof of Lemma 8 is nearly identical to the proof of [25, Lemma 15], establishing the very same result for $\mathcal{A} = \mathcal{P}(\mathcal{X})$ case.

3.3. Nonasymptotic Sphere Packing Bounds for the Composition Constrained Codes

The composition constrained codes on the stationary memoryless channels have been analyzed extensively, but only for the finite input set case. It is often assumed that all codewords of the code have precisely the same composition. If that is not the case, one usually invokes an expurgation based on the compositions of the codewords and focuses on the most populous composition. However, such an expurgation leads to a nontrivial result only when the input set is finite. Bounds derived through the analysis of the most populous composition in such cases can be derived using the Augustin information and mean. The following lemma, on the other hand, use the concepts of the Augustin capacity and center together with Augustin's averaging in order to avoid assuming a finite input set or a composition that has a fraction of the codewords that is no smaller than a polynomial function of the block length.

Lemma 9. *Let* $n \in \mathbb{Z}_+$; $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$; $\mathcal{A}$ *be a convex subset of* $\mathcal{P}(\mathcal{X})$; κ, ϕ, ϵ *be positive parameters satisfying* $\kappa \geq 3$, $\phi < 1$, $\epsilon < 1$; *and* γ *be*

$$\gamma \triangleq 3\sqrt[\kappa]{3n}\left(\left[C_{1/2,U_{\mathcal{A}}^{(n)}} + (1-\phi)\ln(1+n)\right] \vee \kappa\right). \tag{43}$$

If $W_t = W$ *for all* $t \leq n$ *and* $M, L \in \mathbb{Z}_+$ *are such that* $\ln\frac{M}{L} > n\widetilde{C}_{\phi,W,\mathcal{A}}^{\epsilon} + \frac{\gamma}{1-\phi} + \ln\frac{8e^3n^{1.5}}{\epsilon}$, *then any* (M,L) *channel code on* $W_{[1,n]}$ *satisfying* $\Upsilon(\Psi(m)) \in \mathcal{A}$ *for all* $m \in \mathcal{M}$ *satisfies*

$$P_e \geq \left(\frac{\epsilon e^{-2\gamma}}{8e^3n^{1.5}}\right)^{1/\phi} e^{-n\widetilde{E}_{sp}^{\epsilon}\left(\frac{1}{n}\ln\frac{M}{L},W,\mathcal{A}\right)}. \tag{44}$$

The absence of a back-off term for the rate in (44) can be interpreted as an advantage; but this is possible only with a prefactor that makes the bound trivial, i.e., zero, as ϕ converges to zero. By changing the analysis slightly it is possible to obtain the following alternative bound:

$$P_e \geq \frac{\epsilon e^{-2\gamma}}{8n^{1.5}}e^{-n\widetilde{E}_{sp}^{\epsilon}(R,W,\mathcal{A})}, \quad R = \frac{1}{n}\ln\frac{M}{L} - \frac{1}{n}\ln\frac{8e^3n^{1.5}}{\epsilon e^{-2\gamma}}. \tag{45}$$

When considered together with the identity $C_{\alpha,U_{\mathcal{A}(\varrho)}^{(n)}} \leq C_{\alpha,W,n\varrho}$, Lemma 9 and (45) imply Corollary 3 and (46) given in the following.

Corollary 3. *Let* $n, \ell \in \mathbb{Z}_+$; $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$; $\rho\colon \mathcal{X} \to \mathbb{R}_{\geq 0}^{\ell}$; $\varrho \in \Gamma_\rho$; κ, ϕ, ϵ *be positive parameters satisfying* $\kappa \geq 3$, $\phi < 1$, $\epsilon < 1$; *and* γ *be*

$$\gamma \triangleq 3\sqrt[\kappa]{3n}\left(\left[C_{1/2,W,n\varrho} + (1-\phi)\ln(1+n)\right] \vee \kappa\right).$$

If $W_t = W$ *for all* $t \leq n$ *and* $M, L \in \mathbb{Z}_+$ *are such that* $\ln\frac{M}{L} > n\widetilde{C}_{\phi,W,\varrho}^{\epsilon} + \frac{\gamma}{1-\phi} + \ln\frac{8e^3n^{1.5}}{\epsilon}$, *then any* (M,L) *channel code on* $W_{[1,n]}$ *satisfying* $\sum_{t=1}^{n}\rho(\Psi_t(m)) \leq n\varrho$ *for all* $m \in \mathcal{M}$ *satisfies*

$$P_e \geq \left(\frac{\epsilon e^{-2\gamma}}{8e^3n^{1.5}}\right)^{1/\phi} e^{-n\widetilde{E}_{sp}^{\epsilon}\left(\frac{1}{n}\ln\frac{M}{L},W,\varrho\right)}.$$

By changing the the analysis slightly it is possible to obtain the following alternative bound:

$$P_{\rm e} \geq \frac{\epsilon e^{-2\gamma}}{8n^{1.5}} e^{-n\widetilde{E}_{\rm sp}^{\epsilon}(R,W,\varrho)}, \quad R = \frac{1}{n}\ln\frac{M}{L} - \frac{1}{n}\ln\frac{8e^3 n^{1.5}}{\epsilon e^{-2\gamma}}. \tag{46}$$

Proof of Lemma 9 and equation (45). The constrained Augustin center $q_{\alpha,W,\mathcal{A}}$ is continuous in α on $(0,1)$ for the total variation topology on $\mathcal{P}(\mathcal{Y})$ by [26, Lemmas 23(d) and 24]. Thus, $q_{\cdot,W,\mathcal{A}}$ is a transition probability from $((0,1),\mathcal{B}((0,1)))$ to $(\mathcal{Y},\mathcal{Y})$. For each $t \leq n$, we define the averaged center $q_{\alpha,t}^{\epsilon}$ as the $\mathcal{Y}$ marginal of the probability measure $u_{\alpha,\epsilon} \circledast q_{\cdot,W,\mathcal{A}}$, where $u_{\alpha,\epsilon}$ is the uniform probability distribution on $(\alpha - \epsilon\alpha, \alpha + \epsilon(1-\alpha))$:

$$q_{\alpha,t}^{\epsilon} \triangleq \frac{1}{\epsilon} \int\limits_{\alpha-\alpha\epsilon}^{\alpha+(1-\alpha)\epsilon} q_{\eta,W,\mathcal{A}}\, d\eta, \quad \forall t \in \{1,\ldots,n\}. \tag{47}$$

Let $q_{\alpha,t} \in \mathcal{P}(\mathcal{Y}_t)$ and $q_{\alpha} \in \mathcal{P}(\mathcal{Y}_1^n)$ be

$$q_{\alpha,t} \triangleq \frac{n}{n+1}q_{\alpha,t}^{\epsilon} + \frac{1}{n+1}q_{1/2,U_{\mathcal{A}}^{(n)}}, \quad \forall t \in \{1,\ldots,n\},$$

$$q_{\alpha} \triangleq \bigotimes_{t=1}^{n} q_{\alpha,t}.$$

Let us denote the probability measure generated by $\Psi(m)$, i.e., $\bigotimes_{t=1}^{n} W(\Psi_t(m))$, by w^m for brevity. We have

$$D_{\alpha}(w^m \| q_{\alpha}) \overset{(a)}{=} \sum_{t=1}^{n} D_{\alpha}\big(W(\Psi_t(m)) \| q_{\alpha,t}\big)$$

$$\overset{(b)}{\leq} \sum_{t=1}^{n} \left(\ln\frac{n+1}{n} + D_{\alpha}(W(\Psi_t(m)) \| q_{\alpha,t}^{\epsilon})\right)$$

$$\overset{(c)}{\leq} 1 + \sum_{t=1}^{n} D_{\alpha}\big(W(\Psi_t(m)) \| q_{\alpha,t}^{\epsilon}\big)$$

$$\overset{(d)}{\leq} 1 + \sum_{t=1}^{n} \frac{1}{\epsilon} \int\limits_{\alpha(1-\epsilon)}^{\alpha+(1-\alpha)\epsilon} D_{\alpha}\big(W(\Psi_t(m)) \| q_{\eta,W,\mathcal{A}}\big)\, d\eta$$

$$\overset{(e)}{=} 1 + \frac{n}{\epsilon} \int\limits_{\alpha(1-\epsilon)}^{\alpha+(1-\alpha)\epsilon} D_{\alpha}\big(W \| q_{\eta,W,\mathcal{A}} \,|\, \Upsilon(\Psi(m))\big)\, d\eta$$

$$\overset{(f)}{\leq} 1 + \frac{n}{\epsilon} \int\limits_{\alpha(1-\epsilon)}^{\alpha+(1-\alpha)\epsilon} \left(1 \vee \frac{1-\eta}{\eta}\frac{\alpha}{1-\alpha}\right) D_{\eta}\big(W \| q_{\eta,W,\mathcal{A}} \,|\, \Upsilon(\Psi(m))\big)$$

$$\overset{(g)}{\leq} 1 + \frac{n}{\epsilon} \int\limits_{\alpha(1-\epsilon)}^{\alpha+(1-\alpha)\epsilon} \left(1 \vee \frac{1-\eta}{\eta}\frac{\alpha}{1-\alpha}\right) C_{\eta,W,\mathcal{A}}\, d\eta$$

$$\overset{(h)}{=} 1 + n\widetilde{C}_{\alpha,W,\mathcal{A}}^{\epsilon}, \tag{48}$$

where (a) is by [36, Theorem 28]; (b) by [26, Lemma 1], because $\frac{n}{n+1}q_{\alpha,t}^{\epsilon} \leq q_{\alpha,t}$; (c) by $\ln\tau \leq \tau-1$; (d) by the Jensen's inequality and [36, Theorem 12]; (e) by Definition 3 and the definition of Υ;

(f) by [36, Theorem 3 and Proposition 2]; (g) by [26, Theorem 1], because $\Upsilon(\Psi(m)) \in \mathcal{A}$; and ($h$) by the definition of $\widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}}$. Let the probability measure v^m_α be the order α tilted probability measure between w^m and q_α defined in (2). Then v^m_α is of the form $v^m_\alpha = \bigotimes\limits_{t=1}^{n} W^{q_{\alpha,t}}_\alpha(\Psi_t(m))$ as a result of product structure of w^m and q_α. Let the random variables $\xi^m_{\alpha,t}$ and ξ^m_α be

$$\xi^m_{\alpha,t} \triangleq \ln \frac{d[W(\Psi_t(m))]_{\mathrm{ac}}}{dq_{\alpha,t}} - \mathbf{E}_{v^m_\alpha}\left[\ln \frac{d[W(\Psi_t(m))]_{\mathrm{ac}}}{dq_{\alpha,t}}\right],$$

$$\xi^m_\alpha \triangleq \sum_{t=1}^{n} \xi^m_{\alpha,t},$$

where $[W(\Psi_t(m))]_{\mathrm{ac}}$ is the component of $W(\Psi_t(m))$ that is absolutely continuous in $q_{\alpha,t}$. Then for all $\kappa \in \mathbb{R}_+$ and $\alpha \in (0,1)$

$$\mathbf{E}_{v^m_\alpha}[|\xi^m_{\alpha,t}|^\kappa]^{1/\kappa} \overset{(a)}{\leq} 3^{1/\kappa} \frac{\left[(1-\alpha)D_\alpha(W(\Psi_t(m)) \,\|\, q_{\alpha,t})\right] \vee \kappa}{\alpha(1-\alpha)}$$

$$\overset{(b)}{\leq} 3^{1/\kappa} \frac{\left[(1-\alpha)D_\alpha\big(W(\Psi_t(m)) \,\|\, q_{1/2,U^{(n)}_\mathcal{A}}\big) + (1-\alpha)\ln(1+n)\right] \vee \kappa}{\alpha(1-\alpha)}$$

$$\overset{(c)}{\leq} 3^{1/\kappa} \frac{\left[D_{1/2}\big(W(\Psi_t(m)) \,\|\, q_{1/2,U^{(n)}_\mathcal{A}}\big) + (1-\alpha)\ln(1+n)\right] \vee \kappa}{\alpha(1-\alpha)}$$

$$\overset{(d)}{\leq} 3^{1/\kappa} \frac{\left[C_{1/2,U^{(n)}_\mathcal{A}} + (1-\alpha)\ln(1+n)\right] \vee \kappa}{\alpha(1-\alpha)},$$

where (a) is by [25, Lemma 17]; (b) by [26, Lemma 1], because $\frac{q_{1/2,U^{(n)}_\mathcal{A}}}{n+1} \leq q_{\alpha,t}$; ($c$) by [36, Theorem 3 and Proposition 2]; and (d) by [26, Theorem 1], because $\Psi_t(m) \in \mathcal{X}^{(n)}_\mathcal{A}$. Then, using the definition of γ given in (43), we get

$$\left[\sum_{t=1}^{n} \mathbf{E}_{v^m_\alpha}[|\xi^m_{\alpha,t}|^\kappa]\right]^{1/\kappa} \leq \frac{\gamma}{3\alpha(1-\alpha)}, \quad \forall \alpha \in [\phi, 1). \tag{49}$$

On the other hand, [36, Theorem 30] implies

$$D_1(v^m_\alpha \,\|\, q_\alpha) = D_\alpha(w^m \,\|\, q_\alpha) - \frac{\alpha}{1-\alpha} D_1(v^m_\alpha \,\|\, w^m), \quad \forall m \in \mathcal{M},\ \alpha \in (0,1). \tag{50}$$

Thus, we can bound $D_1(v^m_\alpha \,\|\, q_\alpha)$ using the nonnegativity of the Rényi divergence, i.e., [36, Theorem 8], and (48):

$$0 \leq D_1(v^m_\alpha \,\|\, q_\alpha) \leq 1 + n\widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}}.$$

Hence,

$$\lim_{\alpha \downarrow \phi} D_1(v^m_\alpha \,\|\, q_\alpha) + \frac{\gamma}{1-\alpha} < \ln \frac{M}{L} \frac{\epsilon}{8e^2 n^{1.5}},$$

$$\lim_{\alpha \uparrow 1} D_1(v^m_\alpha \,\|\, q_\alpha) + \frac{\gamma}{1-\alpha} = \infty.$$

$D_1(v^m_\alpha \,\|\, q_\alpha)$ is continuous in α by [25, Lemma 16], because q_α is continuous in α for the total variation topology[6] on $\mathcal{P}(\mathcal{Y})$. Then, as a result of the intermediate value theorem [37, Theorem 4.23],

---

[6] In particular, $\|q_\alpha - q_\eta\| \leq \sqrt{8 \ln \frac{\epsilon}{\epsilon - (1-\epsilon)|\alpha - \eta|}}$, because $\|q_\alpha - q_\eta\| \leq \sqrt{4D_{1/2}(q_\alpha \,\|\, q_\eta)}$ by [36, Theorem 31], $D_{1/2}(q_\alpha \,\|\, q_\eta) = nD_{1/2}(q_{\alpha,t} \,\|\, q_{\eta,t})$ by [36, Theorem 28] and the definition of q_α, $D_{1/2}(q_{\alpha,t} \,\|\, q_{\eta,t}) \leq 2\ln \frac{2}{2 - \|q_{\alpha,t} - q_{\eta,t}\|}$ by [26, equation (9)], and $\|q_{\alpha,t} - q_{\eta,t}\| \leq 2\frac{1-\epsilon}{\epsilon}|\eta - \alpha|$ by the definition of q_α.

for each $m \in \mathcal{M}$ there exists an α_m in $(\phi, 1)$ satisfying

$$\left(D_1(v_\alpha^m \,\|\, q_\alpha) + \frac{\gamma}{1-\alpha} \right)\bigg|_{\alpha=\alpha_m} = \ln \frac{M}{L} \frac{\epsilon}{8e^2 n^{1.5}}. \tag{51}$$

For any $K \in \mathbb{Z}_+$, there exists a $1/K$ long closed subinterval of $(0,1)$ with $\left\lceil \dfrac{M}{K} \right\rceil$ or more of these α_m. Let $[\eta, \eta + 1/K]$ be this interval and $\widetilde{q}$ and $\widetilde{q}_t$ be

$$\widetilde{q} \triangleq \bigotimes_{t=1}^{n} \widetilde{q}_t, \qquad \widetilde{q}_t = \frac{n}{n+1} q_{\widetilde{\alpha}, t}^{\widetilde{\epsilon}} + \frac{1}{n+1} q_{1/2, U_{\mathcal{A}}^{(n)}},$$

where $\widetilde{\epsilon}$ and $\widetilde{\alpha}$ are given by

$$\widetilde{\epsilon} = \frac{1}{K} + \epsilon \left(1 - \frac{1}{K} \right), \qquad \widetilde{\alpha} = \frac{1-\epsilon}{1-\widetilde{\epsilon}} \eta.$$

Then for all α in $\left[\eta, \eta + \dfrac{1}{K} \right]$, using the definition of the averaged center $q_{\alpha, t}^\epsilon$ given in (47), we get

$$q_{\alpha, t} \leq \frac{\widetilde{\epsilon}}{\epsilon} \widetilde{q}_t, \qquad q_\alpha \leq \left(\frac{\widetilde{\epsilon}}{\epsilon} \right)^n \widetilde{q}.$$

At least half of the messages with α_m in $[\eta, \eta + 1/K]$, i.e., at least $\left\lceil \dfrac{1}{2} \left\lceil \dfrac{M}{K} \right\rceil \right\rceil$ messages, satisfy $\widetilde{q}(\mathcal{E}_m) \leq 2 \dfrac{L}{\lceil M/K \rceil}$ as a result of Markov's inequality, because $\sum\limits_{m \in \widetilde{\mathcal{M}}} \widetilde{q}(\mathcal{E}_m) \leq L$ by the definition of list decoding, where $\widetilde{\mathcal{M}} = \{ m : \alpha_m \in [\eta, \eta + 1/K] \}$. Then at least $\left\lceil \dfrac{1}{2} \left\lceil \dfrac{M}{K} \right\rceil \right\rceil$ messages with α_m in $[\eta, \eta + 1/K]$ satisfy

$$q_{\alpha_m}(\mathcal{E}_m) \leq \frac{2LK}{M} \left(1 + \frac{1}{K} \frac{1-\epsilon}{\epsilon} \right)^n$$
$$\leq \frac{2LK}{M} \left(1 + \frac{1}{K\epsilon} \right)^n.$$

Note that $\dfrac{n}{\epsilon} > 1$, because we have assumed that $\epsilon < 1$ and $n \geq 1$. If we set K to be $K = \left\lfloor \dfrac{n}{\epsilon} \right\rfloor$ and use the identity $(1+\tau)^{1/\tau} < e$ together with (51), we get

$$q_{\alpha_m}(\mathcal{E}_m) \leq \frac{2LK}{M} \left(1 + \frac{1}{K\epsilon} \right)^{K\epsilon \frac{n}{K\epsilon}}$$
$$\leq \frac{2L}{M} \frac{n}{\epsilon} e^2$$
$$\leq \frac{1}{4\sqrt{n}} e^{-D_1\left(v_{\alpha_m}^m \,\|\, q_{\alpha_m} \right) - \frac{\gamma}{1-\alpha_m}}.$$

Then (49) and Lemma 7 imply

$$P_e^m \geq \frac{1}{4\sqrt{n}} e^{-D_1\left(v_{\alpha_m}^m \,\|\, w^m \right) - \frac{\gamma}{\alpha_m}}. \tag{52}$$

Using (48) and (50)–(52), we get

$$P_e^m \geq \frac{1}{4\sqrt{n}} \left(\frac{\epsilon}{8e^2 n^{1.5}} \frac{M}{L} \right)^{\frac{1-\alpha_m}{\alpha_m}} e^{-\frac{1-\alpha_m}{\alpha_m} \left(1 + n\widetilde{C}_{\alpha_m, W}^\epsilon \right) - \frac{2\gamma}{\alpha_m}}. \tag{53}$$

Hence, for all m satisfying (51), as a result of the definition of $\widetilde{E}_{\mathrm{sp}}^\epsilon(R, W, \mathcal{A})$ given in (40), we have

$$P_e^m \geq \frac{e^{-2\gamma/\phi}}{4\sqrt{n}} \left(\frac{\epsilon}{8e^3 n^{1.5}} \right)^{\frac{1-\phi}{\phi}} e^{-n\widetilde{E}_{\mathrm{sp}}^\epsilon(R, W, \mathcal{A})}, \qquad R = \frac{1}{n} \ln \frac{M}{L}.$$

Since there are at least $\left\lceil \frac{1}{2}\left\lceil \frac{M}{K}\right\rceil\right\rceil$ such messages and $\left\lceil \frac{1}{2}\left\lceil \frac{M}{K}\right\rceil\right\rceil \geq \frac{M\epsilon}{2n}$ by the construction, we get the lower bound given in (44).

Note that $\frac{\epsilon e^{-2\gamma}}{8e^3 n^{1.5}} < 1$, because $\epsilon \in (0,1)$, $n \geq 1$, and $\gamma \geq 0$. Thus, the prefactor in (44) converges to zero as ϕ goes to zero. In order to avoid this phenomena, one can change the analysis after (53) and introduce an approximation error term to the rate of the averaged SPE term: For all for all m satisfying (51), as a result of the definition of $\widetilde{E}_{\mathrm{sp}}^{\epsilon}(R,W,\varrho)$ given in (40) and (53) we have

$$P_{\mathrm{e}}^m \geq \frac{e^{-2\gamma}}{4\sqrt{n}} e^{-n\widetilde{E}_{\mathrm{sp}}^{\epsilon}(R,W)}, \quad R = \frac{1}{n}\ln\frac{M}{L} - \frac{2\gamma}{n} - \frac{1}{n}\ln\frac{8e^3 n^{1.5}}{\epsilon}.$$

Since there are at least $\left\lceil \frac{1}{2}\left\lceil \frac{M}{K}\right\rceil\right\rceil$ such messages and $\left\lceil \frac{1}{2}\left\lceil \frac{M}{K}\right\rceil\right\rceil \geq \frac{M\epsilon}{2n}$ by the construction, we get the lower bound given in (45). △

3.4. Nonasymptotic Sphere Packing Bounds for the Cost Constrained Codes

In Section 3.3, we were primarily interested in the stationary memoryless channels with convex composition constraints; the cost constrained memoryless channels, considered in Corollary 3, were merely an afterthought. For the cost constrained memoryless channels it is possible to establish the SPB even for the nonstationary channels using the convex-conjugation techniques, as we will demonstrate in the following. The only drawback of using the convex-conjugation techniques is that we will be able to establish the SPB for the cost constraints in $\operatorname{int}\Gamma_\rho$, rather than the cost constraints in Γ_ρ.

Lemma 10. *Let* $n \in \mathbb{Z}_+$; $W_{[1,n]}\colon \mathcal{X}_1^n \to \mathcal{P}(\mathcal{Y}_1^n)$ *be a length* n *product channel with an additive cost function* $\rho_{[1,n]}\colon \mathcal{X}_1^n \to \mathbb{R}_{\geq 0}^\ell$ *satisfying* $\rho_{[1,n]}(x_1^n) = \sum\limits_{t=1}^{n} \rho_t(x_t)$ *for some* $\rho_t\colon \mathcal{X}_t \to \mathbb{R}_{\geq 0}^\ell$; ϱ *be a per channel use cost constraint satisfying* $n\varrho \in \operatorname{int}\Gamma_{\rho_{[1,n]}}$; κ, ϕ, ϵ *be arbitrary positive parameters satisfying* $\kappa \geq 3$, $\phi < 1$, $\epsilon < 1$; *and* γ *be*

$$\gamma \triangleq 3\left[3\sum_{t=1}^n \left(\left[C_{1/2,W_t,n\varrho} + (1-\phi)\ln(1+n)\right] \vee \kappa\right)^\kappa\right]^{1/\kappa}. \tag{54}$$

If M *and* L *are integers such that* $\ln\frac{M}{L} > \widetilde{C}_{\phi,W_{[1,n]},n\varrho}^{\epsilon} + \frac{\gamma}{1-\phi} + \ln\frac{8e^3 n^{1.5}}{\epsilon}$, *then any* (M,L) *channel code on* $W_{[1,n]}$ *satisfying* $\bigvee\limits_{m=1}^M \rho_{[1,n]}(\Psi(m)) \leq n\varrho$ *satisfies*

$$P_{\mathrm{e}} \geq \left(\frac{\epsilon e^{-2\gamma}}{8e^3 n^{1.5}}\right)^{1/\phi} e^{-\widetilde{E}_{\mathrm{sp}}^{\epsilon}\left(\ln\frac{M}{L}, W_{[1,n]}, n\varrho\right)}. \tag{55}$$

The bound given in (55) does not have a back-off term for the rate in the averaged SPE but it becomes trivial, i.e., zero, as ϕ converges to zero. By changing the analysis slightly it is possible to obtain the following alternative bound:

$$P_{\mathrm{e}} \geq \frac{\epsilon e^{-2\gamma}}{8n^{1.5}} e^{-\widetilde{E}_{\mathrm{sp}}^{\epsilon}(R, W_{[1,n]}, n\varrho)}, \quad R = \ln\frac{M}{L} - \ln\frac{8e^3 n^{1.5}}{\epsilon e^{-2\gamma}}. \tag{56}$$

The bounds in Lemma 10 and (56) are good enough for proving the asymptotic SPB given in Theorem 2. However, the γ term can be improved by a more careful calculation. The resulting expression, however, includes an additional optimization.

Lemma 11. *Lemma* 10 *and* (56) *are valid as they are if the definition of* γ *given in* (54) *is replaced by the following one:*

$$\gamma \triangleq \max_{\substack{\{\varrho_t\}:\, \varrho_t \geq 0\, \forall t, \\ \sum_{t=1}^{n} \varrho_t \leq n\varrho}} 3\left[3\sum_{t=1}^{n}\left(\left[C_{1/2,W_t}^{\lambda} + \lambda \cdot \varrho_t + (1-\phi)\ln(1+n)\right] \vee \kappa\right)^{\kappa}\right]^{1/\kappa}, \tag{57}$$

where $\lambda = \lambda_{1/2,W_{[1,n]},n\varrho}$. *If* $W_t = W$ *and* $\rho_t = \rho$ *for all* $t \in [1,n]$, *then* γ *defined in* (57) *satisfies*

$$\gamma = \max_{\substack{\{\varrho_t\}:\, \varrho_t \geq 0\, \forall t, \\ \sum_{t=1}^{n} \varrho_t \leq n\varrho}} 3\left[3\sum_{t=1}^{n}\left(\left[C_{1/2,W,\varrho} + \lambda \cdot (\varrho_t - \varrho) + (1-\phi)\ln(1+n)\right] \vee \kappa\right)^{\kappa}\right]^{1/\kappa}. \tag{58}$$

Proof of Lemma 10 and equation (56). For each $\alpha \in (0,1)$ there exists a $\lambda_\alpha \in \mathbb{R}_{\geq 0}^{\ell}$ satisfying

$$C_{\alpha,W_{[1,n]},n\varrho} = C_{\alpha,W_{[1,n]}}^{\lambda_\alpha} + \lambda_\alpha \cdot \varrho n \tag{59}$$

by [26, Lemma 29(c)], because $n\varrho \in \operatorname{int}\Gamma_{\rho_{[1,n]}}$. Hence, $q_{\alpha,W_{[1,n]},n\varrho} = q_{\alpha,W_{[1,n]}}^{\lambda_\alpha}$ by [26, Lemma 31]. Furthermore, $q_{\alpha,W_{[1,n]}}^{\lambda_\alpha}$ is of the form $q_{\alpha,W_{[1,n]}}^{\lambda_\alpha} = \bigotimes_{t=1}^{n} q_{\alpha,W_t}^{\lambda_\alpha}$ by [26, Lemma 32]. Then

$$\|q_{\alpha,W_t}^{\lambda_\alpha} - q_{\phi,W_t}^{\lambda_\phi}\|^2 \overset{(a)}{\leq} 4D_{1/2}(q_{\alpha,W_t}^{\lambda_\alpha} \| q_{\phi,W_t}^{\lambda_\phi})$$

$$\overset{(b)}{\leq} 4D_{1/2}(q_{\alpha,W_{[1,n]}}^{\lambda_\alpha} \| q_{\phi,W_{[1,n]}}^{\lambda_\phi})$$

$$\overset{(c)}{=} 4D_{1/2}(q_{\alpha,W_{[1,n]},n\varrho} \| q_{\phi,W_{[1,n]},n\varrho})$$

$$\overset{(d)}{\leq} 8\ln\frac{2}{2 - \|q_{\alpha,W_{[1,n]},\varrho} - q_{\phi,W_{[1,n]},\varrho}\|},$$

where (a) is by [36, Theorem 31], (b) by [36, Theorems 8 and 28], (c) because $q_{\alpha,W_{[1,n]},n\varrho} = q_{\alpha,W_{[1,n]}}^{\lambda_\alpha}$ $\forall \alpha \in (0,1)$, and (d) by [26, equation (9)]. On the other hand, the constrained Augustin center $q_{\alpha,W_{[1,n]},n\varrho}$ is continuous in α on $(0,1)$ for the total variation topology on $\mathcal{P}(\mathcal{Y}_1^n)$ by [26, Lemmas 23(d) and 24]. Thus, $q_{\alpha,W_t}^{\lambda_\alpha}$ is a continuous function of α for the total variation topology on $\mathcal{P}(\mathcal{Y}_t)$ as well. Then $q_{\cdot,W_t}^{\lambda}$ is a transition probability from $((0,1),\mathcal{B}((0,1)))$ to $(\mathcal{Y}_t,\mathcal{Y}_t)$. We define $q_{\alpha,t}^{\epsilon}$ as the $\mathcal{Y}_t$ marginal of the probability measure $u_{\alpha,\epsilon} \circledast q_{\cdot,W_t}^{\lambda_{\cdot,W,\varrho}}$, where $u_{\alpha,\epsilon}$ is the uniform probability distribution on $(\alpha - \epsilon\alpha, \alpha + \epsilon(1-\alpha))$:

$$q_{\alpha,t}^{\epsilon} = \frac{1}{\epsilon}\int_{\alpha - \alpha\epsilon}^{\alpha + (1-\alpha)\epsilon} q_{\eta,W_t}^{\lambda_\eta}\, d\eta, \quad \forall t \in \{1,\ldots,n\}.$$

Let $q_{\alpha,t} \in \mathcal{P}(\mathcal{Y}_t)$ and $q_\alpha \in \mathcal{P}(\mathcal{Y}_1^n)$ be

$$q_{\alpha,t} \triangleq \frac{n}{n+1}q_{\alpha,t}^{\epsilon} + \frac{1}{n+1}q_{1/2,W_t,n\varrho}, \quad \forall t \in \{1,\ldots,n\},$$

$$q_\alpha \triangleq \bigotimes_{t=1}^{n} q_{\alpha,t}.$$

Let us denote the probability measure generated by $\Psi(m)$, i.e., $\bigotimes_{t=1}^{n} W_t(\Psi_t(m))$, by w^m for brevity. We have

$$
\begin{aligned}
D_\alpha(w^m \,\|\, q_\alpha) &\overset{(a)}{=} \sum_{t=1}^{n} D_\alpha(W_t(\Psi_t(m)) \,\|\, q_{\alpha,t}) \\
&\overset{(b)}{\leq} \sum_{t=1}^{n} \left(\ln \frac{n+1}{n} + D_\alpha\big(W_t(\Psi_t(m)) \,\|\, q_{\alpha,t}^\epsilon\big) \right) \\
&\overset{(c)}{\leq} 1 + \sum_{t=1}^{n} D_\alpha\big(W_t(\Psi_t(m)) \,\|\, q_{\alpha,t}^\epsilon\big) \\
&\overset{(d)}{\leq} 1 + \sum_{t=1}^{n} \frac{1}{\epsilon} \int_{\alpha(1-\epsilon)}^{\alpha+(1-\alpha)\epsilon} D_\alpha\big(W_t(\Psi_t(m)) \,\|\, q_{\eta,W_t}^{\lambda_\eta}\big) \, d\eta \\
&\overset{(e)}{=} 1 + \frac{1}{\epsilon} \int_{\alpha(1-\epsilon)}^{\alpha+(1-\alpha)\epsilon} D_\alpha\big(w^m \,\|\, q_{\eta,W_{[1,n]}}^{\lambda_\eta}\big) \, d\eta \\
&\overset{(f)}{\leq} 1 + \frac{1}{\epsilon} \int_{\alpha(1-\epsilon)}^{\alpha+(1-\alpha)\epsilon} \left(1 \vee \frac{1-\eta}{\eta}\frac{\alpha}{1-\alpha} \right) D_\eta\big(w^m \,\|\, q_{\eta,W_{[1,n]}}^{\lambda_\eta}\big) \, d\eta \\
&\overset{(g)}{\leq} 1 + \frac{1}{\epsilon} \int_{\alpha(1-\epsilon)}^{\alpha+(1-\alpha)\epsilon} \left(1 \vee \frac{1-\eta}{\eta}\frac{\alpha}{1-\alpha} \right) \big(C_{\eta,W}^{\lambda_\eta} + \lambda_\eta \cdot \varrho n \big) \, d\eta \\
&\overset{(h)}{=} 1 + \widetilde{C}_{\alpha,W,\varrho}^\epsilon,
\end{aligned}
\tag{60}
$$

where (a) is by [36, Theorem 28]; (b) by [26, Lemma 1], because $\frac{n}{n+1} q_{\alpha,t}^\epsilon \leq q_{\alpha,t}$; (c) by $\ln \tau \leq \tau - 1$; (d) by the Jensen's inequality and [36, Theorem 12]; (e) by [36, Theorem 28] and [26, Lemma 32]; (f) by [36, Theorem 3 and Proposition 2]; (g) by [26, Theorem 2], because $\rho_{[1,n]}(\Psi(m)) \leq n\varrho$; and (h) by (39) and (59). Let v_α^m be the order α tilted probability measure between w^m and q_α defined in (2). Then v_α^m is a probability measure of the form $v_\alpha^m = \bigotimes_{t=1}^{n} [W_t]_\alpha^{q_{\alpha,t}}(\Psi_t(m))$ as a result of the product structure of w^m and q_α. Let the random variables $\xi_{\alpha,t}^m$ and ξ_α^m be

$$
\begin{aligned}
\xi_{\alpha,t}^m &\triangleq \ln \frac{d[W_t(\Psi_t(m))]_{\mathrm{ac}}}{dq_{\alpha,t}} - \mathbf{E}_{v_\alpha^m}\left[\ln \frac{d[W_t(\Psi_t(m))]_{\mathrm{ac}}}{dq_{\alpha,t}} \right], \\
\xi_\alpha^m &\triangleq \sum_{t=1}^{n} \xi_{\alpha,t}^m,
\end{aligned}
$$

where $[W_t(\Psi_t(m))]_{\mathrm{ac}}$ is the component of $W_t(\Psi_t(m))$ that is absolutely continuous in $q_{\alpha,t}$. Then for all $\kappa \in \mathbb{R}_+$ and $\alpha \in (0,1)$

$$
\begin{aligned}
\mathbf{E}_{v_\alpha^m}[|\xi_{\alpha,t}^m|^\kappa]^{1/\kappa} &\overset{(a)}{\leq} 3^{1/\kappa} \frac{[(1-\alpha)D_\alpha(W_t(\Psi_t(m)) \,\|\, q_{\alpha,t})] \vee \kappa}{\alpha(1-\alpha)} \\
&\overset{(b)}{\leq} 3^{1/\kappa} \frac{\left[(1-\alpha)D_\alpha\big(W_t(\Psi_t(m)) \,\|\, q_{1/2,W_t,n\varrho}\big) + (1-\alpha)\ln(1+n) \right] \vee \kappa}{\alpha(1-\alpha)} \\
&\overset{(c)}{\leq} 3^{1/\kappa} \frac{\left[D_{1/2}\big(W_t(\Psi_t(m)) \,\|\, q_{1/2,W_t,n\varrho}\big) + (1-\alpha)\ln(1+n) \right] \vee \kappa}{\alpha(1-\alpha)}
\end{aligned}
$$

$$\overset{(d)}{\leq} 3^{1/\kappa} \frac{\left[C_{1/2,W_t,n\varrho} + (1-\alpha)\ln(1+n)\right] \vee \kappa}{\alpha(1-\alpha)},$$

where (a) is by [25, Lemma 17]; (b) by [26, Lemma 1], because $\dfrac{q_{1/2,W_t,n\varrho}}{n+1} \leq q_{\alpha,t}$; (c) by [36, Theorem 3 and Proposition 2]; and (d) by [26, Theorem 1], because $\rho_t(\Psi_t(m)) \leq n\varrho$. Then, using the definition of γ given in (54), we get

$$\left[\sum_{t=1}^{n} \mathbf{E}_{v_\alpha^m}\left[|\xi_{\alpha,t}^m|^\kappa\right]\right]^{1/\kappa} \leq \frac{\gamma}{3\alpha(1-\alpha)}, \quad \forall \alpha \in [\phi,1). \tag{61}$$

The rest of the proof is analogous to the proof of Lemma 9 and (45) after (49). The differences that might worth mentioning are that bounds given in (60) and (61) are invoked instead of (48) and (49), and $\widetilde{q}_t = \dfrac{n}{n+1}q_{\alpha,t}^{\widetilde{\epsilon}} + \dfrac{1}{n+1}q_{1/2,W_t,\varrho}$ is used instead of $\widetilde{q}_t = \dfrac{n}{n+1}q_{\alpha,t}^{\widetilde{\epsilon}} + \dfrac{1}{n+1}q_{1/2,U_{\mathcal{A}}^{(n)}}$. $\triangle$

Proof of Lemma 11. The proof is identical to that of Lemma 10 and (56) except for the definition of $q_{\alpha,t}$ and the bound on $\mathbf{E}_{v_\alpha^m}\left[|\xi_{\alpha,t}^m|^\kappa\right]^{1/\kappa}$. In particular, we set $q_{\alpha,t} \in \mathcal{P}(\mathcal{Y}_t)$ to be $\dfrac{n}{n+1}q_{\alpha,t}^{\epsilon} + \dfrac{1}{n+1}q_{1/2,W_t}^{\lambda_{1/2}}$ for all $t \leq n$ and bound $\mathbf{E}_{v_\alpha^m}\left[|\xi_{\alpha,t}^m|^\kappa\right]^{1/\kappa}$ as follows:

$$\mathbf{E}_{v_\alpha^m}\left[|\xi_{\alpha,t}^m|^\kappa\right]^{1/\kappa} \overset{(a)}{\leq} 3^{1/\kappa} \frac{\left[(1-\alpha)D_\alpha(W_t(\Psi_t(m)) \,\|\, q_{\alpha,t})\right] \vee \kappa}{\alpha(1-\alpha)}$$

$$\overset{(b)}{\leq} 3^{1/\kappa} \frac{\left[(1-\alpha)D_\alpha\big(W_t(\Psi_t(m)) \,\|\, q_{1/2,W_t}^{\lambda_{1/2}}\big) + (1-\alpha)\ln(1+n)\right] \vee \kappa}{\alpha(1-\alpha)}$$

$$\overset{(c)}{\leq} 3^{1/\kappa} \frac{\left[D_{1/2}\big(W_t(\Psi_t(m)) \,\|\, q_{1/2,W_t}^{\lambda_{1/2}}\big) + (1-\alpha)\ln(1+n)\right] \vee \kappa}{\alpha(1-\alpha)}$$

$$\overset{(d)}{\leq} 3^{1/\kappa} \frac{\left[C_{1/2,W_t}^{\lambda_{1/2}} + \lambda_{1/2} \cdot \rho_t(\Psi_t(m)) + (1-\alpha)\ln(1+n)\right] \vee \kappa}{\alpha(1-\alpha)},$$

where (a) is by [25, Lemma 17]; (b) by [26, Lemma 1], because $\dfrac{q_{1/2,W_t}^{\lambda_{1/2}}}{n+1} \leq q_{\alpha,t}$; (c) by [36, Theorem 3 and Proposition 2]; and (d) by [26, Theorem 2]. Then, using the definition of γ given in (57), we get

$$\left[\sum_{t=1}^{n} \mathbf{E}_{v_\alpha^m}\left[|\xi_{\alpha,t}^m|^\kappa\right]\right]^{1/\kappa} \leq \frac{\gamma}{3\alpha(1-\alpha)}, \quad \forall \alpha \in [\phi,1).$$

The rest of the proof is identical to that of Lemma 10 and (56).

If $W_t = W$ and $\rho_t = \rho$ for all $t \in [1,n]$, then $C_{\alpha,W_{[1,n]},n\varrho} = nC_{\alpha,W,\varrho}$ for all $\varrho \in \Gamma_\rho$ and $C_{\alpha,W_{[1,n]}}^\lambda = nC_{\alpha,W}^\lambda$ for all $\lambda \in \mathbb{R}_{\geq 0}^\ell$ by [26, Lemmas 28 and 32]. Then (58) follows from $C_{\alpha,W,\varrho} = C_{\alpha,W}^{\lambda_{\alpha,W,\varrho}} + \lambda_{\alpha,W,\varrho} \cdot \varrho$, established in [26, Lemma 29(c)]. $\triangle$

4. EXAMPLES

As a result of Sections 2.4 and 3, we can conclude that the SPE governs the exponential decay rate of the error probability of channel codes with list decoding on memoryless channels under rather mild hypotheses. The calculation of the SPE itself, however, is a separate issue that is essential from a practical standpoint. In this section, we derive the SPE for various Gaussian and Poisson channels and demonstrate that it is possible to obtain parametric forms for these channels similar to the one given in Lemma 2 for $E_{\mathrm{sp}}(R,W,p)$. We believe these parametric forms are more straightforward and intuitive than commonly used equivalent parametric forms that were previously derived.

4.1. Gaussian Channels

We denote the probability density function of the zero mean Gaussian random variable with variance σ^2 by φ_{σ^2}, i.e.,

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

With a slight abuse of notation, we denote the corresponding probability measure on $\mathcal{B}(\mathbb{R})$ by φ_{σ^2} as well.

Example 1 (scalar Gaussian channel). Let W be the scalar Gaussian channel with noise variance σ^2 and the associated cost function ρ be the quadratic one:

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

$$\rho(x) = x^2, \qquad\qquad \forall x \in \mathbb{R}.$$

The cost constrained Augustin capacity and center of this channel are determined in [26, Example 4]:

$$C_{\alpha,W,\varrho} = \begin{cases} \dfrac{\alpha\varrho}{2(\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2)} + \dfrac{1}{\alpha - 1} \ln \dfrac{(\theta_{\alpha,\sigma,\varrho})^{\alpha/2}\sigma^{(1-\alpha)}}{\sqrt{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}}, & \alpha \in \mathbb{R}_+ \setminus \{1\}, \\[4mm] \dfrac{1}{2} \ln \left(1 + \dfrac{\varrho}{\sigma^2}\right), & \alpha = 1, \end{cases} \tag{62}$$

$$q_{\alpha,W,\varrho} = \varphi_{\theta_{\alpha,\sigma,\varrho}}, \tag{63}$$

$$\theta_{\alpha,\sigma,\varrho} \triangleq \sigma^2 + \frac{\varrho}{2} - \frac{\sigma^2}{2\alpha} + \sqrt{\left(\frac{\varrho}{2} - \frac{\sigma^2}{2\alpha}\right)^2 + \varrho\sigma^2}. \tag{64}$$

It is worth mentioning that $C_{\alpha,W,\varrho} = I_\alpha(\varphi_\varrho; W)$ and $q_{\alpha,W,\varrho} = q_{\alpha,\varphi_\varrho}$ for all positive orders α, i.e., zero mean Gaussian distribution with variance ϱ is the optimal input distribution for all orders. Thus, $E_{\mathrm{sp}}(R, W, \varrho) = E_{\mathrm{sp}}(R, W, \varphi_\varrho)$.

The SPE of the scalar Gaussian channel can be characterized using Lemma 4. To see how, first note that for any $\theta > 0$ and the corresponding the Gaussian probability measure φ_θ, the order α tilted channel $W_\alpha^{\varphi_\theta}$, defined in (3), is given by

$$W_\alpha^{\varphi_\theta}(\mathcal{E} \,|\, x) = \int_{\mathcal{E}} \varphi_{\frac{\sigma^2\theta}{\alpha\theta + (1-\alpha)\sigma^2}}\left(y - \frac{\alpha\theta}{\alpha\theta + (1-\alpha)\sigma^2}x\right) dy, \quad \forall \mathcal{E} \in \mathcal{B}(\mathbb{R}). \tag{65}$$

Since $\theta_{\alpha,\sigma,\varrho}$ is a root of the equality $\theta^2 - \theta\left[\varrho + \left(2 - \dfrac{1}{\alpha}\right)\sigma^2\right] + \left(1 - \dfrac{1}{\alpha}\right)\sigma^4 = 0$ for θ by [26, equation (132) and (133)], one can confirm using [26, equation (131)] by substitution that

$$D_1\big(W_\alpha^{\varphi_{\theta_{\alpha,\sigma,\varrho}}} \,\|\, \varphi_{\theta_{\alpha,\sigma,\varrho}} \,|\, p\big) = \frac{\alpha^2\theta_{\alpha,\sigma,\varrho}}{2(\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2)^2}(\mathbf{E}_p[\rho] - \varrho)$$

$$+ \frac{1}{2} \ln \frac{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}{\sigma^2}, \tag{66}$$

$$D_1\big(W_\alpha^{\varphi_{\theta_{\alpha,\sigma,\varrho}}} \,\|\, W \,|\, p\big) = \frac{(1-\alpha)^2\sigma^2}{2(\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2)^2}(\mathbf{E}_p[\rho] - \varrho)$$

$$+ \frac{(1-\alpha)\varrho}{2(\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2)} + \frac{1}{2} \ln \frac{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}{\theta_{\alpha,\sigma,\varrho}}. \tag{67}$$

Thus, for each $\alpha \in (0,1)$, $V_p = W_\alpha^{\varphi_{\theta_{\alpha,\sigma,\varrho}}}$ satisfies the hypotheses of Lemma 4 given in (21) and (22) for $R = \dfrac{1}{2} \ln \dfrac{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}{\sigma^2}$ as a result of (62) and the constraint $\mathbf{E}_p[\rho] \leq \varrho$. Furthermore,

$f(\alpha) \triangleq \frac{1}{2} \ln \frac{\alpha \theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}{\sigma^2}$ is a continuous and increasing function of α satisfying $\lim_{\alpha \downarrow 0} f(\alpha) = 0$ and $f(1) = C_{1,W,\varrho}$. Thus, the SPE can be written in the following parametric form in terms of $\alpha \in [0,1]$ for all rates in $[0, C_{1,W,\varrho}]$:

$$R = \frac{1}{2} \ln \frac{\alpha \theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}{\sigma^2}, \tag{68}$$

$$E_{\mathrm{sp}}(R, W, \varrho) = \frac{(1-\alpha)\varrho}{2(\alpha \theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2)} + \frac{1}{2} \ln \frac{\alpha \theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}{\theta_{\alpha,\sigma,\varrho}}. \tag{69}$$

Using (66) and (67), we can express both the rate and the SPE in terms of the tilted channel $W_\alpha^{\varphi_{\theta_{\alpha,\sigma,\varrho}}}$. On the other hand, $\varphi_{\theta_{\alpha,\sigma,\varrho}}$ is the output distribution for the input distribution φ_ϱ on $W_\alpha^{\varphi_{\theta_{\alpha,\sigma,\varrho}}}$, because $\varphi_{\theta_{\alpha,\sigma,\varrho}}$ is the Augustin mean $q_{\alpha,\varphi_\varrho}$ for the input distribution φ_ϱ satisfying the fixed point property $T_{\alpha,\varphi_\varrho}(q_{\alpha,\varphi_\varrho}) = q_{\alpha,\varphi_\varrho}$ as well. Thus, we can rewrite (68) and (69), using (66) and (67), as follows:

$$R = I_1(\varphi_\varrho; W_\alpha^{\varphi_{\theta_{\alpha,\sigma,\varrho}}}), \tag{70}$$

$$E_{\mathrm{sp}}(R, W, \varrho) = D_1(W_\alpha^{\varphi_{\theta_{\alpha,\sigma,\varrho}}} \| W \,|\, \varphi_\varrho). \tag{71}$$

To obtain an expression for the SPE that does not depend on $\theta_{\alpha,\sigma,\varrho}$ explicitly, we first note that (64) and (68) imply

$$\alpha = \frac{e^{2R} - 1}{2}\left(\sqrt{1 + \frac{4\sigma^2}{\varrho} \frac{e^{2R}}{e^{2R} - 1}} - 1\right). \tag{72}$$

On the other hand, $\varphi_{\theta_{\alpha,\sigma,\varrho}}$ is the output distribution for the input distribution φ_ϱ on channel $W_\alpha^{\varphi_{\theta_{\alpha,\sigma,\varrho}}}$. Thus, (65) implies

$$\frac{\sigma^2 \theta_{\alpha,\sigma,\varrho}}{\alpha \theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2} + \left(\frac{\alpha \theta_{\alpha,\sigma,\varrho}}{\alpha \theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}\right)^2 \varrho = \theta_{\alpha,\sigma,\varrho}.$$

Thus,

$$\frac{\sigma^2}{\theta_{\alpha,\sigma,\varrho}} = 1 - \frac{\alpha \varrho}{\alpha \theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}$$

$$= 1 - \frac{\varrho \alpha}{\sigma^2 e^{2R}}, \tag{73}$$

where (73) follows from (68).

Using first (68) and (73) in (69), and then invoking (72), we get the following expression for the SPE:

$$E_{\mathrm{sp}}(R, W, \varrho) = \frac{1}{2} \frac{(1-\alpha)\varrho}{\sigma^2 e^{2R}} + R + \frac{1}{2} \ln\left(1 - \frac{\alpha \varrho}{\sigma^2 e^{2R}}\right)$$

$$= \frac{\varrho}{4\sigma^2}\left[1 + \frac{1}{e^{2R}} - \left(1 - \frac{1}{e^{2R}}\right)\sqrt{1 + \frac{4\sigma^2}{\varrho} \frac{e^{2R}}{e^{2R} - 1}}\right]$$

$$+ \frac{1}{2} \ln\left[e^{2R} - \frac{\varrho}{\sigma^2}\left(\frac{e^{2R} - 1}{2}\right)\left(\sqrt{1 + \frac{4\sigma^2}{\varrho} \frac{e^{2R}}{e^{2R} - 1}} - 1\right)\right]. \tag{74}$$

The expression given in (74) for the SPE is equivalent to [8, equation (7.4.33)].

The parametric characterization given in (68) and (69) can be obtained by a more direct approach using the differentiability of $\theta_{\alpha,\sigma,\varrho}$ and $C_{\alpha,W,\varrho}$ in α as well. In particular, since $\theta_{\alpha,\sigma,\varrho}$ is a root of the

equality $\theta^2 - \theta\left[\varrho + \left(2 - \frac{1}{\alpha}\right)\sigma^2\right] + \left(1 - \frac{1}{\alpha}\right)\sigma^4 = 0$ for θ, we get the following closed form expression for the derivative of the Augustin capacity with respect to the order:

$$
\begin{aligned}
\frac{\partial}{\partial\alpha}C_{\alpha,W,\varrho} &= \frac{1}{2(1-\alpha)^2}\left[\frac{(1-\alpha)\varrho}{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2} + \ln\frac{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}{\theta_{\alpha,\sigma,\varrho}}\right] \\
&\quad + \frac{\theta_{\alpha,\sigma,\varrho}^2 - \theta_{\alpha,\sigma,\varrho}\left[\varrho + \left(2 - \frac{1}{\alpha}\right)\sigma^2\right] + \left(1 - \frac{1}{\alpha}\right)\sigma^4}{2(\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2)^2}\left[\frac{\alpha}{1-\alpha} + \frac{\alpha^2}{\theta_{\alpha,\sigma,\varrho}}\frac{\partial}{\partial\alpha}\theta_{\alpha,\sigma,\varrho}\right] \\
&= \frac{1}{2(1-\alpha)^2}\left[\frac{(1-\alpha)\varrho}{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2} + \ln\frac{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}{\theta_{\alpha,\sigma,\varrho}}\right]. \quad (75)
\end{aligned}
$$

Using first (75) and then (62), we get

$$
\begin{aligned}
&\frac{d}{d\alpha}\frac{1-\alpha}{\alpha}(C_{\alpha,W,\varrho} - R) \\
&= \frac{1-\alpha}{\alpha}\left(\frac{1}{2(1-\alpha)^2}\left[\frac{(1-\alpha)\varrho}{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2} + \ln\frac{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}{\theta_{\alpha,\sigma,\varrho}}\right]\right) - \frac{1}{\alpha^2}(C_{\alpha,W,\varrho} - R) \\
&= \frac{1}{\alpha^2}\left(R - \frac{1}{2}\ln\frac{\alpha\theta_{\alpha,\sigma,\varrho} + (1-\alpha)\sigma^2}{\sigma^2}\right).
\end{aligned}
$$

Then the derivative test implies the parametric form given in (68) and (69) as a result of (62).

Example 2 (parallel Gaussian channels). Let $W_{[1,n]}$ be the product of scalar Gaussian channels with noise variance σ_i^2 for $i \in \{1, \ldots, n\}$ and the cost function $\rho_{[1,n]}$ be additive and quadratic, i.e.,

$$
W_{[1,n]}(\mathcal{E} \,|\, x_1^n) = \int_{\mathcal{E}}\left[\prod_{i=1}^n \varphi_{\sigma_i^2}(y_i - x_i)\right]dy_1^n, \qquad \forall\mathcal{E} \in \mathcal{B}(\mathbb{R}^n),
$$

$$
\rho_{[1,n]}(x_1^n) = \sum_{i=1}^n x_i^2, \qquad\qquad \forall x_1^n \in \mathbb{R}^n.
$$

The constrained Augustin capacity and center of $W_{[1,n]}$ were determined in [26, Example 5]:

$$
C_{\alpha,W_{[1,n]},\varrho} = \sum_{i=1}^n C_{\alpha,W_i,\varrho_{\alpha,i}}, \tag{76}
$$

$$
q_{\alpha,W_{[1,n]},\varrho} = \bigotimes_{i=1}^n \varphi_{\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}}}, \tag{77}
$$

$$
\varrho_{\alpha,i} = \frac{\left|\alpha - 2\sigma_i^2\lambda_\alpha\right|^+}{2\lambda_\alpha(\alpha + 2(\alpha - 1)\sigma_i^2\lambda_\alpha)}, \tag{78}
$$

where $\theta_{\alpha,\sigma,\varrho}$ is defined in (64) and λ_α is determined by $\sum_i \varrho_{\alpha,i} = \varrho$ uniquely.[7] Furthermore, $\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}}$ can be expressed in terms of σ_i and λ_α without explicitly referring to $\varrho_{\alpha,i}$ as follows:

$$
\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}} = \sigma_i^2 + \left|\frac{1}{2\lambda_\alpha} - \frac{\sigma_i^2}{\alpha}\right|^+. \tag{79}
$$

---

[7] The constraint $\sum_i \varrho_{\alpha,i} = \varrho$ determines λ_α uniquely, because the expression on the right-hand side of (78) is a nonincreasing function of λ_α.

On the other hand, $\dfrac{d}{d\varrho_i}C_{\alpha,W_i,\varrho_i}\Big|_{\varrho_i=\varrho_{\alpha,i}} = \lambda_\alpha$ for all i with a positive $\varrho_{\alpha,i}$ and $\dfrac{d}{d\varrho_i}C_{\alpha,W_i,\varrho_i}\Big|_{\varrho_i=\varrho_{\alpha,i}} \le \lambda_\alpha$ for all i. Then, using the chain rule of derivatives together with (75) and (62), we get

$$
\begin{aligned}
&\frac{d}{d\alpha}\frac{1-\alpha}{\alpha}(C_{\alpha,W_{[1,n]},\varrho} - R) \\
&= \frac{1-\alpha}{\alpha}\sum_{i=1}^{n}\left[\frac{\partial}{\partial\phi}C_{\phi,W_i,\varrho_{\alpha,i}}\Big|_{\phi=\alpha} + \frac{\partial}{\partial\varrho}C_{\alpha,W_i,\varrho}\Big|_{\varrho=\varrho_{\alpha,i}}\frac{d}{d\alpha}\varrho_{\alpha,i}\right] - \frac{1}{\alpha^2}(C_{\alpha,W_{[1,n]},\varrho} - R) \\
&= \frac{1}{\alpha^2}\left[R - \frac{1}{2}\sum_{i=1}^{n}\ln\frac{\alpha\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}} + (1-\alpha)\sigma_i^2}{\sigma_i^2}\right] + \frac{1-\alpha}{\alpha}\lambda_\alpha\sum_{i=1}^{n}\frac{d}{d\alpha}\varrho_{\alpha,i} \\
&= \frac{1}{\alpha^2}\left[R - \frac{1}{2}\sum_{i=1}^{n}\ln\frac{\alpha\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}} + (1-\alpha)\sigma_i^2}{\sigma_i^2}\right].
\end{aligned}
$$

Thus, we obtain the following parametric form for $E_{\mathrm{sp}}(R,W_{[1,n]},\varrho)$ in terms of α for all $R \in [0, C_{\alpha,W_{[1,n]},\varrho}]$:

$$R = \frac{1}{2}\sum_{i=1}^{n}\ln\frac{\alpha\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}} + (1-\alpha)\sigma_i^2}{\sigma_i^2}, \tag{80}$$

$$E_{\mathrm{sp}}(R,W_{[1,n]},\varrho) = \frac{1}{2}\sum_{i=1}^{n}\left[\frac{(1-\alpha)\varrho_{\alpha,i}}{\alpha\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}} + (1-\alpha)\sigma_i^2} + \ln\frac{\alpha\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}} + (1-\alpha)\sigma_i^2}{\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}}}\right]. \tag{81}$$

Thus, $E_{\mathrm{sp}}(R,W_{[1,n]},\varrho) = D_1(V_\alpha \,\|\, W_{[1,n]} \,|\, \Phi_\alpha)$ for $R = I_1(\Phi_\alpha; V_\alpha)$ where the input distribution Φ_α is the zero mean Gaussian distribution with the diagonal covariance matrix whose eigenvalues are $\varrho_{\alpha,1}, \ldots, \varrho_{\alpha,n}$ and V_α is the order α tilted channel between $W_{[1,n]}$ and q_{α,Φ_α}.

If $\sigma_i^2 \ge \dfrac{\alpha}{2\lambda_\alpha}$ for an i, then $\varrho_{\alpha,i} = 0$, and the corresponding terms in the sums given in (80) and (81) are zero. In [19], Ebert provided an alternative parametric form for the SPE relying on this observation. To obtain Ebert's characterization, first note that

$$\alpha\theta_{\alpha,\sigma_i,\varrho_{\alpha,i}} + (1-\alpha)\sigma_i^2 = \frac{\alpha}{2\lambda_\alpha} \vee \sigma_i^2, \quad \forall i \in \{1,\ldots,n\}$$

by (79). Thus, (79)–(81) and the constraint $\sum_i \varrho_{\alpha,i} = \varrho$ imply the following parametric form in terms of $N = \dfrac{\alpha}{2\lambda_\alpha}$, which is equivalent to Ebert's characterization [8, equations (7.5.28), (7.5.32), and (7.5.34)], [18, p. 294], [19, equation (20)]:

$$R = \frac{1}{2}\sum_{i:\,\sigma_i^2 \le N}\ln\frac{N}{\sigma_i^2}, \tag{82}$$

$$\varrho = \frac{1}{\alpha}\sum_{i=1}^{n}\frac{|N - \sigma_i^2|^+}{1 + (\alpha-1)\frac{\sigma_i^2}{N}}, \tag{83}$$

$$E_{\mathrm{sp}}(R,W_{[1,n]},\varrho) = \frac{(1-\alpha)\varrho}{2N} + \frac{1}{2}\sum_{i:\,\sigma_i^2 \le N}\ln\frac{\alpha N}{N - (1-\alpha)\sigma_i^2}. \tag{84}$$

Note that one does not need to determine λ_α or invoke $N = \dfrac{\alpha}{2\lambda_\alpha}$ in the above expressions. One can first determine N using (82) and then determine α using (83) in order to determine $E_{\mathrm{sp}}(R,W_{[1,n]},\varrho)$, as was noted by Ebert in [18,19].

4.2. Poisson Channels

Let $T \in \mathbb{R}_+$ and $a, b \in \mathbb{R}_{\geq 0}$ such that $a \leq b$. Then for the Poisson channel $\Lambda \colon \mathcal{F} \to \mathcal{P}(\mathcal{Y})$, the input set $\mathcal{F}$ is the set of all measurable functions of the form $f \colon (0, T] \to [a, b]$, the output set $\mathcal{Y}$ is the set of all nondecreasing, right-continuous, integer valued functions on $(0, T]$, the σ-algebra of the output events $\mathcal{Y}$ is the Borel σ-algebra for the topology generated by the Skorokhod metric on $\mathcal{Y}$, and $\Lambda(f)$ is the Poisson point process with deterministic intensity function f for all $f \in \mathcal{F}$. With a slight abuse of notation, we denote the Poisson process with constant intensity γ by $\Lambda(\gamma)$. The cost function $\rho \colon \mathcal{F} \to \mathbb{R}_{\geq 0}$ is

$$\rho(f) \triangleq \frac{1}{T} \int_0^T f(t) \, dt.$$

In [27, Section V-C], the (unconstrained) Rényi capacities and centers of various Poisson channels are determined. These expressions are equal to the corresponding Augustin capacities and centers, because $C_{\alpha,W}^0 = C_{\alpha,W}^{g0}$ for any W and $q_{\alpha,W}^0 = q_{\alpha,W}^{g0}$ for any W with finite $C_{\alpha,W}^0$ by [26, Theorems 2 and 3].

Example 3 (Poisson channels with given average intensity). [27, Example 9] considers channels $\Lambda^\varrho \colon \mathcal{F}^\varrho \to \mathcal{P}(\mathcal{Y})$, where $\mathcal{F}^\varrho = \{f \in \mathcal{F} : \rho(f) = \varrho\}$ and $\Lambda^\varrho(f) = \Lambda(f)$ for all $f \in \mathcal{F}^\varrho$. The Rényi capacity and center of Λ^ϱ—hence the Augustin capacity and center of Λ^ϱ—are given in [27, equations (73) and (74)] to be

$$C_{\alpha,\Lambda^\varrho} = \begin{cases} \dfrac{\alpha}{\alpha - 1} (\zeta_{\alpha,\varrho} - \varrho) T, & \alpha \neq 1, \\[2mm] \left(\dfrac{\varrho - a}{b - a} b \ln \dfrac{b}{\varrho} + \dfrac{b - \varrho}{b - a} a \ln \dfrac{a}{\varrho} \right) T, & \alpha = 1, \end{cases} \tag{85}$$

$$q_{\alpha,\Lambda^\varrho} = \Lambda(\zeta_{\alpha,\varrho}), \tag{86}$$

$$\zeta_{\alpha,\varrho} \triangleq \left(\frac{\varrho - a}{b - a} b^\alpha + \frac{b - \varrho}{b - a} a^\alpha \right)^{1/\alpha}. \tag{87}$$

Since the expression for $C_{\alpha,\Lambda^\varrho}$ is differentiable in α, we obtain the following parametric expression for the SPE using the derivative test:

$$R = \left(\frac{\varrho - a}{b - a} b^\alpha \zeta_{\alpha,\varrho}^{1-\alpha} \ln \frac{b^\alpha \zeta_{\alpha,\varrho}^{1-\alpha}}{\zeta_{\alpha,\varrho}} + \frac{b - \varrho}{b - a} a^\alpha \zeta_{\alpha,\varrho}^{1-\alpha} \ln \frac{a^\alpha \zeta_{\alpha,\varrho}^{1-\alpha}}{\zeta_{\alpha,\varrho}} \right) T, \tag{88}$$

$$E_{\mathrm{sp}}(R, \Lambda^\varrho) = \left(\varrho - \zeta_{\alpha,\varrho} + \frac{\varrho - a}{b - a} b^\alpha \zeta_{\alpha,\varrho}^{1-\alpha} \ln \frac{b^\alpha \zeta_{\alpha,\varrho}^{1-\alpha}}{b} + \frac{b - \varrho}{b - a} a^\alpha \zeta_{\alpha,\varrho}^{1-\alpha} \ln \frac{a^\alpha \zeta_{\alpha,\varrho}^{1-\alpha}}{a} \right) T. \tag{89}$$

There is an alternative parametric characterization, which is considerably easier to remember in terms of the tilted channels. In order to derive that expression, first note that [27, equations (66) and (68)] and the definition of the tilted channel given in (3) imply that

$$\Lambda_\alpha^{\Lambda(g)}(f) = \Lambda\big(f^\alpha g^{(1-\alpha)}\big). \tag{90}$$

Then, using [27, equation (68)] and (86)–(89), we get

$$R = D_1\big(\Lambda_\alpha^{q_{\alpha,\Lambda^\varrho}}(f_{\mathrm{opt}}) \,\|\, q_{\alpha,\Lambda^\varrho}\big), \tag{91}$$

$$E_{\mathrm{sp}}(R, \Lambda^\varrho) = D_1\big(\Lambda_\alpha^{q_{\alpha,\Lambda^\varrho}}(f_{\mathrm{opt}}) \,\|\, \Lambda(f_{\mathrm{opt}})\big), \tag{92}$$

where f_{opt} is any $\{a, b\}$ valued function in $\mathcal{F}^\varrho$, i.e., any function $f_{\mathrm{opt}} \colon (0, T] \to \{a, b\}$ satisfying $\rho(f_{\mathrm{opt}}) = \varrho$.

Example 4 (Poisson channels with constrained average intensity). Let us first confirm that the constrained Augustin capacity $C_{\alpha,\Lambda,\varrho}$ and the constrained Augustin center $q_{\alpha,\Lambda,\varrho}$ are given by[8]

$$C_{\alpha,\Lambda,\varrho} = C_{\alpha,\Lambda^{\varrho \wedge \varrho_\alpha}}, \tag{93}$$

$$q_{\alpha,\Lambda,\varrho} = q_{\alpha,\Lambda^{\varrho \wedge \varrho_\alpha}}, \tag{94}$$

where $C_{\alpha,\Lambda^\varrho}$ and $q_{\alpha,\Lambda^\varrho}$ are determined by (85)–(87) and ϱ_α is a decreasing function of the order α defined as

$$\varrho_\alpha \triangleq \begin{cases} \alpha^{\frac{\alpha}{1-\alpha}} \left(\dfrac{b-a}{b^\alpha - a^\alpha}\right)^{\frac{1}{1-\alpha}} + \dfrac{ab^\alpha - ba^\alpha}{b^\alpha - a^\alpha}, & \alpha \neq 1, \\ e^{-1} b^{\frac{b}{b-a}} a^{-\frac{a}{b-a}}, & \alpha = 1. \end{cases} \tag{95}$$

To establish (93) and (94), first note that $C_{\alpha,\Lambda^{\varrho \wedge \varrho_\alpha}} \leq C_{\alpha,\Lambda,\varrho}$, because $C_{\alpha,\Lambda^{\varrho \wedge \varrho_\alpha}} = C_{\alpha,\Lambda,\mathcal{P}(\mathcal{F}^{\varrho \wedge \varrho_\alpha})}$ and $\mathcal{P}(\mathcal{F}^{\varrho \wedge \varrho_\alpha}) \subset \mathcal{A}(\varrho)$, where $\mathcal{A}(\varrho) = \{p \in \mathcal{P}(\mathcal{F}) : \mathbf{E}_p[\rho(f)] \leq \varrho\}$. On the other hand, invoking first [27, equation (82)] and then [27, equation (76)], we get

$$D_\alpha\big(\Lambda(f) \,\|\, \Lambda(\zeta_{\alpha,\varrho \wedge \varrho_\alpha})\big)$$
$$\leq \frac{b - \rho(f)}{b-a} D_\alpha\big(\Lambda(a) \,\|\, \Lambda(\zeta_{\alpha,\varrho \wedge \varrho_\alpha})\big) + \frac{\rho(f) - a}{b-a} D_\alpha\big(\Lambda(b) \,\|\, \Lambda(\zeta_{\alpha,\varrho \wedge \varrho_\alpha})\big)$$
$$= C_{\alpha,\Lambda^{\varrho \wedge \varrho_\alpha}} + \frac{\varrho \wedge \varrho_\alpha - \rho(f)}{b-a} \left[D_\alpha\big(\Lambda(a) \,\|\, \Lambda(\zeta_{\alpha,\varrho \wedge \varrho_\alpha})\big) - D_\alpha\big(\Lambda(b) \,\|\, \Lambda(\zeta_{\alpha,\varrho \wedge \varrho_\alpha})\big) \right], \quad \forall f \in \mathcal{F}.$$

Since $D_\alpha(\Lambda(a) \,\|\, \Lambda(\zeta_{\alpha,\varrho})) = D_\alpha(\Lambda(b) \,\|\, \Lambda(\zeta_{\alpha,\varrho}))$ by [27, equations (83) and (84)], we get

$$D_\alpha\big(\Lambda(f) \,\|\, \Lambda(\zeta_{\alpha,\varrho \wedge \varrho_\alpha})\big)$$
$$\leq C_{\alpha,\Lambda^{\varrho \wedge \varrho_\alpha}} + \mathbb{1}_{\{\varrho < \varrho_\alpha\}} \frac{\varrho - \rho(f)}{b-a} \left[D_\alpha\big(\Lambda(a) \,\|\, \Lambda(\zeta_{\alpha,\varrho})\big) - D_\alpha\big(\Lambda(b) \,\|\, \Lambda(\zeta_{\alpha,\varrho})\big) \right], \quad \forall f \in \mathcal{F}.$$

On the other hand, $D_\alpha(\Lambda(a) \,\|\, \Lambda(\zeta_{\alpha,\varrho})) \leq D_\alpha(\Lambda(b) \,\|\, \Lambda(\zeta_{\alpha,\varrho}))$ for all $\varrho \leq \varrho_\alpha$ by [27, equation (83)]; consequently, we have

$$D_\alpha\big(\Lambda(f) \,\|\, \Lambda(\zeta_{\alpha,\varrho \wedge \varrho_\alpha}) \,|\, p\big) \leq C_{\alpha,\Lambda^{\varrho \wedge \varrho_\alpha}}, \quad \forall p : \mathbf{E}_p[\rho] \leq \varrho.$$

Thus, (93) and (94) follow from [26, Lemma 25], because $\Lambda(\zeta_{\alpha,\varrho \wedge \varrho_\alpha}) = q_{\alpha,\Lambda^{\varrho \wedge \varrho_\alpha}}$ by (86).

Since the expression for $C_{\alpha,\Lambda,\varrho}$ given in (93) is differentiable in α, we can use the derivative test to determine optimal order for the SPE defined in (10). We obtain the following parametric form as a result:

$$R = \left(\frac{\varrho \wedge \varrho_\alpha - a}{b-a} b^\alpha \zeta_{\alpha,\varrho \wedge \varrho_\alpha}^{1-\alpha} \ln \frac{b^\alpha \zeta_{\alpha,\varrho \wedge \varrho_\alpha}^{1-\alpha}}{\zeta_{\alpha,\varrho \wedge \varrho_\alpha}} + \frac{b - \varrho \wedge \varrho_\alpha}{b-a} a^\alpha \zeta_{\alpha,\varrho \wedge \varrho_\alpha}^{1-\alpha} \ln \frac{a^\alpha \zeta_{\alpha,\varrho \wedge \varrho_\alpha}^{1-\alpha}}{\zeta_{\alpha,\varrho \wedge \varrho_\alpha}} \right) T, \tag{96}$$

$$E_{\mathrm{sp}}(R, \Lambda, \varrho) = \left(\varrho \wedge \varrho_\alpha - \zeta_{\alpha,\varrho \wedge \varrho_\alpha} + \frac{\varrho \wedge \varrho_\alpha - a}{b-a} b^\alpha \zeta_{\alpha,\varrho \wedge \varrho_\alpha}^{1-\alpha} \ln \frac{b^\alpha \zeta_{\alpha,\varrho \wedge \varrho_\alpha}^{1-\alpha}}{b} \right.$$
$$\left. + \frac{b - \varrho \wedge \varrho_\alpha}{b-a} a^\alpha \zeta_{\alpha,\varrho \wedge \varrho_\alpha}^{1-\alpha} \ln \frac{a^\alpha \zeta_{\alpha,\varrho \wedge \varrho_\alpha}^{1-\alpha}}{a} \right) T. \tag{97}$$

Using [27, equation (68)], (86), (90), (94), (96), and (97), we get the following parametric characterization:

$$R = D_1\big(\Lambda_\alpha^{q_{\alpha,\Lambda,\varrho}}(f_\alpha) \,\|\, q_{\alpha,\Lambda,\varrho}\big), \tag{98}$$

$$E_{\mathrm{sp}}(R, \Lambda, \varrho) = D_1\big(\Lambda_\alpha^{q_{\alpha,\Lambda,\varrho}}(f_\alpha) \,\|\, \Lambda(f_\alpha)\big), \tag{99}$$

---

[8] Note that $C_{\alpha,\Lambda,\varrho} = C_{\alpha,\Lambda^{\leq \varrho}}$ and $q_{\alpha,\Lambda,\varrho} = q_{\alpha,\Lambda^{\leq \varrho}}$ for the Poisson channel $\Lambda^{\leq \varrho} : \{f \in \mathcal{F} : \rho(f) \leq \varrho\} \to \mathcal{P}(\mathcal{Y})$ considered in [27, Example 10].

where f_α is any $\{a, b\}$ valued function in $\mathcal{F}^{\varrho \wedge \varrho_\alpha}$, i.e., any function $f_\alpha \colon (0, T] \to \{a, b\}$ satisfying $\rho(f_\alpha) = \varrho \wedge \varrho_\alpha$.

5. DISCUSSION

We have applied Augustin's method to derive SPBs for two families of memoryless channels. For the stationary memoryless channels with convex composition constraints, the novel observation behind Augustin's method is:

$$\limsup_{\substack{\phi \to \alpha \\ p \in \mathcal{A}}} D_\alpha(W \,\|\, q_{\phi, W, \mathcal{A}} \,|\, p) = C_{\alpha, W, \mathcal{A}}, \quad \forall \alpha \in (0, 1). \tag{100}$$

Note that the results established for the convex composition constrained stationary memoryless channels also hold for the cost constrained stationary memoryless channels, because any cost constraint on a stationary memoryless channel can be expressed as a convex composition constraint as well. For the nonstationary cost constrained memoryless channels we have employed (100) together with the convex conjugation techniques. Theorem 2 improves a similar result by Augustin, i.e., [24, Theorem 36.6], in terms of the approximation error terms. The prefactor of Theorem 2 is of the form $e^{-O(\ln n)}$, rather than $e^{-O(\sqrt{n})}$ similar to [24, Theorem 36.6]. Also, unlike [24, Theorem 36.6], Theorem 2 does not assume the cost functions to be bounded and thus holds for the Gaussian models considered in [17–20] as well.

For classical-quantum channels, the SPB was established in [40]. Following this breakthrough, there has been a reviewed interest in the SPB for classical-quantum channels [41–44]. Augustin's method, however, has not been applied to any quantum information theoretic model. Successful applications of Augustin's method will allow us to get rid of the stationarity, and finite input set hypotheses, at the very least.

The Augustin's variant of Gallager's bound discussed in Section 2.4 is not widely known. Our main aim in Section 2.4 was to present this approach in its simplest form. Thus, while bounding $\mathbf{E}[P_e]$, we were content with passing from (29) to (30). Using more careful analysis and bounding the deviation of the order α A-L information random variable, i.e., $\ln \dfrac{f_x}{h_x}$ for $q = q_{\alpha, p}$, one can obtain sharper bounds similar to the ones in [45–47]. The random coding bound for the classical-quantum channels has been established in [48]. It is suggested in [42, p. 5606] that the reliance of [48] on the codes generated with i.i.d. symbols might make it hard to modify the proof to the constant composition, i.e., composition constrained, case. We think that the Augustin's variant of Gallager's bound might be helpful in overcoming this issue.

As a side note, let us point out that Sections 2.4 and 3 imply that SPE is the reliability function for certain fading channels, i.e., for certain channels with state, provided that the list decoding is allowed, even in the nonstationary case. In particular, both the fast fading channels with no state information (i.e., with statistical state information) and the fast fading channels with state information only at the receiver are cost constrained memoryless channels. Thus, for the channels considered in [20] and the ones considered in [49, Section 4] the reliability function under list decoding is equal to the SPE. This is the case for the models with per antenna power constraints considered in [50–52] as well, because the channels considered in [50–52] are cost constrained memoryless channels, albeit with multiple constraints. The determination of the SPE for these channels, however, is a separate issue, as we have noted in Section 4.

The optimal prefactor of the SPB was known for specific channels since the early days of information theory; see, for example, [3, 4, 17]. In recent years, there has been a reviewed interest in establishing such sharp SPBs under various symmetry hypotheses [5, 6, 13, 21, 22]. The parametric characterization of the SPE given in (13) and (14) is used to establish such bounds for constant

composition codes in [14]. A refined SPB can be established in every single one of the cases considered in [3–6, 13, 17, 21, 22] using analogous parametric characterizations, as demonstrated by [28]. This is one of the reasons for us to present the parametric characterizations given in (70), (71), (91), (92), (98), and (99). It turns out that one can strengthen the strong converses in terms of their prefactors using analogous parametric characterizations under appropriate symmetry hypothesis as well; see [53].

APPENDIX A: BLAHUT'S APPROACH

In [31], Blahut derives a lower bound to the error probability of the channel codes without using constant composition arguments. Blahut claims that the exponential decay rate of his bound is equal to the SPE, see [31, Theorem 19]. Blahut claims the equality of the aforementioned exponent and the SPE in other publications too, see [32, Lemma 1] and [33, Theorem 10.1.4]. We show in the following that for the Z-channel the exponent of the Blahut's bound is infinite for any rate less than the channel capacity. Hence, [31, Theorem 19], [32, Lemma 1], and [33, Theorem 10.1.4] are all incorrect. More importantly, we show that even the best bound that can be obtained using Blahut's method is strictly inferior to the SPB in terms of its exponential decay rate with the block length.

For any $W\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ and $R \in [0, C_{1,W}]$, let $G(R, W, p, q)$ be

$$G(R, W, p, q) \triangleq \inf_{V\colon D_1(V\|q|p)\leq R} D_1(V\|W\,|\,p), \quad \forall p \in \mathcal{P}(\mathcal{X}),\ q \in \mathcal{P}(\mathcal{Y}).$$

Since $D_1(V\|(pV)\,|\,p) = I_1(p; V)$, where $(pV) = \sum_x p(x)V(x)$, from the alternative expression for $E_{\mathrm{sp}}(R, W)$ given in (20) it follows that

$$E_{\mathrm{sp}}(R, W) = \sup_{p\in\mathcal{P}(\mathcal{X})}\ \inf_{q\in\mathcal{P}(\mathcal{Y})} G(R, W, p, q), \quad \forall R \in [0, C_{1,W}]. \tag{A.1}$$

Thus, the max-min inequality implies

$$E_{\mathrm{sp}}(R, W) \leq \inf_{q\in\mathcal{P}(\mathcal{Y})}\ \sup_{p\in\mathcal{P}(\mathcal{X})} G(R, W, p, q), \quad \forall R \in [0, C_{1,W}].$$

The initial part of the proof of [31, Theorem 19] establishes the following bound on the error probability of the codes on a stationary product channel with the component channel W whose input set $\mathcal{X}$ and output set $\mathcal{Y}$ are finite:

$$P_{\mathrm{e}}^{\max,n} \geq O(1)e^{-o(n) - n\sup_{p\in\mathcal{P}(\mathcal{X})} G(R, W, p, q)}, \quad \forall q \in \mathcal{P}(\mathcal{Y}). \tag{A.2}$$

The second half of the proof of [31, Theorem 19] claims that $\sup_{p\in\mathcal{P}(\mathcal{X})} G(R, W, p, q_{\alpha_R, p_R})$ is equal to $E_{\mathrm{sp}}(R, W)$ for some (α_R, p_R) pair satisfying the following equalities[9]:

$$E_{\mathrm{sp}}(R, W) = E_{\mathrm{sp}}(R, W, p_R)$$
$$= \frac{1 - \alpha_R}{\alpha_R}(I_{\alpha_R}(p_R; W) - R).$$

When considered together with (A.1), the second half of the proof of [31, Theorem 19] asserts that

$$\sup_{p\in\mathcal{P}(\mathcal{X})} G(R, W, p, q_{\alpha_R, p_R}) \overset{?}{=} \sup_{p\in\mathcal{P}(\mathcal{X})}\ \inf_{q\in\mathcal{P}(\mathcal{Y})} G(R, W, p, q). \tag{A.3}$$

---

[9] Blahut mentions only the first equality explicitly.

Note that [32, Lemma 1] and [33, Theorem 10.1.4] imply the same equality when considered together with (A.1) as well. In order to disprove (A.3), we consider the Z-channel, which is a discrete channel with the input set $\mathcal{X} = \{1, 2\}$ and the output set $\mathcal{Y} = \{a, b\}$ such that

$$W = \begin{bmatrix} 1 & 0 \\ \varepsilon & 1 - \varepsilon \end{bmatrix}.$$

We determine the order α Augustin capacity and the order α Augustin center using the identities $C_{\alpha,W} = D_\alpha(W(1) \,\|\, q_{\alpha,W})$ and $C_{\alpha,W} = D_\alpha(W(2) \,\|\, q_{\alpha,W})$:

$$C_{\alpha,W} = \ln\left(1 + \left(\frac{1 - \varepsilon^\alpha}{(1 - \varepsilon)^\alpha}\right)^{\frac{1}{1-\alpha}}\right), \tag{A.4}$$

$$q_{\alpha,W}(a) = \frac{(1 - \varepsilon)^{\frac{\alpha}{1-\alpha}}}{(1 - \varepsilon)^{\frac{\alpha}{1-\alpha}} + (1 - \varepsilon^\alpha)^{\frac{1}{1-\alpha}}}. \tag{A.5}$$

Then for $q = q_{\alpha,W}$ the tilted channel W_α^q defined in (3) is

$$W_\alpha^{q_{\alpha,W}} = \begin{bmatrix} 1 & 0 \\ \varepsilon^\alpha & 1 - \varepsilon^\alpha \end{bmatrix}.$$

Furthermore, one can confirm by substitution that the order α Augustin center $q_{\alpha,W}$ is the fixed point of the order α Augustin operator defined in (4) for the prior $p_{\alpha,W}$ satisfying

$$p_{\alpha,W}(1) = \frac{(1 - \varepsilon)^{\frac{\alpha}{1-\alpha}} - \varepsilon^\alpha(1 - \varepsilon^\alpha)^{\frac{\alpha}{1-\alpha}}}{(1 - \varepsilon)^{\frac{\alpha}{1-\alpha}} + (1 - \varepsilon^\alpha)^{\frac{1}{1-\alpha}}}.$$

Thus, the order α Augustin center $q_{\alpha,W}$ is equal to the order α Augustin mean for the prior $p_{\alpha,W}$, i.e., $q_{\alpha,W} = q_{\alpha,p_{\alpha,W}}$, by [26, Lemma 13(c),(d)] and $I_\alpha(p_{\alpha,W}; W) = D_\alpha(W \,\|\, q_{\alpha,W} \,|\, p_{\alpha,W})$. Consequently, $I_\alpha(p_{\alpha,W}; W) = C_{\alpha,W}$ as well.

Note that $C_{\alpha,W}$ given in (A.4) is a differentiable function of α such that

$$\frac{\partial}{\partial \alpha} C_{\alpha,W} = \frac{q_\alpha(b)}{(1 - \alpha)^2}\left(\ln\frac{1 - \varepsilon^\alpha}{1 - \varepsilon} + \frac{\varepsilon^\alpha}{1 - \varepsilon^\alpha}\ln\frac{\varepsilon^\alpha}{\varepsilon}\right)$$

$$= \frac{1}{(1 - \alpha)^2} D_1(W_\alpha^{q_{\alpha,W}} \,\|\, W \,|\, p_{\alpha,W}).$$

Then, using first the identity $I_\alpha(p_{\alpha,W}; W) = C_{\alpha,W}$ and then [26, equation (35)], we get

$$\frac{\partial}{\partial \alpha}\frac{1 - \alpha}{\alpha}(C_{\alpha,W} - R) = \frac{1}{\alpha^2}\left(R - C_{\alpha,W} + (1 - \alpha)\alpha\frac{\partial}{\partial \alpha}C_{\alpha,W}\right)$$

$$= \frac{1}{\alpha^2}\left(R - I_\alpha(p_{\alpha,W}; W) + \frac{\alpha}{1 - \alpha}D_1(W_\alpha^{q_{\alpha,W}} \,\|\, W \,|\, p_{\alpha,W})\right)$$

$$= \frac{1}{\alpha^2}\left(R - I_1(p_{\alpha,W}; W_\alpha^{q_{\alpha,W}})\right).$$

One can confirm numerically that $I_1(p_{\alpha,W}; W_\alpha^{q_{\alpha,W}})$ is an increasing function of α for any $\varepsilon \in (0, 1)$. Thus, we can express the rate and the corresponding SPE in the following parametric form:

$$R(\alpha) = I_1(p_{\alpha,W}; W_\alpha^{q_{\alpha,W}}),$$

$$E_{\mathrm{sp}}(R(\alpha), W) = D_1(W_\alpha^{q_{\alpha,W}} \,\|\, W \,|\, p_{\alpha,W}).$$

Thus, for $R = R(\phi)$ we have $(\alpha_R, p_R) = (\phi, p_{\phi,W})$ and $q_{\alpha_R, p_R} = q_{\phi,W}$. Then

$$\sup_{p \in \mathcal{P}(\mathcal{X})} G(R, W, p, q_{\alpha_R, p_R}) \geq G(R, W, p, q_{\alpha_R, W})\big|_{p: p(1)=1}$$
$$= \inf_{v: D_1(v \| q_{\alpha_R, W}) \leq R} D_1(v \| W(1)).$$

On the other hand, $v(b) > 0$ for all v satisfying $D_1(v \| q_{\alpha_R, W}) \leq R$, because $D_1(v \| q_{\alpha_R, W}) = \ln \dfrac{1}{q_{\alpha_R, W}(a)} = C_{\alpha_R, W}$ whenever $v(b) = 0$ and $C_{\alpha_R, W} > R$. Furthermore, $D_1(v \| W(1)) = \infty$ whenever $v(b) > 0$. Thus,

$$\inf_{v: D_1(v \| q_{\alpha_R, W}) \leq R} D_1(v \| W(1)) = \infty.$$

Thus, the bound established in the initial part of the proof of [31, Theorem 19], i.e., (A.2), is trivial for $q = q_{\alpha_R, p_R}$. Furthermore, [31, Theorem 19], [32, Lemma 1], and [33, Theorem 10.1.4] are all incorrect, because $E_{\mathrm{sp}}(R, W) < \infty$ for all $R \in [0, C_{1,W}]$ for the Z-channels.[10]

Using other values of q, one can obtain nontrivial bounds from (A.2); those bounds, however, do not imply the SPB either. If $q(a) < e^{-R}$, then $\sup\limits_{p \in \mathcal{P}(\mathcal{X})} G(R, W, p, q_{\alpha_R, p_R}) = \infty$ as a result of the analysis presented above. We analyze the case $q(a) \geq e^{-R}$ in the following:

$$\sup_{p \in \mathcal{P}(\mathcal{X})} G(R, W, p, q) \geq \inf_{V: D_1(V \| q | p_R) \leq R} D_1(V \| W | p_R)$$
$$\geq \inf_V D_1(V \| W | p_R) + \frac{1 - \alpha_R}{\alpha_R}(D_1(V \| q | p_R) - R)$$
$$\stackrel{(a)}{=} \frac{1 - \alpha_R}{\alpha_R}(D_{\alpha_R}(W \| q | p_R) - R)$$
$$\stackrel{(b)}{\geq} \frac{1 - \alpha_R}{\alpha_R}(I_{\alpha_R}(p_R; W) + D_{\alpha_R}(q_{\alpha_R, p_R} \| q) - R)$$
$$= E_{\mathrm{sp}}(R, W) + \frac{1 - \alpha_R}{\alpha_R} D_{\alpha_R}(q_{\alpha_R, p_R} \| q)$$
$$\stackrel{(c)}{\geq} E_{\mathrm{sp}}(R, W) + \frac{1 - \alpha_R}{2} \| q - q_{\alpha_R, p_R} \|^2,$$

where (a) is by [36, Theorem 30], (b) by [26, Lemma 13(c)], and (c) by [36, Theorem 31]. On the other hand, $q_{\alpha_R, p_R}(a) < e^{-R}$, because $q_{\alpha_R, p_R} = q_{\alpha_R, W}$ and $C_{\alpha_R, W} > R$. Thus,

$$\inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{p \in \mathcal{P}(\mathcal{X})} G(R, W, p, q) \geq E_{\mathrm{sp}}(R, W) + 2(1 - \alpha_R)(e^{-R} - q_{\alpha_R, W}(a))^2$$
$$= \sup_{p \in \mathcal{P}(\mathcal{X})} \inf_{q \in \mathcal{P}(\mathcal{Y})} G(R, W, p, q) + 2(1 - \alpha_R)(e^{-R} - q_{\alpha_R, W}(a))^2.$$

Hence, it is not possible to derive the SPB using Blahut's method, as it is presented in [31]. When the input set is finite, one can overcome this problem by employing composition based expurgations. But that approach had been presented by Haroutunian in [10], before [31].

---

[10] Recently, Yang argued that Blahut's method can be used to derive the SPB if the minimax equality given in [54, equation (3.63)] holds. Thus, as a result of our analysis we can conclude that [54, equation (3.63)] does not holds in general. This fact can be derived using the absence of the minimax equality for $G(R, W, p, q)$ without relying on the reasoning in [54] as well.

APPENDIX B: *R-G INFORMATION MEASURES*

The order one R-G information measures are equal to the corresponding order one A-L information measures by definition. Thus, our discussion will be confined to orders other than one.

Definition 12. For any $\alpha \in \mathbb{R}_+ \setminus \{1\}$, channel $W\colon \mathfrak{X} \to \mathcal{P}(\mathcal{Y})$ with a cost function $\rho\colon \mathfrak{X} \to \mathbb{R}_{\geq 0}^\ell$, $p \in \mathcal{P}(\mathfrak{X})$, and $\lambda \in \mathbb{R}_{\geq 0}^\ell$, the *order α Rényi–Gallager (R-G) information for the input distribution p and the Lagrange multiplier λ* is

$$I_\alpha^{g\lambda}(p; W) \triangleq \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha\big(p \circledast W e^{\frac{1-\alpha}{\alpha}\lambda \cdot \rho} \,\|\, p \otimes q\big).$$

If λ is a vector of zeros, then the R-G information is the Rényi information. Similarly to the Rényi information, the R-G information has a closed form expression, described in terms of a mean achieving the infimum in its definition.

Definition 13. For any $\alpha \in \mathbb{R}_+$, channel $W\colon \mathfrak{X} \to \mathcal{P}(\mathcal{Y})$ with a cost function $\rho\colon \mathfrak{X} \to \mathbb{R}_{\geq 0}^\ell$, $p \in \mathcal{P}(\mathfrak{X})$, and $\lambda \in \mathbb{R}_{\geq 0}^\ell$, the *order α mean measure for the input distribution p and the Lagrange multiplier λ* is

$$\frac{d\mu_{\alpha,p}^\lambda}{d\nu} \triangleq \left[\sum_x p(x) e^{(1-\alpha)\lambda \cdot \rho(x)} \left(\frac{dW(x)}{d\nu}\right)^\alpha\right]^{\frac{1}{\alpha}}. \tag{B.1}$$

The *order α Rényi–Gallager (R-G) mean for the input distribution p and the Lagrange multiplier λ* is

$$q_{\alpha,p}^{g\lambda} \triangleq \frac{\mu_{\alpha,p}^\lambda}{\|\mu_{\alpha,p}^\lambda\|}.$$

Both $\mu_{\alpha,p}^\lambda$ and $q_{\alpha,p}^{g\lambda}$ depend on the Lagrange multiplier λ for $\alpha \in \mathbb{R}_+ \setminus \{1\}$. Furthermore, one can confirm by substitution that

$$D_\alpha\big(p \circledast W e^{\frac{1-\alpha}{\alpha}\lambda \cdot \rho} \,\|\, p \otimes q\big) = I_\alpha^{g\lambda}(p; W) + D_\alpha\big(q_{\alpha,p}^{g\lambda} \,\|\, q\big), \quad \alpha \in \mathbb{R}_+ \setminus \{1\}.$$

Then as a result of [26, Lemma 2] we have

$$\begin{aligned}
I_\alpha^{g\lambda}(p; W) &= D_\alpha\big(p \circledast W e^{\frac{1-\alpha}{\alpha}\lambda \cdot \rho} \,\|\, p \otimes q_{\alpha,W}^{g\lambda}\big) \\
&= \frac{\alpha}{\alpha-1} \ln \|\mu_{\alpha,p}^\lambda\|, \quad \alpha \in \mathbb{R}_+ \setminus \{1\}.
\end{aligned} \tag{B.2}$$

Using the definitions of the A-L information and the R-G information together with the Jensen's inequality and the concavity of the natural logarithm function, we get

$$I_\alpha^\lambda(p; W) \geq I_\alpha^{g\lambda}(p; W), \quad \alpha \in (0, 1], \tag{B.3}$$

$$I_\alpha^\lambda(p; W) \leq I_\alpha^{g\lambda}(p; W), \quad \alpha \in [1, \infty). \tag{B.4}$$

Definition 14. For any $\alpha \in \mathbb{R}_+$, channel $W\colon \mathfrak{X} \to \mathcal{P}(\mathcal{Y})$ with a cost function $\rho\colon \mathfrak{X} \to \mathbb{R}_{\geq 0}^\ell$, and $\lambda \in \mathbb{R}_{\geq 0}^\ell$, the *order α Rényi–Gallager (R-G) capacity for the Lagrange multiplier λ* is

$$C_{\alpha,W}^{g\lambda} \triangleq \sup_{p \in \mathcal{P}(\mathfrak{X})} I_\alpha^{g\lambda}(p; W). \tag{B.5}$$

Although inequalities in (B.3) and (B.4) are strict for most input distributions, as a result of [26, Theorem 3], we have

$$C_{\alpha,W}^{g\lambda} = S_{\alpha,W}^\lambda. \tag{B.6}$$

Thus, $C_{\alpha,W}^{g\lambda} = C_{\alpha,W}^{\lambda}$ by (9). This is the reason why in terms of determining the optimal performance either family can be used. The following lemma is, in essence, a restatement of [29, Theorem 8], which is the result that popularized the use of R-G information measures in cost constrained problems; see [18–20].

Lemma 12. *For any* $\ell, M, L \in \mathbb{Z}_+$ *such that* $L < M$, $W: \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $\rho: \mathcal{X} \to \mathbb{R}_{\geq 0}^{\ell}$, $p \in \mathcal{P}(\mathcal{X})$, $\mathcal{B} \subset \mathcal{X}$, *and* $\alpha \in \left[\frac{1}{1+L}, 1\right)$ *there exists an* (M, L) *channel code with an encoding function of the form* $\Psi: \mathcal{M} \to \mathcal{B}$ *such that*

$$\ln P_e \leq \frac{\alpha - 1}{\alpha} \left[I_\alpha^{g\lambda}(p; W) + \left(\inf_{x \in \mathcal{B}} \lambda \cdot \rho(x) \right) - \ln \frac{(M-1)e}{L} \right] - \frac{\ln p(\mathcal{B})}{\alpha}. \tag{B.7}$$

Proof of Lemma 12. We follow the proof of Lemma 6 up to (30). As a result of (26) and (30), we have

$$\ln \mathbf{E}[P_e] \leq \ln \mathbf{E}_q \left[\left(\sum_{x \in \mathcal{B}} \frac{p(x)}{p(\mathcal{B})} e^{(1-\alpha)\lambda \cdot \rho(x)} \left(\frac{dW(x)}{dq} \right)^\alpha \right)^{\frac{1}{\alpha}} \right]$$

$$+ \frac{\alpha - 1}{\alpha} \left[\left(\inf_{x \in \mathcal{B}} \lambda \cdot \rho(x) \right) - \ln \frac{(M-1)e}{L} \right]$$

$$\leq \ln \mathbf{E}_q \left[\left(\sum_{x \in \mathcal{X}} p(x) e^{(1-\alpha)\lambda \cdot \rho(x)} \left(\frac{dW(x)}{dq} \right)^\alpha \right)^{\frac{1}{\alpha}} \right]$$

$$+ \frac{\alpha - 1}{\alpha} \left[\left(\inf_{x \in \mathcal{B}} \lambda \cdot \rho(x) \right) - \ln \frac{(M-1)e}{L} \right] - \frac{\ln p(\mathcal{B})}{\alpha}.$$

Since there exists a code with P_e less than or equal to $\mathbf{E}[P_e]$, the existence of a code satisfying (B.7) with an encoding function of the form $\Psi: \mathcal{M} \to \mathcal{B}$ follows from (B.1) and (B.2). △

APPENDIX C: OMITTED PROOFS

Proof of Lemma 1. $E_{sp}(R, W, \mathcal{A})$ is convex in R, because the pointwise supremum of a family of convex functions is convex and $\frac{1-\alpha}{\alpha}(C_{\alpha,W,\mathcal{A}} - R)$ is convex in R for any $\alpha \in (0, 1)$. $E_{sp}(R, W, \mathcal{A})$ is nonincreasing in R as a result of an analogous argument. The continuity and finiteness claims are proved while establishing (11).

Recall that $C_{\alpha,W,\mathcal{A}}$ is a nondecreasing function of the order α by [26, Lemma 23(a)].

- If $C_{0^+,W,\mathcal{A}} = \infty$, then $C_{1/2,W} = \infty$ and $E_{sp}(R, W, \mathcal{A}) = \infty$ for all $R \in \mathbb{R}_{\geq 0}$. On the other hand, $R < C_{0^+,W,\mathcal{A}}$ for all $R \in \mathbb{R}_{\geq 0}$. Hence, (11) holds, and claims about the continuity and the finiteness of $E_{sp}(R, W, \mathcal{A})$ are void.

- If $C_{0^+,W,\mathcal{A}} < \infty$ and $C_{0^+,W,\mathcal{A}} = C_{1,W,\mathcal{A}}$, then $E_{sp}(R, W, \mathcal{A}) = \infty$ for all $R \in [0, C_{1,W,\mathcal{A}})$ and $E_{sp}(R, W, \mathcal{A}) = 0$ for all $R \geq C_{1,W,\mathcal{A}}$. Hence, (11) and claims about the continuity and the finiteness of $E_{sp}(R, W, \mathcal{A})$ hold.

- If $C_{0^+,W,\mathcal{A}} < \infty$ and $C_{0^+,W,\mathcal{A}} \neq C_{1,W,\mathcal{A}}$, then $E_{sp}(R, W, \mathcal{A}) = \infty$ for all $R \in [0, C_{0^+,W,\mathcal{A}})$. For $R \geq C_{0^+,W,\mathcal{A}}$, the nonnegativity of $\frac{1-\alpha}{\alpha}(C_{\alpha,W,\mathcal{A}} - R)$ imply the restrictions given in (11).

  As a result of (11), $E_{sp}(R, W, \mathcal{A})$ is finite for all $R > \lim_{\phi \downarrow 0} C_{\phi,W,\mathcal{A}}$. Thus, $E_{sp}(R, W, \mathcal{A})$ is continuous on $(C_{0^+,W,\mathcal{A}}, \infty)$ by [35, Theorem 6.3.3]. In order to extent the continuity to $[C_{0^+,W,\mathcal{A}}, \infty)$, note that the function $\frac{1-\alpha}{\alpha}(C_{\alpha,W,\mathcal{A}} - R)$ is decreasing and continuous in R for each α in $(0, 1)$. Thus, $E_{sp}(R, W, \mathcal{A})$ is a nonincreasing and lower semicontinuous function of R. Hence, $E_{sp}(R, W, \mathcal{A})$ is continuous from the right, and hence at $R = C_{0^+,W,\mathcal{A}}$. △

Proof of Corollary 2. $\frac{1-\alpha}{\alpha}\left(C_{\alpha,W_{[1,n]},n\varrho} - \ln\frac{M}{L}\right) \geq E_{\mathrm{sp}}\left(\ln\frac{M}{L}, W_{[1,n]}, n\varrho\right) - \frac{1}{n}$ for an $\alpha \in$
$(\eta, 1)$ by Lemma 1. There exists a $p \in \mathcal{P}(\widetilde{\mathcal{X}}_1^n)$ of the form $p = \bigotimes\limits_{t=1}^{n} p_t$ satisfying $\mathbf{E}_p[\rho_{[1,n]}] \leq n(\widetilde{\varrho} - \delta)$.
Let $\widehat{\varrho} \triangleq \varrho - \frac{3e\varsigma}{n}$. There exists a $\widetilde{p} \in \mathcal{P}(\widetilde{\mathcal{X}}_1^n)$ of the form $\widetilde{p} = \bigotimes\limits_{t=1}^{n} \widetilde{p}_t$ satisfying both $I_\alpha(\widetilde{p}; \widetilde{W}_{[1,n]}) \geq$
$C_{\alpha,\widetilde{W}_{[1,n]},n\widehat{\varrho}} - \epsilon$ and $\mathbf{E}_{\widetilde{p}}[\rho_{[1,n]}] \leq n\widehat{\varrho}$ by [26, Lemmas 14 and 28]. Let $\widehat{W}_{[1,n]}\colon \widehat{\mathcal{X}}_1^n \to \mathcal{P}(\mathcal{Y}_1^n)$ be a
product channel satisfying $\widehat{W}_{[1,n]}(x_1^n) = W_{[1,n]}(x_1^n)$ for all $x_1^n \in \widehat{\mathcal{X}}_1^n$ and $\widehat{\mathcal{X}}_t = \operatorname{supp}\widetilde{p}_t \cup \operatorname{supp}p_t$ for
all $t \in \{1,\ldots,n\}$. Then $C_{\alpha,\widehat{W}_{[1,n]},n\widehat{\varrho}} \geq C_{\alpha,\widetilde{W}_{[1,n]},n\widehat{\varrho}} - \epsilon$ and $\widetilde{\varrho} \leq \widehat{\varrho}$ by the construction. There exists
a $\widehat{p} \in \mathcal{P}(\widehat{\mathcal{X}}_1^n)$ satisfying both $I_\alpha(\widehat{p}; \widehat{W}_{[1,n]}) = C_{\alpha,\widehat{W}_{[1,n]},n\widehat{\varrho}}$ and $\mathbf{E}_{\widehat{p}}[\rho_{[1,n]}] \leq n\widehat{\varrho}$ by [26, Lemma 19],
because $\widehat{\mathcal{X}}_1^n$ is finite. Furthermore, we can assume that $\widehat{p}$ is of the form $\widehat{p} = \bigotimes\limits_{t=1}^{n} \widehat{p}_t$ without loss of
generality by [26, Lemma 14], because the cost function $\rho_{[1,n]}$ is additive.

Note that $\mathbf{E}_{\widehat{p}_t}\left[\left|\rho_t - \mathbf{E}_{\widehat{p}_t}[\rho_t]\right|^\kappa\right]^{1/\kappa} \leq \varsigma$ for all $\kappa \in \mathbb{R}_+$ and $t \in \{1,\ldots,n\}$, because ρ_t is positive
and less than ς with probability one under $\widehat{p}_t$. Then, using [25, Lemma 19] for $\kappa = \ln n$, we get

$$\widehat{p}\left(\left|\rho_{[1,n]}(x) - \mathbf{E}_{\widehat{p}}[\rho_{[1,n]}]\right| < 3\varsigma e\right) \geq \frac{1}{2\sqrt{n}}.$$

On the other hand, there exists a $\widehat{\lambda} = \lambda_{\alpha,\widehat{W}_{[1,n]},n\widehat{\varrho}}$ satisfying $C_{\alpha,\widehat{W}_{[1,n]},n\widehat{\varrho}} = C_{\alpha,\widehat{W}_{[1,n]}}^{\widehat{\lambda}} + \widehat{\lambda}n\widehat{\varrho}$ by [26,
Lemma 29(c)], because $n\widehat{\varrho}$ is in the interior of the feasible cost constraints $\widehat{W}_{[1,n]}$ by the construction.
Furthermore, $I_\alpha^{\widehat{\lambda}}(\widehat{p}; \widehat{W}_{[1,n]}) = C_{\alpha,\widehat{W}_{[1,n]}}^{\widehat{\lambda}}$ by [26, Lemma 29(d)], because $I_\alpha(\widehat{p}; \widehat{W}_{[1,n]}) = C_{\alpha,\widehat{W}_{[1,n]},n\widehat{\varrho}}$
and $\mathbf{E}_{\widehat{p}}[\rho_{[1,n]}] \leq n\widehat{\varrho}$. As a result, $D_\alpha(W_{[1,n]}(x_1^n) \| q_{\alpha,\widehat{p}}) - \widehat{\lambda}\rho_{[1,n]}(x_1^n) = C_{\alpha,\widehat{W}_{[1,n]}}^{\widehat{\lambda}}$ for all $x_1^n \in \operatorname{supp}\widehat{p}$
and $\widehat{\lambda}n\widehat{\varrho} = \widehat{\lambda}\,\mathbf{E}_{\widehat{p}}[\rho_{[1,n]}]$. Applying Lemma 6 for $\mathcal{B} = \left\{x_1^n \in \operatorname{supp}\widehat{p}: \left|\rho_{[1,n]}(x_1^n) - \mathbf{E}_{\widehat{p}}[\rho_{[1,n]}]\right| < 3\varsigma e\right\}$
we can conclude that there exists an (M, L) channel code satisfying

$$\ln P_{\mathrm{e}} \leq \frac{\alpha-1}{\alpha}\left[C_{\alpha,\widehat{W}_{[1,n]}}^{\widehat{\lambda}} + \inf_{x_1^n \in \mathcal{B}} \widehat{\lambda}\rho_{[1,n]}(x_1^n) - \ln\frac{(M-1)e}{L}\right] + \frac{\ln 4n}{2\alpha}$$

$$\leq \frac{\alpha-1}{\alpha}\left[C_{\alpha,\widehat{W}_{[1,n]},n\widehat{\varrho}} - 3\varsigma e\widehat{\lambda} - \ln\frac{(M-1)e}{L}\right] + \frac{\ln 4n}{2\alpha}.$$

Since $C_{\alpha,\widehat{W}_{[1,n]},n\widehat{\varrho}} + \widehat{\lambda}n(\varrho - \widehat{\varrho}) \geq C_{\alpha,\widehat{W}_{[1,n]},n\varrho}$ by [26, Lemma 27(b)], we get

$$\ln P_{\mathrm{e}} \leq -E_{\mathrm{sp}}\left(\ln\frac{M}{L}, W_{[1,n]}, \varrho\right) + \frac{1}{n} + \frac{1-\alpha}{\alpha}(6\varsigma e\widehat{\lambda} + 2\epsilon + 1) + \frac{\ln 4n}{2\alpha}.$$

Then (25) holds, because $\frac{1-\alpha}{\alpha}C_{\alpha,W_{[1,n]},n\varrho}$ is nonincreasing in α by [26, Lemma 23(b)], provided
that $\widehat{\lambda} \leq \frac{C_{\alpha,W_{[1,n]},n\varrho}}{n\delta}$. In order to see why such a bound holds first note that $\widehat{\varrho} \leq \varrho$ by definition
and thus $C_{\alpha,\widehat{W}_{[1,n]},n\widehat{\varrho}} \leq C_{\alpha,\widehat{W}_{[1,n]},n\varrho}$. Furthermore,

$$C_{\alpha,\widehat{W}_{[1,n]},n\widehat{\varrho}} = \inf_{\lambda \geq \widehat{\lambda}} C_{\alpha,\widehat{W}_{[1,n]}}^{\lambda} + \lambda n\widehat{\varrho}$$

$$> n\delta\widehat{\lambda} + \inf_{\lambda \geq \widehat{\lambda}} C_{\alpha,\widehat{W}_{[1,n]}}^{\lambda} + n\lambda(\widehat{\varrho} - \delta)$$

$$\geq n\delta\widehat{\lambda} + C_{\alpha,\widehat{W}_{[1,n]},n(\widehat{\varrho}-\delta)}$$

$$\geq n\delta\widehat{\lambda}. \quad \triangle$$

Proof of Lemma 8. $C_{\alpha,W,\mathcal{A}} \leq \widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}}$ by (39) and [26, Lemma 23(a),(b)]. Then, as a result of the expressions for $E_{\mathrm{sp}}(R,W,\mathcal{A})$ given in (11) and the definition of $\widetilde{E}^{\epsilon}_{\mathrm{sp}}(R,W,\mathcal{A})$ given in (40), we have

$$E_{\mathrm{sp}}(R,W,\mathcal{A}) \leq \widetilde{E}^{\epsilon}_{\mathrm{sp}}(R,W,\mathcal{A}), \quad \forall R \in \mathbb{R}_{\geq 0}. \tag{C.1}$$

Let us proceed with bounding $\widetilde{E}^{\epsilon}_{\mathrm{sp}}(R,W,\mathcal{A}) - E_{\mathrm{sp}}(R,W,\mathcal{A})$ from above for $R \in [C_{\phi,W,\mathcal{A}},\infty)$. First,

$$\frac{1-\alpha}{\alpha}(\widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}} - R)$$

$$= \frac{1}{\epsilon} \int\limits_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} \left(\frac{1-\alpha}{\alpha} \vee \frac{1-\eta}{\eta} \right) C_{\eta,W,\mathcal{A}}\, d\eta - \frac{1-\alpha}{\alpha}R$$

$$= \frac{1}{\epsilon}\frac{1-\alpha}{\alpha} \int\limits_{\alpha}^{\alpha+\epsilon(1-\alpha)} (C_{\eta,W,\mathcal{A}} - R)\, d\eta + \frac{1}{\epsilon} \int\limits_{\alpha-\epsilon\alpha}^{\alpha} \frac{1-\eta}{\eta}(C_{\eta,W,\mathcal{A}} - R)\, d\eta + \frac{R}{\epsilon} \int\limits_{\alpha-\epsilon\alpha}^{\alpha} \frac{\alpha-\eta}{\eta\alpha}\, d\eta$$

$$\leq \frac{1}{\epsilon}\frac{1-\alpha}{\alpha} \int\limits_{\alpha}^{\alpha+\epsilon(1-\alpha)} (C_{\eta,W,\mathcal{A}} - R)\, d\eta + \frac{1}{\epsilon} \int\limits_{\alpha-\epsilon\alpha}^{\alpha} \frac{1-\eta}{\eta}(C_{\eta,W,\mathcal{A}} - R)\, d\eta + \frac{\epsilon}{1-\epsilon}R. \tag{C.2}$$

We bound $\widetilde{E}^{\epsilon}_{\mathrm{sp}}(R,W,\mathcal{A})$ by bounding the expression in (C.2) separately on two intervals for α.

To bound the expression in (C.2) for $\alpha \in [\phi,1)$, we use the fact that $\sup\limits_{\eta\in(0,1)} \frac{1-\eta}{\eta}(C_{\eta,W,\mathcal{A}} - R) = E_{\mathrm{sp}}(R,W,\mathcal{A})$. We have

$$\frac{1-\alpha}{\alpha}(\widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}} - R)$$

$$\leq \frac{1}{\epsilon}\frac{1-\alpha}{\alpha} \int\limits_{\alpha}^{\alpha+\epsilon(1-\alpha)} (C_{\eta,W,\mathcal{A}} - R)d\eta + \frac{1}{\epsilon} \int\limits_{\alpha-\epsilon\alpha}^{\alpha} \frac{1-\eta}{\eta}(C_{\eta,W,\mathcal{A}} - R)d\eta + \frac{\epsilon}{1-\epsilon}R$$

$$\leq \frac{1}{\epsilon}\frac{1-\alpha}{\alpha} \int\limits_{\alpha}^{\alpha+\epsilon(1-\alpha)} \frac{\eta}{1-\eta}E_{\mathrm{sp}}(R,W,\mathcal{A})d\eta + \frac{1}{\epsilon} \int\limits_{\alpha-\epsilon\alpha}^{\alpha} E_{\mathrm{sp}}(R,W,\mathcal{A})d\eta + \frac{\epsilon}{1-\epsilon}R$$

$$\leq E_{\mathrm{sp}}(R,W,\mathcal{A}) + \frac{\epsilon}{1-\epsilon}\frac{1-\alpha}{\alpha}E_{\mathrm{sp}}(R,W,\mathcal{A}) + \frac{\epsilon}{1-\epsilon}R.$$

Since $(1-\alpha)E_{\mathrm{sp}}(R,W,\mathcal{A}) + \alpha R \leq (R \vee E_{\mathrm{sp}}(R,W,\mathcal{A}))$ for all $\alpha \in (0,1)$, we have

$$\frac{1-\alpha}{\alpha}(\widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}} - R) \leq E_{\mathrm{sp}}(R,W,\mathcal{A}) + \frac{\epsilon}{1-\epsilon}\frac{R \vee E_{\mathrm{sp}}(R,W,\mathcal{A})}{\phi}, \quad \alpha \in [\phi,1). \tag{C.3}$$

In order to bound the expression in (C.2) for $\alpha \in (0,\phi]$, recall that $C_{\alpha,W,\mathcal{A}}$ is nondecreasing in α by [26, Lemma 23(a)]. Thus, for any $R \geq C_{\phi,W,\mathcal{A}}$ we have

$$\frac{1-\alpha}{\alpha}(\widetilde{C}^{\epsilon}_{\alpha,W,\mathcal{A}} - R)$$

$$\leq \frac{1}{\epsilon}\frac{1-\alpha}{\alpha} \int\limits_{\phi}^{\alpha+\epsilon(1-\alpha)} \frac{\eta}{1-\eta}E_{\mathrm{sp}}(R,W,\mathcal{A})\, d\eta\, \mathbb{1}_{\left\{\alpha\in\left[\frac{\phi-\epsilon}{1-\epsilon},\phi\right]\right\}} + \frac{\epsilon}{1-\epsilon}R$$

$$\leq \frac{1}{\epsilon}\frac{\alpha+\epsilon(1-\alpha)}{\alpha(1-\epsilon)} \int\limits_{\phi}^{\alpha+\epsilon(1-\alpha)} E_{\mathrm{sp}}(R,W,\mathcal{A})\, d\eta\, \mathbb{1}_{\left\{\alpha\in\left[\frac{\phi-\epsilon}{1-\epsilon},\phi\right]\right\}} + \frac{\epsilon}{1-\epsilon}R$$

$$= \left[\frac{\alpha(1-\epsilon) + 2\epsilon - \phi}{\epsilon} - \frac{\phi - \epsilon}{\alpha(1-\epsilon)} \right] E_{\mathrm{sp}}(R, W, \mathcal{A}) \mathbb{1}_{\left\{ \alpha \in \left[\frac{\phi - \epsilon}{1 - \epsilon}, \phi \right] \right\}} + \frac{\epsilon}{1 - \epsilon} R$$

$$\leq (1 - \phi) E_{\mathrm{sp}}(R, W, \mathcal{A}) + \frac{\epsilon}{1 - \epsilon} \frac{(1 - \phi) E_{\mathrm{sp}}(R, W, \mathcal{A}) + \phi R}{\phi}, \quad \alpha \in (0, \phi]. \tag{C.4}$$

Now (41) follows from (C.1), (C.3), and (C.4).

On the other hand, $C_{\alpha, W, \mathcal{A}}$ is nondecreasing and $\frac{1 - \alpha}{\alpha} C_{\alpha, W, \mathcal{A}}$ is nonincreasing in α by [26, Lemma 23(a),(b)]. Then, as a result of the expression for $E_{\mathrm{sp}}(R, W, \mathcal{A})$ given in (11), we have $E_{\mathrm{sp}}(R, W, \mathcal{A}) \leq \frac{1 - \phi}{\phi} R$ for all $R \in [C_{\phi, W, \mathcal{A}}, \infty)$. Hence,

$$R \vee E_{\mathrm{sp}}(R, W, \mathcal{A}) \leq R/\phi, \quad \forall R \in [C_{\phi, W, \mathcal{A}}, \infty). \tag{C.5}$$

Now (42) follows from (41) and (C.5). $\triangle$

Proof of Theorem 1. We prove Theorem 1 using Lemmas 8 and 9. We are free to choose different values for κ and ϵ for different values of n provided that the hypotheses of Lemmas 8 and 9 are satisfied.

As a result of Assumption 1, there exists a $K \in [1, \infty)$ and an $n_0 \in \mathbb{Z}_+$ such that

$$\max_{t : t \leq n} C_{1/2, U_{\mathcal{A}}^{(t)}} \leq K \ln(n), \quad \forall n \geq n_0.$$

Let κ_n be $\kappa_n = K \ln(1 + n)$. Then

$$\gamma_n \leq 40(K + 1) \ln(1 + n), \quad \forall n \geq n_0. \tag{C.6}$$

$C_{\alpha, W, \mathcal{A}}$ is nondecreasing in α by [26, Lemma 23(a)] and $\frac{1 - \alpha}{\alpha} C_{\alpha, W, \mathcal{A}}$ is nonincreasing in α on $(0, 1)$ by [26, Lemma 23(b)]. Thus, we can bound $\widetilde{C}_{\alpha, W, \mathcal{A}}^{\epsilon}$ using its definition given in (39):

$$\widetilde{C}_{\alpha, W, \mathcal{A}}^{\epsilon} = \frac{1}{\epsilon} \int_{\alpha - \epsilon \alpha}^{\alpha + \epsilon(1 - \alpha)} \left[1 \vee \left(\frac{\alpha}{1 - \alpha} \frac{1 - \eta}{\eta} \right) \right] C_{\eta, W, \mathcal{A}} \, d\eta$$

$$\leq \frac{1}{\epsilon} \int_{\alpha - \epsilon \alpha}^{\alpha + \epsilon(1 - \alpha)} \left[1 \vee \left(\frac{\alpha}{1 - \alpha} \frac{1 - \eta}{\eta} \right) \right] \left[1 \vee \left(\frac{1 - \alpha}{\alpha} \frac{\eta}{1 - \eta} \right) \right] C_{\alpha, W, \mathcal{A}} \, d\eta$$

$$= \frac{C_{\alpha, W, \mathcal{A}}}{\epsilon} \int_{\alpha - \epsilon \alpha}^{\alpha + \epsilon(1 - \alpha)} \left[\left(\frac{\alpha}{1 - \alpha} \frac{1 - \eta}{\eta} \right) \vee \left(\frac{1 - \alpha}{\alpha} \frac{\eta}{1 - \eta} \right) \right] d\eta$$

$$\leq \left(1 + \frac{\epsilon}{1 - \epsilon} \frac{\alpha^2 + (1 - \alpha)^2}{\alpha(1 - \alpha)} \right) C_{\alpha, W, \mathcal{A}}.$$

Then, for $\epsilon_n = 1/n$, (C.6) imply that

$$n\widetilde{C}_{\phi, W, \mathcal{A}}^{\epsilon} + \frac{\gamma_n}{1 - \phi} + \ln \frac{8e^3 n^{1.5}}{\epsilon}$$

$$\leq nC_{\phi, W, \mathcal{A}} + \frac{n}{n - 1} \frac{C_{\phi, W, \mathcal{A}}}{\phi(1 - \phi)} + \frac{40(K + 1) \ln(1 + n)}{1 - \phi} + 3 \ln(2en), \quad \forall n \geq n_0. \tag{C.7}$$

Thus, the hypothesis of Theorem 1 implies the hypothesis of Lemma 9 for all n large enough. Consequently, for all n large enough Lemma 9 and (C.6) implies that

$$P_{\mathrm{e}} \geq \left(\frac{(1 + n)^{-80(K + 1)}}{8e^3 n^{2.5}} \right)^{1/\phi} e^{-n \widetilde{E}_{\mathrm{sp}}^{\epsilon_n} \left(\frac{1}{n} \ln \frac{M_n}{L_n}, W, \mathcal{A} \right)}. \tag{C.8}$$

On the other hand, Lemma 8, (31), and the monotonicity of $C_{\alpha,W,\mathcal{A}}$ in α imply that for n large enough

$$\widetilde{E}_{\mathrm{sp}}^{\epsilon_n}\Big(\frac{1}{n}\ln\frac{M_n}{L_n},W,\mathcal{A}\Big) \leq E_{\mathrm{sp}}\Big(\frac{1}{n}\ln\frac{M_n}{L_n},W,\mathcal{A}\Big) + \frac{C_{\alpha_1,W,\mathcal{A}}}{(n-1)\phi^2}. \tag{C.9}$$

Now (32) follows from (C.8) and (C.9). $\triangle$

Proof of Theorem 2. We prove Theorem 2 using Lemmas 8 and 10. We are free to choose different values for κ and ϵ for different values of n provided that the hypotheses of Lemmas 8 and 10 are satisfied.

As a result of Assumption 2, there exists a $K \in [1,\infty)$ and an $n_0 \in \mathbb{Z}_+$ such that

$$\max_{t:\,t\leq n} C_{1/2,W_t,n\varrho} \leq K\ln n, \quad \forall n \geq n_0. \tag{C.10}$$

Let κ_n be $\kappa_n = K\ln(1+n)$. Then

$$\gamma_n \leq 40(K+1)\ln(1+n), \quad \forall n \geq n_0. \tag{C.11}$$

$C_{\alpha,W,\varrho}$ is nondecreasing in α by [26, Lemma 23(a)] and $\dfrac{1-\alpha}{\alpha}C_{\alpha,W,\varrho}$ is nonincreasing in α on $(0,1)$ by [26, Lemma 23(b)]. Thus, we can bound $\widetilde{C}_{\alpha,W,\varrho}^{\epsilon}$ using its definition given in (39):

$$\widetilde{C}_{\alpha,W_{[1,n]},n\varrho}^{\epsilon} \leq \Big(1 + \frac{\epsilon}{1-\epsilon}\frac{\alpha^2 + (1-\alpha)^2}{\alpha(1-\alpha)}\Big) C_{\alpha,W_{[1,n]},n\varrho}.$$

Then for $\epsilon_n = \dfrac{1}{n}$, (C.10), (C.11), and [26, Lemmas 23(a),(b), 27(a), and 28] imply that

$$\widetilde{C}_{\alpha,W_{[1,n]},n\varrho}^{\epsilon} + \frac{\gamma_n}{1-\phi} + \ln\frac{8e^3 n^{1.5}}{\epsilon}$$

$$\leq C_{\alpha,W_{[1,n]},n\varrho} + \frac{n}{n-1}\frac{K\ln(n)}{\phi(1-\phi)} + \frac{40(K+1)\ln(1+n)}{1-\phi} + 3\ln(2en), \quad \forall n \geq n_0. \tag{C.12}$$

Thus, the hypothesis of Theorem 2 implies the hypothesis of Lemma 10 for all n large enough. Consequently, for all n large enough Lemma 10 and (C.11) implies that

$$P_{\mathrm{e}} \geq \Big(\frac{(1+n)^{-80(K+1)}}{8e^3 n^{2.5}}\Big)^{1/\phi} e^{-\widetilde{E}_{\mathrm{sp}}^{\epsilon_n}\big(\ln\frac{M_n}{L_n},W_{[1,n]},n\varrho\big)}. \tag{C.13}$$

On the other hand, Lemma 8, (33), and the monotonicity of $C_{\alpha,W,\varrho}$ in α imply that for n large enough

$$\widetilde{E}_{\mathrm{sp}}^{\epsilon_n}\Big(\ln\frac{M_n}{L_n},W_{[1,n]},n\varrho\Big) \leq E_{\mathrm{sp}}\Big(\ln\frac{M_n}{L_n},W_{[1,n]},n\varrho\Big) + \frac{C_{\alpha_1,W_{[1,n]},n\varrho}}{(n-1)\phi^2}. \tag{C.14}$$

Note that $C_{\alpha,W_{[1,n]},n\varrho}$ is nondecreasing in α by [26, Lemma 23(a)] and $\dfrac{1-\alpha}{\alpha}C_{\alpha,W_{[1,n]},n\varrho}$ is nonincreasing in α on $(0,1)$ by [26, Lemma 23(b)]. Thus,

$$C_{\alpha_1,W_{[1,n]},n\varrho} \leq \Big(\frac{\alpha_1}{1-\alpha_1} \vee 1\Big)C_{1/2,W_{[1,n]},n\varrho}. \tag{C.15}$$

Since $C_{\alpha,W,\varrho}$ is nondecreasing in ϱ by [26, Lemma 27(a)], (C.10) and [26, Lemma 28] imply for all n large enough

$$C_{1/2,W_{[1,n]},n\varrho} \leq Kn\ln n. \tag{C.16}$$

For n large enough, (C.14)–(C.16) imply

$$\widetilde{E}_{\mathrm{sp}}^{\epsilon_n}\Big(\ln\frac{M_n}{L_n},W_{[1,n]},n\varrho\Big) \leq E_{\mathrm{sp}}\Big(\ln\frac{M_n}{L_n},W_{[1,n]},n\varrho\Big) + \frac{2}{\phi^2}\Big(\frac{\alpha_1}{1-\alpha_1} \vee 1\Big)K\ln n. \tag{C.17}$$

Now (34) follows from (C.13) and (C.17). $\triangle$

ACKNOWLEDGEMENTS

ADDITIONAL INFORMATION

This paper was presented in part at the 2017 IEEE International Symposium on Information Theory [1].

REFERENCES

1. Nakiboğlu, B., The Augustin Center and the Sphere Packing Bound for Memoryless Channels, in *Proc. 2017 IEEE Int. Symp. on Information Theory (ISIT'2017), Aachen, Germany, June 25–30, 2017*, pp. 1401–1405.

2. Elias, P., Coding for Noisy Channels, *IRE Conv. Rec.*, 1955, vol. 4, pp. 37–46. Reprinted in *Key Papers in the Development of Information Theory*, Slepian, D., Ed., New York: IEEE Press, 1974, pp. 102–111.

3. Elias, P., Coding for Two Noisy Channels, *Information Theory: Third London Symposium (London, England, Sept. 1955)*, Cherry C., Ed., London: Butterworth Sci., 1956, pp. 61–74.

4. Dobrushin, R.L., Asymptotic Estimates of the Probability of Error for Transmission of Messages over a Discrete Memoryless Communication Channel with a Symmetric Transition Probability Matrix, *Teor. Veroyatn. Primen.*, 1962, vol. 7, no. 3, pp. 283–311 [*Theory Probab. Appl.* (Engl. Transl.), 1962, vol. 7, no. 3, pp. 270–300].

5. Altuğ, Y. and Wagner, A.B., On Exact Asymptotics of the Error Probability in Channel Coding: Symmetric Channels, `arXiv:1908.11419 [cs.IT]`, 2019.

6. Altuğ, Y. and Wagner, A.B., Refinement of the Sphere Packing Bound for Symmetric Channels, in *Proc. 49th Annu. Allerton Conf. on Communication, Control, and Computing, Monticello, IL, Sept. 27–30, 2011*, pp. 30–37.

7. Shannon, C.E., Gallager, R.G., and Berlekamp, E.R., Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels. I, II, *Inform. Control*, 1967, vol. 10, no. 1, pp. 65–103; no. 5, pp. 522–552.

8. Gallager, R.G., *Information Theory and Reliable Communication*, New York: Wiley, 1968.

9. Fano, R.M., *Transmission of Information: A Statistical Theory of Communications*, New York: M.I.T. Press, 1961.

10. Haroutunian, E.A., Bounds for the Exponent of the Probability of Error for a Semicontinuous Memoryless Channel, *Probl. Peredachi Inf.*, 1968, vol. 4, no. 4, pp. 37–48 [*Probl. Inf. Transm.* (Engl. Transl.), 1968, vol. 4, no. 4, pp. 29–39].

11. Omura, J.K., A Lower Bounding Method for Channel and Source Coding Probabilities, *Inform. Control*, 1975, vol. 27, no. 2, pp. 148–177.

12. Csiszár, I. and Körner, J., *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Cambridge, UK: Cambridge Univ. Press, 2011, 2nd ed.

13. Altuğ, Y. and Wagner, A.B., Refinement of the Sphere-Packing Bound: Asymmetric Channels, *IEEE Trans. Inform. Theory*, 2014, vol. 60, no. 3, pp. 1592–1614.

14. Nakiboğlu, B., A Simple Derivation of the Refined SPB for the Constant Composition Codes, in *Proc. 2019 IEEE Int. Symp. on Information Theory (ISIT'2019), Paris, France, July 7–12, 2019*, pp. 2659–2663.

15. Wyner, A.D., Capacity and Error Exponent for the Direct Detection Photon Channel. II, *IEEE Trans. Inform. Theory*, 1988, vol. 34, no. 6, pp. 1462–1471.

16. Burnashev, M.V. and Kutoyants, Yu.A., On the Sphere-Packing Bound, Capacity, and Similar Results for Poisson Channels, *Probl. Peredachi Inf.*, 1999, vol. 35, no. 2, pp. 3–22 [*Probl. Inf. Transm.* (Engl. Transl.), 1999, vol. 35, no. 2, pp. 95–111].

17. Shannon, C.E., Probability of Error for Optimal Codes in a Gaussian Channel, *Bell Syst. Tech. J.*, 1959, vol. 38, no. 3, pp. 611–656.

18. Ebert, P.M., Error Bounds for Gaussian Noise Channels, in *Quarterly Progress Rep. of the Research Lab. of Electronics, MIT*, Cambridge, MA, 1965, no. 77, pp. 292–302. Available at `https://dspace.mit.edu/handle/1721.1/55609`.

19. Ebert, P.M., Error Bounds for Parallel Communication Channels, *Tech. Rep. of the Research Lab. of Electronics, MIT*, Cambridge, MA, 1966, no. 448. Available at `https://dspace.mit.edu/handle/1721.1/4295`.

20. Richters, J.S., Communication over Fading Dispersive Channels, *Tech. Rep. of the Research Lab. of Electronics, MIT*, Cambridge, MA, 1967, no. 464. Available at `https://dspace.mit.edu/handle/1721.1/4279`.

21. Lancho Serrano, A., Fundamental Limits of Short-Packet Wireless Communications, *PhD Thesis*, Departamento de Teoría de la Señal y Comunicaciones, Univ. Carlos III de Madrid, Spain, 2019. Available at `https://e-archivo.uc3m.es/handle/10016/29596`.

22. Lancho, A., Östman, J., Durisi, G., Koch, T., and Vazquez-Vilar, G., Saddlepoint Approximations for Short-Packet Wireless Communications, *IEEE Trans. Wireless Commun.*, 2020, vol. 19, no. 7, pp. 4831–4846.

23. Augustin, U., Error Estimates for Low Rate Codes, *Z. Wahrsch. Verw. Gebiete*, 1969, vol. 14, no. 1, pp. 61–88.

24. Augustin, U., Noisy Channels, *Habilitation Thesis*, Univ. Erlangen-Nürnberg, Germany, 1978. Avaliable at `http://bit.ly/2ID8h7m`.

25. Nakiboğlu, B., The Sphere Packing Bound via Augustin's Method, *IEEE Trans. Inform. Theory*, 2019, vol. 65, no. 2, pp. 816–840.

26. Nakiboğlu, B., The Augustin Capacity and Center, *Probl. Peredachi Inf.*, 2019, vol. 55, no. 4, pp. 3–51 [*Probl. Inf. Transm.* (Engl. Transl.), 2019, vol. 55, no. 4, pp. 299–342].

27. Nakiboğlu, B., The Rényi Capacity and Center, *IEEE Trans. Inform. Theory*, 2019, vol. 65, no. 2, pp. 841–860; see also `arXiv:1608.02424 [cs.IT]` (extended version).

28. Nakiboğlu, B., A Simple Derivation of the Refined Sphere Packing Bound under Certain Symmetry Hypotheses, *Turkish J. Math.*, 2020, vol. 44, no. 3, pp. 919–948.

29. Gallager, R.G., A Simple Derivation of the Coding Theorem and Some Applications, *IEEE Trans. Inform. Theory*, 1965, vol. 11, no. 1, pp. 3–18.

30. Poltyrev, G.Sh., Random Coding Bounds for Discrete Memoryless Channels, *Probl. Peredachi Inf.*, 1982, vol. 18, no. 1, pp. 12–26 [*Probl. Inf. Transm.* (Engl. Transl.), 1982, vol. 18, no. 1, pp. 9–21].

31. Blahut, R.E., Hypothesis Testing and Information Theory, *IEEE Trans. Inform. Theory*, 1974, vol. 20, no. 4, pp. 405–417.

32. Blahut, R.E., Information Bounds of the Fano–Kullback Type, *IEEE Trans. Inform. Theory*, 1976, vol. 22, no. 4, pp. 410–421.

33. Blahut, R.E., *Principles and Practice of Information Theory*, Reading, MA: Addison-Wesley, 1987.

34. Vazquez-Vilar, G., Martinez, A., and Guillén i Fàbregas, A., A Derivation of the Cost-Constrained Sphere-Packing Exponent, in *Proc. 2015 IEEE Int. Symp. on Information Theory (ISIT'2015), Hong Kong, China, June 14–19, 2015*, pp. 929–933.

35. Dudley, R.M., *Real Analysis and Probability*, Cambridge: Cambridge Univ. Press, 2002.

36. van Erven, T. and Harremoës, P., Rényi Divergence and Kullback–Leibler Divergence, *IEEE Trans. Inform. Theory*, 2014, vol. 60, no. 7, pp. 3797–3820.

37. Rudin, W., *Principles of Mathematical Analysis*, New York: McGraw-Hill, 1976, 3rd ed.

38. Bogachev, V.I., *Measure Theory*, Berlin: Springer, 2007.

39. Elias, P., List Decoding for Noisy Channels, *Tech. Rep. of the Research Lab. of Electronics, MIT*, Cambridge, MA, 1957, no. 335. Available at `https://dspace.mit.edu/handle/1721.1/4484`.

40. Dalai, M., Lower Bounds on the Probability of Error for Classical and Classical-Quantum Channels, *IEEE Trans. Inform. Theory*, 2013, vol. 59, no. 12, pp. 8027–8056.

41. Dalai, M., Some Remarks on Classical and Classical-Quantum Sphere Packing Bounds: Rényi vs. Kullback–Leibler, *Entropy*, 2017, vol. 19, no. 7, p. 355 (11 pp.).

42. Dalai, M. and Winter, A., Constant Compositions in the Sphere Packing Bound for Classical-Quantum Channels, *IEEE Trans. Inform. Theory*, 2017, vol. 63, no. 9, pp. 5603–5617.

43. Cheng, H.-C. and Hsieh, M.H., Moderate Deviation Analysis for Classical-Quantum Channels and Quantum Hypothesis Testing, *IEEE Trans. Inform. Theory*, 2018, vol. 64, no. 2, pp. 1385–1403.

44. Cheng, H.-C., Hsieh, M.H., and Tomamichel, M., Quantum Sphere-Packing Bounds with Polynomial Prefactors, *IEEE Trans. Inform. Theory*, 2019, vol. 65, no. 5, pp. 2872–2898.

45. Altuğ, Y. and Wagner, A.B., Refinement of the Random Coding Bound, *IEEE Trans. Inform. Theory*, 2014, vol. 60, no. 10, pp. 6005–6023.

46. Scarlett, J., Martinez, A., and Guillén i Fàbregas, A., Mismatched Decoding: Error Exponents, Second-Order Rates and Saddlepoint Approximations, *IEEE Trans. Inform. Theory*, 2014, vol. 60, no. 5, pp. 2647–2666.

47. Honda, J., Exact Asymptotics for the Random Coding Error Probability, in *Proc. 2015 IEEE Int. Symp. on Information Theory (ISIT'2015), Hong Kong, China, June 14–19, 2015*, pp. 91–95.

48. Burnashev, M.V. and Holevo, A.S., On the Reliability Function for a Quantum Communication Channel, *Probl. Peredachi Inf.*, 1998, vol. 34, no. 2, pp. 3–15 [*Probl. Inf. Transm.* (Engl. Transl.), 1998, vol. 34, no. 2, pp. 97–107].

49. Telatar, E., Capacity of Multi-antenna Gaussian Channels, *Eur. Trans. Telecommun.*, 1999, vol. 10, no. 6, pp. 585–595.

50. Vu, M., MISO Capacity with Per-Antenna Power Constraint, *IEEE Trans. Commun.*, 2011, vol. 59, no. 5, pp. 1268–1274.

51. Cao, P.L., Oechtering, T.J., Schaefer, R.F., and Skoglund, M., Optimal Transmit Strategy for MISO Channels with Joint Sum and Per-Antenna Power Constraints, *IEEE Trans. Signal Process.*, 2016, vol. 64, no. 16, pp. 4296–4306.

52. Loyka, S., The Capacity of Gaussian MIMO Channels under Total and Per-Antenna Power Constraints, *IEEE Trans. Commun.*, 2017, vol. 65, no. 3, pp. 1035–1043.

53. Cheng, H.-C. and Nakiboğlu, B., Refined Strong Converse for the Constant Composition Codes, in *Proc. 2020 IEEE Int. Symp. on Information Theory (ISIT'2020), Los Angeles, CA, June 21–26, 2020*, pp. 2149–2154.

54. Yang, W., Fading Channels: Capacity and Channel Coding Rate in the Finite-Blocklength Regime, *PhD Thesis*, Communication Systems Group, Dept. of Signals and Systems, Chalmers Univ. of Technology, Gothenburg, Sweden, 2015.