

“© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

# Quantum Sphere-Packing Bounds with Polynomial Prefactors

Hao-Chung Cheng<sup>1,2,3</sup>, Min-Hsiu Hsieh<sup>1</sup>, and Marco Tomamichel<sup>1,4</sup>

<sup>1</sup>*Centre for Quantum Software and Information (UTS:QISI),*

*Faculty of Engineering and Information Technology, University of Technology Sydney, Australia*

<sup>2</sup>*Graduate Institute Communication Engineering, National Taiwan University, Taiwan (R.O.C.)*

<sup>3</sup>*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, United Kingdom*

<sup>4</sup>*School of Physics, The University of Sydney, Australia*

**ABSTRACT.** We study lower bounds on the optimal error probability in classical coding over classical-quantum channels at rates below the capacity, commonly termed quantum sphere-packing bounds. Winter and Dalai have derived such bounds for classical-quantum channels; however, the exponents in their bounds only coincide when the channel is classical. In this paper, we show that these two exponents admit a variational representation and are related by the Golden-Thompson inequality, reaffirming that Dalai's expression is stronger in general classical-quantum channels. Second, we establish a finite blocklength sphere-packing bound for classical-quantum channels, which significantly improves Dalai's prefactor from the order of subexponential to polynomial. Furthermore, the gap between the obtained error exponent for constant composition codes and the best known classical random coding exponent vanishes in the order of  $o(\log n/n)$ , indicating our sphere-packing bound is almost exact in the high rate regime. Finally, for a special class of symmetric classical-quantum channels, we can completely characterize its optimal error probability without the constant composition code assumption. The main technical contributions are two converse Hoeffding bounds for quantum hypothesis testing and the saddle-point properties of error exponent functions.

## 1. INTRODUCTION

Shannon's noisy coding theorem [1] states that a message in an appropriately coded form can be reliably transmitted through a discrete memoryless channel  $\mathcal{W}$ , provided the coding rate  $R$  is below the channel capacity  $C_{\mathcal{W}}$ . More precisely, the probability of decoding errors can be made arbitrarily small as the coding blocklength grows. Later, Shannon himself pioneered the study of the exponential dependency of the optimal error probability  $\epsilon^*(n, R)$  for a blocklength  $n$  and transmission rate  $R$  [2]. He defined the *reliability function* to be, for any fixed coding rate  $R < C_{\mathcal{W}}$ ,

$$E(R) := \limsup_{n \rightarrow +\infty} -\frac{1}{n} \log \epsilon^*(n, R).$$

The quantity  $E(R)$  then provides a measure of how rapidly the error probability approaches zero with an increase in blocklength. This characterization of the reliability function is hence called the *reliability function analysis* or the *error exponent analysis*.

For a classical channel, lower bounds for the reliability function can be established by random coding arguments [3, 4, 5, 6]. However, upper bounds require different techniques since the code-dependent bounds on the error probability need to be optimized over all codebooks. The first result—the *sphere-packing bound*  $E(R) \leq E_{\text{sp}}(R)$ —was studied by Fano [4] and Shannon, Gallager, and Berlekamp [7], and proved by the later paper. The *sphere-packing exponent*  $E_{\text{sp}}(R)$  is defined as

$$E_{\text{sp}}(R) := \sup_{s \geq 0} \left\{ \max_P E_0(s, P) - sR \right\}, \quad (1)$$

where  $P$  is maximized over all probability distributions on the input alphabet, and  $E_0(s, P)$  is the *auxiliary function* or *Gallager's function* [5]. Unlike Shannon-Gallager-Berlekamp's technique which relates channel

---

*E-mail address:* HaoChung.Ch@gmail.com, Min-Hsiu.Hsieh@uts.edu.au, marco.tomamichel@uts.edu.au.

coding to binary hypothesis testing, Haroutunian [8, 9] employed a combinatorial method and obtained an upper bound for the reliability function in terms of the following expression

$$\tilde{E}_{\text{sp}}(R) := \max_P \min_{\mathcal{V}} \{D(\mathcal{V}||\mathcal{W}|P) : I(P, \mathcal{V}) \leq R\}, \quad (2)$$

where  $\mathcal{V}$  is minimized over all channels with the same output alphabet as  $\mathcal{W}$ ,  $D(\mathcal{V}||\mathcal{W}|P)$  is the conditional relative entropy between the dummy channel  $\mathcal{V}$  and the true channel  $\mathcal{W}$ , and  $I(P, \mathcal{V})$  is the mutual information of the channel  $\mathcal{V}$  (the detailed definitions are given in Section 2). It was then realized that the two quantities in Eqs. (1) and (2) are equivalent: they are related by convex program duality [8, 10, 11, 12]. Therefore, these two expressions, Eqs. (1) or (2), are both called sphere-packing exponents.

Error exponent analysis in classical-quantum (c-q) channels is more challenging because of the noncommutative nature of quantum mechanics. Burnashev and Holevo [13] introduced a quantum version of the auxiliary function [14, 15] and initialized the study of reliability functions in c-q channels. Winter [16] derived a sphere-packing bound for c-q channels in the form of  $\tilde{E}_{\text{sp}}(R)$  in Eq. (2), generalizing Haroutunian's idea [8]. Dalai [17] employed Shannon-Gallager-Berlekamp's approach [7] to establish a sphere-packing bound with Gallager's exponent in Eq. (1). In the follow-up work [18], Dalai and Winter pointed out that these two exponents are not equal in c-q channels. In this work, we explicitly demonstrate a relationship between the two quantities. Precisely, we show that they individually admit a variational representation (Theorem 6 in Section 3):

$$E_{\text{sp}}(R) = \max_P \sup_{0 < \alpha \leq 1} \min_{\sigma} \left\{ \frac{1 - \alpha}{\alpha} \left( \sum_x P(x) D_{\alpha}(W_x || \sigma) - R \right) \right\}; \quad (3)$$

$$\tilde{E}_{\text{sp}}(R) = \max_P \sup_{0 < \alpha \leq 1} \min_{\sigma} \left\{ \frac{1 - \alpha}{\alpha} \left( \sum_x P(x) D_{\alpha}^b(W_x || \sigma) - R \right) \right\}, \quad (4)$$

where  $\sigma$  is minimized over all density operators on some Hilbert space  $\mathcal{H}$ ;  $W_x$  is the channel output state on  $\mathcal{H}$ ;  $D_{\alpha}$  is the (Petz)  $\alpha$ -Rényi divergence [19]; and  $D_{\alpha}^b$  is the *log-Euclidean*  $\alpha$ -Rényi divergence.

Since  $D_{\alpha} \leq D_{\alpha}^b$  for all  $\alpha \in (0, 1]$ , as a simple consequence of the Golden-Thompson inequality [20, 21], the exponent  $E_{\text{sp}}(R)$  in Eq. (1) is stronger than  $\tilde{E}_{\text{sp}}(R)$  in Eq. (2), i.e.

$$E(R) \leq E_{\text{sp}}(R) \leq \tilde{E}_{\text{sp}}(R).$$

These two exponents coincide for all  $R$  only when all the channel output states commute<sup>1</sup> (i.e. for classical channels). Thus, we call  $E_{\text{sp}}(R)$  and  $\tilde{E}_{\text{sp}}(R)$  the *strong sphere-packing exponent* and the *weak sphere-packing exponent*, respectively. The lower bounds for the optimal error probability in terms of these two quantities are called the strong sphere-packing bound

$$\epsilon^*(n, R) \geq f(n) \exp \{-n [E_{\text{sp}}(R - g(n))]\}, \quad (5)$$

and the weak sphere-packing bound

$$\epsilon^*(n, R) \geq f(n) \exp \left\{ -n \left[ \tilde{E}_{\text{sp}}(R - g(n)) \right] \right\}, \quad (6)$$

where  $f(n)$  is the prefactor of the bound, and  $g(n)$  is a rate back-off term. We note that  $g(n) = 0$  in our main result, and hence we only study  $f(n)$  in the following discussion.

The strong sphere-packing bound obtained by Dalai [17] had a prefactor  $f(n) = e^{-O(\sqrt{n})}$ , which is loose for small blocklength  $n$  or in the situation where the transmission rate is close to channel capacity. Furthermore, such bound only holds asymptotically, i.e. when  $n$  tends to infinity. The main contribution of this paper is to establish a sphere-packing bound with a better prefactor  $f(n) = O(n^{-t})$  for some  $t > 1/2$ , which not only holds for finite blocklength  $n$  but also notably improves Dalai's bound [17] from the order of subexponential to polynomial (Corollary 9). When restricting to constant composition codes, we can be more explicit about the obtained prefactor, namely,  $f(n) = n^{-\frac{1}{2}(1 + |E'_{\text{sp}}(R)| + o(1))}$  (Theorem 8)<sup>2</sup>. Moreover,

<sup>1</sup>For the coding rates above channel capacity, these two exponents are both zero ( $\alpha$  attains 1 in Eqs. (1) and (2)). We exclude this trivial case and only consider the rate being strictly below capacity.

<sup>2</sup>The notion  $E'_{\text{sp}}(R)$  means the left-derivative of  $E_{\text{sp}}(R)$ . Note that  $E_{\text{sp}}(R)$  is not necessarily differentiable.

this sphere-packing bound and the best known random coding upper bound [22, 23, 24, 25] in the classical case coincide up to the third-order term (see the discussion in Section 4)). Hence, our result yields a tight asymptotics of the sphere-packing bound for constant composition codes. Our second contribution is to show that, for a class of symmetric c-q channels, the prefactor  $f(n) = O(n^{-\frac{1}{2}(1+|E'_{\text{sp}}(R)|)})$ , holds for general codes. In other words, we are able to obtain a tight sphere-packing bound for general codes, by exploiting a symmetric property of the channel.

Our main ingredients are a tight concentration inequality in strong large deviation theory [26], [27, Theorem 3.7.4], [28, Section III.D] (Appendix B), an one-shot converse via quantum hypothesis testing [4, 7, 10], and a uniform continuity property (Proposition 18 in Appendix C). The strategy of the proof consists of three steps: (i) formulate the error probability of a certain codebook to a hypothesis testing problem; (ii) prove lower (or called the converse) bounds on type-I error in quantum hypothesis testing; and (iii) relate the error with the strong sphere-packing exponent. In Section 4.1, we first prove a one-shot converse bound to relate the channel coding problem into a binary quantum hypothesis (Proposition 10). We then employ Bahadur-Ranga Rao's inequality [26] to establish converse Hoeffding bounds for quantum hypothesis testing (Theorem 11). Next, we apply a uniform continuity property (Proposition 18 in Appendix C) and the established sharp Hoeffding bound (Theorem 11) to prove two finite blocklength converse bounds on the optimal type-I error for a fixed-composition codeword. The first bound is a Chebyshev-type bound with a subexponential prefactor (Proposition 12), while the second bound is a sharp converse bound with a polynomial prefactor (Proposition 13). Finally, we combine these two results to obtain a refined strong sphere-packing bound with a polynomial prefactor and finite blocklength.

Table 1 collects major proof approaches of classical sphere-packing bounds, Eqs. (1) and (1), and discusses their generalizations to c-q channels. We remark that the established finite blocklength bounds and the polynomial prefactor are crucial for the analysis of coding performance in the medium error probability regime (more commonly known as moderate deviation analysis) [28, 29, 30], classical data compression with quantum side information [31, 32], and joint source-channel coding with quantum side information [33].

The remaining part of the paper is organized as follows. Section 2 introduces the notation and necessary preliminaries. The relationship between the weak and strong sphere-packing exponents is proved in Section 3. In Section 4, we prove a refined sphere-packing bound for c-q channels. We consider a symmetric c-q channel and establish an exact sphere-packing bound in Section 5. Lastly, we conclude this paper in Section 6.

## 2. NOTATION AND PRELIMINARIES

Throughout this paper, we consider a finite-dimensional Hilbert space  $\mathcal{H}$ . The set of density operators (i.e. positive semi-definite operators with unit trace) and the set of full-rank density operators on  $\mathcal{H}$  are defined as  $\mathcal{S}(\mathcal{H})$  and  $\mathcal{S}_{>0}(\mathcal{H})$ , respectively. For  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ , we write  $\rho \ll \sigma$  if the support of  $\rho$  is contained in the support of  $\sigma$ . The identity operator on  $\mathcal{H}$  is denoted by  $\mathbb{1}_{\mathcal{H}}$ . If there is no possibility of confusion, we will skip the subscript  $\mathcal{H}$ . We use  $\text{Tr}[\cdot]$  to denote the trace. Let  $\mathbb{N}$ ,  $\mathbb{R}$ ,  $\mathbb{R}_{\geq 0}$ , and  $\mathbb{R}_{>0}$  denote the set of integers, real numbers, non-negative real numbers, and positive real numbers, respectively. Define  $[n] := \{1, 2, \dots, n\}$  for  $n \in \mathbb{N}$ .

For a positive semi-definite operator  $A$  whose spectral decomposition is  $A = \sum_i a_i P_i$ , where  $(a_i)_i$  and  $(P_i)_i$  are the eigenvalues and eigenprojections of  $A$ , its power is defined as:  $A^p := \sum_{i:a_i \neq 0} a_i^p P_i$ . In particular,  $A^0$  denotes the projection onto  $\text{supp}(A)$ , where we use  $\text{supp}(A)$  to denote the support of the operator  $A$ . Further,  $A \perp B$  means  $\text{supp}(A) \cap \text{supp}(B) = \emptyset$ , and  $A \ll B$  indicates  $\text{supp}(A) \subseteq \text{supp}(B)$ . We denote by  $\log$  the natural logarithm. We use  $f \vee g$  (resp.  $a \wedge b$ ) to denote the pointwise maximum (resp. minimum) between two functions  $f$  and  $g$ .

**2.1. Information Quantities and Error-Exponent Functions.** Given a pair of positive semi-definite operators  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ , we define quantum relative entropy [44, 45] and relative variance [46, 47, 48],

<sup>3</sup>Blahut in Ref. [10] claimed that the sphere-packing bound can be derived without using constant composition argument. However, there is a non-trivial gap. We refer readers to the discussion by Nakiboğlu in Ref. [42, Appendix A]

Bounds\Settings	Asymptotics	Composition dependent	Pre-factor $f(n)$	Rate back-off $g(n)$	Classical-Quantum channels	Expression
(a) Shannon-Gallager-Berlekamp [7]	Asymptotic	Yes	$e^{-O(\sqrt{n})}$	$O\left(\frac{\log n}{n}\right)$	Dalai [17]	$E_{\text{sp}}(R)$
(b) Haroutunian [8] Omura [34] Csisár-Korner [12]	Asymptotic	Yes	$e^{-o(n)}$	$o(1)$	Winter [16]	$\tilde{E}_{\text{sp}}(R)$
(c) Blahut [10]	Asymptotic	Yes <sup>3</sup>	$e^{-O(\sqrt{n})}$	$O\left(n^{-\frac{1}{2}}\right)$	Eqs. (4.2) & (4.2)	$E_{\text{sp}}(R)$
(d) Altuğ-Wagner [35]	Finite Blocklength	Yes	$n^{-\frac{1}{2}(1+ E'_{\text{sp}}(R) +o(1))}$	0	Theorem 8	$E_{\text{sp}}(R)$
(e) Elkayam-Feder [36]	Asymptotic	Yes	$O(n^{-t})$	$O\left(\frac{\log n}{n}\right)$	Unknown	Unknown
(f) Agustin-Nakiboğlu [37, 38, 39, 40, 41, 42]	Finite Blocklength	No	$O(n^{-t})$	0	Unknown	Unknown

TABLE 1. Different sphere-packing bounds are compared by (i) whether the bounds hold for finite blocklength  $n$  or hold asymptotically as  $n \rightarrow +\infty$ ; (ii) whether or not they are dependent on the constant composition codes; (iii) & (iv) the asymptotics  $f(n)$  and  $g(n)$ ; (v) the corresponding generalizations to classical-quantum channel coding; The parameter  $t$  in rows (e) and (f) is some value in the range  $t > 1/2$ ; and (vi) whether their error exponent expressions for c-q channels are expressed by  $E_{\text{sp}}(R)$  given in Eq. (1) or by  $\tilde{E}_{\text{sp}}(R)$  defined in Eq. (1).

respectively as

$$D(\rho\|\sigma) := \text{Tr} [\rho (\log \rho - \log \sigma)];$$

$$V(\rho\|\sigma) := \text{Tr} \left[ \rho (\log \rho - \log \sigma)^2 \right] - D(\rho\|\sigma)^2,$$

when  $\rho \ll \sigma$ , and  $+\infty$  otherwise.

For two positive definite operators  $\rho, \sigma > 0$  on  $\mathcal{H}$ , and every  $\alpha \in (0, 1)$ , we define the following two families of quantum Rényi divergences [19, 49, 50]:

$$D_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log Q_\alpha(\rho\|\sigma), \quad Q_\alpha(\rho\|\sigma) := \text{Tr} [\rho^\alpha \sigma^{1-\alpha}]; \quad (7)$$

$$D_\alpha^b(\rho\|\sigma) := \frac{1}{\alpha - 1} \log Q_\alpha^b(\rho\|\sigma), \quad Q_\alpha^b(\rho\|\sigma) := \text{Tr} \left[ e^{\alpha \log \rho + (1-\alpha) \log \sigma} \right]. \quad (8)$$

We term the above quantities as the *(Petz)  $\alpha$ -Rényi divergence*, and the *log-Euclidean  $\alpha$ -Rényi divergence*, respectively. The log-Euclidean Rényi divergence arises from the *log-Euclidean operator mean* (also called the *chaotic mean*):  $A \diamond_\alpha B := \exp((1-\alpha) \log A + \alpha \log B)$  for  $0 \leq \alpha \leq 1$ . For general positive semi-definite operators  $\rho, \sigma \geq 0$ , the above definitions can be extended as

$$Q_\alpha(\rho\|\sigma) := \lim_{\delta \downarrow 0} Q_\alpha(\rho + \delta \mathbb{1} \|\sigma + \delta \mathbb{1}) \quad \text{and} \quad Q_\alpha^b(\rho\|\sigma) := \lim_{\delta \downarrow 0} Q_\alpha^b(\rho + \delta \mathbb{1} \|\sigma + \delta \mathbb{1}).$$

From the Golden-Thompson inequality [20, 21]:

$$\text{Tr} [e^{A+B}] \leq \text{Tr} [e^A e^B], \quad \forall A, B \geq 0,$$

these two quantities are related by

$$Q_\alpha^b(\rho\|\sigma) \leq Q_\alpha(\rho\|\sigma), \quad \forall \alpha \in (0, 1). \quad (9)$$

For  $\alpha = 1$  and  $\alpha = 0$ , we define (see e.g. [50, Lemma 3.5]):

$$D_1(\rho\|\sigma) := \lim_{\alpha \uparrow 1} D_\alpha(\rho\|\sigma) = D(\rho\|\sigma), \quad D_1^b(\rho\|\sigma) := \lim_{\alpha \uparrow 1} D_\alpha^b(\rho\|\sigma) = D(\rho\|\sigma);$$

$$D_0(\rho\|\sigma) := \lim_{\alpha \downarrow 0} D_\alpha(\rho\|\sigma), \quad D_0^b(\rho\|\sigma) := \lim_{\alpha \downarrow 0} D_\alpha^b(\rho\|\sigma).$$

We will need the properties of the Rényi divergence for the next section.

**Lemma 1.** *The following hold:*

- (a) *For every  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ , the map  $\alpha \mapsto D_\alpha(\rho\|\sigma)$  is continuous and monotone increasing on  $[0, 1]$ .*
- (b) *Let  $\rho \in \mathcal{S}(\mathcal{H})$ , positive semi-definite operators  $\sigma_1$  and  $\sigma_2$  on  $\mathcal{H}$ , and  $\alpha \in [0, 1]$ . If  $\sigma_1 \leq \sigma_2$ , then  $D_\alpha(\rho\|\sigma_1) \geq D_\alpha(\rho\|\sigma_2)$ . Moreover, if  $\sigma_1 = \gamma\sigma_2$  for some  $\gamma > 0$ , then  $D_\alpha(\rho\|\sigma_1) = D_\alpha(\rho\|\sigma_2) - \log \gamma$ .*
- (c) *For every  $\rho \in \mathcal{S}(\mathcal{H})$  and  $\alpha \in [0, 1]$ , the map  $\sigma \mapsto D_\alpha(\rho\|\sigma)$  is convex and lower semi-continuous on  $\mathcal{S}(\mathcal{H})$ .*

We note that item (a) was proved in [50, Lemma 3.12, Corollary 3.15]; item (b) was proved in [50, Lemma 3.24]; item (c) was shown in [51, 52, 19] [50, Theorem 3.16]<sup>4</sup>.

Let  $\mathcal{X} = \{1, 2, \dots, |\mathcal{X}|\}$  be a finite alphabet, and let  $\mathcal{P}(\mathcal{X})$  be the set of probability distributions on  $\mathcal{X}$ . A classical-quantum (c-q) channel  $\mathcal{W}$  maps elements of the finite set  $\mathcal{X}$  to density operators in  $\mathcal{S}(\mathcal{H})$ , i.e.  $\mathcal{W} : x \mapsto W_x$ . For a c-q channel  $\mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$  and  $P \in \mathcal{P}(\mathcal{X})$ , it is convenient to denote the corresponding c-q state:

$$P \circ \mathcal{W} := \sum_{x \in \mathcal{X}} P(x) |x\rangle\langle x| \otimes W_x.$$

We also express the input distribution  $P \in \mathcal{P}(\mathcal{X})$  as a diagonal matrix with respect to the computational basis  $\{|x\rangle\}_{x \in \mathcal{X}}$ , i.e.  $P = \sum_{x \in \mathcal{X}} P(x) |x\rangle\langle x|$ . Denote the conditional relative entropy of two c-q channels  $\mathcal{V}, \mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$  with a prior distribution  $P \in \mathcal{P}(\mathcal{X})$  by

$$D(\mathcal{V}\|\mathcal{W}|P) := \sum_{x \in \mathcal{X}} P(x) D(V_x\|W_x).$$

Similarly, we define the following conditional entropic quantities for  $\mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$ ,  $\sigma \in \mathcal{S}(\mathcal{H})$  and  $P \in \mathcal{P}(\mathcal{X})$ :

$$\begin{aligned} D(\mathcal{W}\|\sigma|P) &:= \sum_{x \in \mathcal{X}} P(x) D(W_x\|\sigma), \\ D_\alpha(\mathcal{W}\|\sigma|P) &:= \sum_{x \in \mathcal{X}} P(x) D_\alpha(W_x\|\sigma), \\ D_\alpha^\flat(\mathcal{W}\|\sigma|P) &:= \sum_{x \in \mathcal{X}} P(x) D_\alpha^\flat(W_x\|\sigma). \end{aligned} \tag{10}$$

The *mutual information* of the prior distribution  $P \in \mathcal{P}(\mathcal{X})$  and the c-q channel  $\mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$  is defined as

$$I(P, \mathcal{W}) := \inf_{\sigma \in \mathcal{S}(\mathcal{H})} D(\mathcal{W}\|\sigma|P) = D(\mathcal{W}\|P\mathcal{W}|P), \tag{11}$$

where  $P\mathcal{W} := \sum_{x \in \mathcal{X}} P(x) W_x$  and the second equality can be found in Ref. [53]. The (classical) *capacity* of the channel  $\mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$  is denoted by [54, 55]:

$$C_{\mathcal{W}} := \max_{P \in \mathcal{P}(\mathcal{X})} I(P, \mathcal{W}).$$

We define two related information quantities: for every  $\alpha \in [0, 1]$ ,

$$\begin{aligned} I_\alpha^{(1)}(P, \mathcal{W}) &:= \inf_{\sigma \in \mathcal{S}(\mathcal{H})} D_\alpha(P \circ \mathcal{W}\|P \otimes \sigma); \\ I_\alpha^{(2)}(P, \mathcal{W}) &:= \inf_{\sigma \in \mathcal{S}(\mathcal{H})} D_\alpha(\mathcal{W}\|\sigma|P). \end{aligned} \tag{12}$$

The term  $I_\alpha^{(1)}(P, \mathcal{W})$  is called the *order  $\alpha$  Rényi information* [56, 57, 39] or the *generalized Holevo quantity*. The second term  $I_\alpha^{(2)}(P, \mathcal{W})$  can be viewed as a variant of the  $\alpha$ -Rényi mutual information, and we call it

<sup>4</sup>It was shown in [50, Corollary 3.27] that the map  $\sigma \mapsto D_\alpha(\rho\|\sigma)$  is lower semi-continuous on  $\mathcal{S}(\mathcal{H})$  for all  $\alpha \in (0, 1)$ . The argument can be extended to the range  $\alpha \in [0, 1]$  by the same method in [50, Lemma 3.26, Corollary 3.27].

the *order  $\alpha$  Augustin information* [41, 42, 85]. It can be verified that these two functions are related by Jensen's inequality:

$$I_\alpha^{(1)}(P, \mathcal{W}) \leq I_\alpha^{(2)}(P, \mathcal{W}). \quad (13)$$

For the case of  $\alpha = 1$ , they both equal conventional mutual information, i.e.  $I_1^{(1)}(P, \mathcal{W}) = I_1^{(2)}(P, \mathcal{W}) = I(P, \mathcal{W})$ . Mosonyi and Ogawa [50, Proposition 4.2] showed that for all  $\alpha \in [0, 1]$ ,

$$C_{\alpha, \mathcal{W}} := \sup_{P \in \mathcal{P}(\mathcal{X})} I_\alpha^{(1)}(P, \mathcal{W}) = \sup_{P \in \mathcal{P}(\mathcal{X})} I_\alpha^{(2)}(P, \mathcal{W}), \quad (14)$$

and it is termed the *Rényi radius* or the *Rényi capacity* of order  $\alpha$ . It is not hard to verify that  $C_{\alpha, \mathcal{W}}$  equals the usual channel capacity  $C_{\mathcal{W}}$  as  $\alpha = 1$ . Moreover, Proposition 2 below and the compactness of  $\mathcal{P}(\mathcal{X})$  show that the suprema in Eq. (2.1) can be replaced with maxima. The following proposition presents important properties of  $\alpha$ -Augustin mutual information and radius. The proof is given in Appendix D.

**Proposition 2** (Properties of order  $\alpha$  Augustin Information and Radius). *Given any classical-quantum channel  $\mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$  with  $|\mathcal{X}| < \infty$ , the following hold:*

- (a) *For every  $P \in \mathcal{P}(\mathcal{X})$ ,  $\alpha \mapsto I_\alpha^{(2)}(P, \mathcal{W})$  is monotone increasing on  $[0, 1]$ , and  $I_\alpha^{(2)}(P, \mathcal{W}) \leq \log |\mathcal{X}|$  for all  $\alpha \in [0, 1]$ .*
- (b) *For every  $(\alpha, P) \in (0, 1] \times \mathcal{P}(\mathcal{X})$ , there exists a unique  $\sigma_{\alpha, P} \in \mathcal{S}(\mathcal{H})$ , termed Augustin mean, such that*

$$I_\alpha^{(2)}(P, \mathcal{W}) = D_\alpha(\mathcal{W} \| \sigma_{\alpha, P} | P),$$

and

$$\mathsf{T}_{\alpha, P}(\sigma) = \sigma \text{ and } \sigma \gg P\mathcal{W} \text{ if and only if } \sigma = \sigma_{\alpha, P}, \quad (15)$$

where the map  $\mathsf{T}_{\alpha, P} : \mathcal{S}_{P, \mathcal{W}}(\mathcal{H}) \rightarrow \mathcal{S}(\mathcal{H})$  is defined as

$$\mathsf{T}_{\alpha, P}(\sigma) = \sum_{x \in \mathcal{X}} P(x) \frac{\sigma^{\frac{1-\alpha}{2}} W_x^\alpha \sigma^{\frac{1-\alpha}{2}}}{\text{Tr}[W_x^\alpha \sigma^{1-\alpha}]}.$$

- (c) *For every  $\alpha \in [0, 1]$ , the map  $P \mapsto I_\alpha^{(2)}(P, \mathcal{W})$  is concave on  $\mathcal{P}(\mathcal{X})$ .*
- (d) *For every  $P \in \mathcal{P}(\mathcal{X})$ ,  $\alpha \mapsto \frac{1-\alpha}{\alpha} I_\alpha^{(2)}(P, \mathcal{W})$  is concave on  $(0, 1]$ .*
- (e) *For every  $P \in \mathcal{P}(\mathcal{X})$ ,  $\alpha \mapsto I_\alpha^{(2)}(P, \mathcal{W})$  is continuous on  $[0, 1]$ .*
- (f) *The family of functions  $\{I_\alpha^{(2)}(P, \mathcal{W})\}_{\alpha \in [0, 1]}$  is uniformly equicontinuous in  $P \in \mathcal{P}(\mathcal{X})$ . Moreover, The map  $(\alpha, P) \mapsto I_\alpha^{(2)}(P, \mathcal{W})$  is jointly continuous on  $[0, 1] \times \mathcal{P}(\mathcal{X})$ .*
- (g) *The map  $(\alpha, P) \mapsto \sigma_{\alpha, P}$  is jointly continuous on  $(0, 1] \times \mathcal{P}(\mathcal{X})$ .*
- (h) *The map  $\alpha \mapsto C_{\alpha, \mathcal{W}}$  is continuous and monotone increasing on  $[0, 1]$ .*

The *strong sphere-packing exponent* [17] of a c-q channel  $\mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$  and a rate  $R \geq 0$  is defined by

$$E_{\text{sp}}(R) := \max_{P \in \mathcal{P}(\mathcal{X})} E_{\text{sp}}(R, P), \quad (16)$$

where

$$E_{\text{sp}}(R, P) := \sup_{s \geq 0} \{E_0(s, P) - sR\}, \quad (17)$$

and  $E_0$  is the *auxiliary function* of the c-q channel  $\mathcal{W}$  (see [13, 14, 15]):

$$E_0(s, P) := -\log \text{Tr} \left[ \left( \sum_{x \in \mathcal{X}} P(x) \cdot W_x^{1/(1+s)} \right)^{1+s} \right]$$

for all  $P \in \mathcal{P}(\mathcal{X})$  and  $s \geq 0$ .



The *weak sphere-packing exponent* [16] is defined as

$$\tilde{E}_{\text{sp}}(R) := \max_{P \in \mathcal{P}(\mathcal{X})} \tilde{E}_{\text{sp}}(R, P),$$

where

$$\tilde{E}_{\text{sp}}(R, P) := \min_{\mathcal{V}: \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})} \{D(\mathcal{V} \parallel \mathcal{W} | P) : I(P, \mathcal{V}) \leq R\}. \quad (18)$$

We also need the following definitions: for any  $R \geq 0$  and  $P \in \mathcal{P}(\mathcal{X})$ ,

$$E_{\text{sp}}^{(1)}(R, P) := \sup_{0 < \alpha \leq 1} \frac{1 - \alpha}{\alpha} \left( I_{\alpha}^{(1)}(P, \mathcal{W}) - R \right); \quad (19)$$

$$E_{\text{sp}}^{(2)}(R, P) := \sup_{0 < \alpha \leq 1} \frac{1 - \alpha}{\alpha} \left( I_{\alpha}^{(2)}(P, \mathcal{W}) - R \right), \quad (20)$$

Eq. (2.1) implies that (see also Theorem 6)  $E_{\text{sp}}^{(1)}(R, P) \leq E_{\text{sp}}^{(2)}(R, P)$ . By quantum Sibson's identity [58]<sup>5</sup>, one finds

$$E_{\text{sp}}^{(1)}(R, P) = E_{\text{sp}}(R, P). \quad (21)$$

Proposition 2 and Eq. (2.1) imply that the two quantities given in Eqs. (2.1) and (2.1) are equal to the strong sphere-packing exponent by maximizing over the input distributions:

$$E_{\text{sp}}(R) = \max_{P \in \mathcal{P}(\mathcal{X})} E_{\text{sp}}^{(1)}(R, P) = \max_{P \in \mathcal{P}(\mathcal{X})} E_{\text{sp}}^{(2)}(R, P). \quad (22)$$

In Proposition 4 below, one has  $E_{\text{sp}}(R) = +\infty$  for  $R < C_{0, \mathcal{W}}$ , and  $E_{\text{sp}}(R) = 0$  as  $R > C_{\mathcal{W}}$ . Throughout this paper, we further assume that the considered c-q channel  $\mathcal{W}$  satisfies  $C_{0, \mathcal{W}} < C_{\mathcal{W}}$ . We note that  $C_{0, \mathcal{W}}$  gives an upper bound to the *zero-error capacity* of the channel. For details, we refer the readers to Ref. [17].

As we will show in Section 4, the quantity  $E_{\text{sp}}^{(2)}(R, P)$  plays a significant role in the connection between hypothesis testing and channel coding. Moreover, Proposition 3 below shows that the optimizer in Eqs. (2.1) and (2.1) forms a saddle-point. The proof closely follows Altuğ and Wagner [35, Proposition 1], and is given in Appendix E.

**Proposition 3 (Saddle-Point).** *Consider a classical-quantum channel  $\mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$ , any  $R \in (C_{0, \mathcal{W}}, C_{\mathcal{W}})$ , and  $P \in \mathcal{P}(\mathcal{X})$ . Let<sup>6</sup>*

$$\mathcal{S}_{P, \mathcal{W}}(\mathcal{H}) := \{\sigma \in \mathcal{S}(\mathcal{H}) : \forall x \in \text{supp}(P), W_x \not\perp \sigma\}.$$

Define

$$F_{R, P}(\alpha, \sigma) := \begin{cases} \frac{1 - \alpha}{\alpha} (D_{\alpha}(\mathcal{W} \parallel \sigma | P) - R), & \alpha \in (0, 1) \\ 0, & \alpha = 1 \end{cases}, \quad (23)$$

on  $(0, 1] \times \mathcal{S}(\mathcal{H})$ , and denote by

$$\mathcal{P}_R(\mathcal{X}) := \left\{ P \in \mathcal{P}(\mathcal{X}) : \sup_{0 < \alpha \leq 1} \inf_{\sigma \in \mathcal{S}(\mathcal{H})} F_{R, P}(\alpha, \sigma) \in \mathbb{R}_{>0} \right\}. \quad (24)$$

The following holds

(a) For any  $P \in \mathcal{P}(\mathcal{X})$ ,  $F_{R, P}(\cdot, \cdot)$  has a saddle-point on  $(0, 1] \times \mathcal{S}_{P, \mathcal{W}}(\mathcal{H})$  with the saddle-value:

$$\min_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{0 < \alpha \leq 1} F_{R, P}(\alpha, \sigma) = \sup_{0 < \alpha \leq 1} \min_{\sigma \in \mathcal{S}(\mathcal{H})} F_{R, P}(\alpha, \sigma) = E_{\text{sp}}^{(2)}(R, P).$$

<sup>5</sup> For joint state  $\rho_{AB} \in \mathcal{S}(AB)$  with marginal states denoted by  $\rho_A$  and  $\rho_B$ , quantum Sibson's identity [58] states that the minimizer of  $\min_{\sigma \in \mathcal{S}(\mathcal{H})} D_{\alpha}(\rho_{AB} \parallel \rho_A \otimes \sigma_B)$  is  $(\text{Tr}_A[\rho_{AB}^{\alpha}])^{1/\alpha} / \text{Tr}[(\text{Tr}_A[\rho_{AB}^{\alpha}])^{1/\alpha}]$ . For the case of classical-quantum channels, the minimizer of  $\min_{\sigma \in \mathcal{H}} D_{\alpha}(P \circ \mathcal{W} \parallel P \otimes \sigma)$  is then  $(\sum_{x \in \mathcal{X}} P(x) W_x^{\alpha})^{1/\alpha} / \text{Tr}[(\sum_{x \in \mathcal{X}} P(x) W_x^{\alpha})^{1/\alpha}]$ .

<sup>6</sup>For  $\alpha \in (0, 1]$ ,  $\mathcal{S}_{P, \mathcal{W}}(\mathcal{H})$  is the *effective domain* of  $F_{R, P}(\alpha, \cdot)$  [75]. Namely,  $F_{R, P}(\alpha, \sigma)$  is finite for all  $\sigma \in \mathcal{S}_{P, \mathcal{W}}(\mathcal{H})$ . Proposition 3 then aims to find the saddle-points within its domain  $(0, 1] \times \mathcal{S}_{P, \mathcal{W}}(\mathcal{H})$ . Otherwise, the saddle-point makes no sense.



(b) Fix  $P \in \mathcal{P}_R(\mathcal{X})$ . Any saddle-point  $(\alpha_{R,P}^*, \sigma_{R,P}^*)$  of  $F_{R,P}(\cdot, \cdot)$  satisfies  $\alpha_{R,P}^* \in (0, 1)$  and

$$\sigma_{R,P}^* \gg W_x, \quad \forall x \in \text{supp}(P).$$

(c) For  $P \in \mathcal{P}_R(\mathcal{X})$ , the saddle-point is unique.

(d) For any  $\underline{R} \in (C_{0,W}, R]$ , both  $\alpha_{r,P}^*$  and  $\sigma_{r,P}^*$  are jointly continuous functions of  $(r, P)$  on  $[\underline{R}, R] \times \mathcal{P}(\mathcal{X})$ .

The following proposition discusses the continuity and differentiability of the error-exponent functions. The proof is shown in Appendix F.

**Proposition 4** (Properties of Error-Exponent Functions). *Consider a classical-quantum channel  $\mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$  with  $C_{0,W} < C_W$ . We have*

(a) *Given every  $P \in \mathcal{P}(\mathcal{X})$ ,  $E_{\text{sp}}^{(2)}(\cdot, P)$  is convex and non-increasing on  $[0, +\infty]$ , and continuous on  $[I_0^{(2)}(P, \mathcal{W}), +\infty]$ . For every  $R > C_{0,W}$ ,  $E_{\text{sp}}^{(2)}(R, \cdot)$  is continuous on  $\mathcal{P}(\mathcal{X})$ . Further,*

$$E_{\text{sp}}^{(2)}(R, P) = \begin{cases} +\infty, & R < I_0^{(2)}(P, \mathcal{W}) \\ 0, & R \geq I_1^{(2)}(P, \mathcal{W}) \end{cases}.$$

(b)  $E_{\text{sp}}(\cdot)$  is convex and non-increasing on  $[0, +\infty]$ , and continuous on  $[C_{0,W}, +\infty]$ . Further,

$$E_{\text{sp}}(R) = \begin{cases} +\infty, & R < C_{0,W} \\ 0, & R \geq C_{1,W} \end{cases}.$$

(c) Consider any  $R \in (C_{0,W}, C_W)$  and  $P \in \mathcal{P}_R(\mathcal{X})$  (see Eq. (3)). The function  $E_{\text{sp}}^{(2)}(\cdot, P)$  is differentiable with

$$s_{R,P}^* := \frac{1 - \alpha_{R,P}^*}{\alpha_{R,P}^*} = - \left. \frac{\partial E_{\text{sp}}^{(2)}(r, P)}{\partial r} \right|_{r=R} \in \mathbb{R}_{>0},$$

where  $\alpha_{R,P}^*$  is the optimizer in Eq. (2.1). Moreover,

$$I_{\alpha_{R,P}^*}(P, \mathcal{W}) > R.$$

Given any  $R \in (C_{0,W}, C_W)$  and  $P \in \mathcal{P}_R(\mathcal{X})$ , we define  $E'_{\text{sp}}(R)$  as the left derivative of  $E_{\text{sp}}(R)$ :

$$E'_{\text{sp}}(R) := \lim_{\delta \downarrow 0} \frac{E_{\text{sp}}(R - \delta) - E_{\text{sp}}(R)}{-\delta}.$$

Since  $E_{\text{sp}}(R)$  is continuous in  $R$  by Proposition 4-(b), the above definition is well-defined. Moreover, Eq. (2.1) and Proposition 4-(c) imply that the absolute value of  $E'_{\text{sp}}(R)$  means the maximum absolute value of the slope of the sphere-packing exponent at  $R$  by

$$|E'_{\text{sp}}(R)| = \max_{P: E_{\text{sp}}^{(2)}(R, P) = E_{\text{sp}}(R)} s_{R,P}^*. \quad (25)$$

Note that the term  $|E'_{\text{sp}}(R)|$  in Eq. (2.1) is well-defined and finite by item (d) in Proposition 3.

**2.2. Quantum Hypothesis Testing and Channel Coding.** Consider a binary hypothesis whose null and alternative hypotheses are  $\rho \in \mathcal{S}(\mathcal{H})$  and  $\sigma \in \mathcal{S}(\mathcal{H})$ , respectively. The *type-I error* and *type-II error* of the hypothesis testing, for an operator  $0 \leq Q \leq \mathbb{1}$ , are defined as:

$$\begin{aligned} \alpha(Q; \rho) &:= \text{Tr}[(\mathbb{1} - Q)\rho], \\ \beta(Q; \sigma) &:= \text{Tr}[Q\sigma]. \end{aligned}$$

There is a trade-off relation between these two errors. Thus we can define the minimum Type-I error when the type-II error is below  $\mu \in (0, 1)$  as

$$\hat{\alpha}_\mu(\rho \| \sigma) := \min_{0 \leq Q \leq \mathbb{1}} \{ \alpha(Q; \rho) : \beta(Q; \sigma) \leq \mu \}.$$

We define an error-exponent function [59, 60, 61] for two sequences of states

$$\begin{aligned} \mathbf{H}_0 : \rho^n &= \rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_n, \\ \mathbf{H}_1 : \sigma^n &= \sigma_1 \otimes \sigma_2 \otimes \cdots \otimes \sigma_n, \end{aligned}$$

by

$$\phi_n(r|\rho^n\|\sigma^n) := \sup_{\alpha \in (0,1]} \left\{ \frac{1-\alpha}{\alpha} \left( \frac{1}{n} D_\alpha(\rho^n\|\sigma^n) - r \right) \right\}, \quad r \geq 0. \quad (26)$$

It is known that [61, Lemma 4]

$$\phi_n(r|\rho^n\|\sigma^n) = +\infty, \quad \forall r \in \left[ 0, -\frac{1}{n} D_0(\rho^n\|\sigma^n) \right).$$

Let  $\mathcal{M}$  be a finite alphabetical set with size  $M = |\mathcal{M}|$ . An  $(n$ -block) *encoder* is a map  $f_n : \mathcal{M} \rightarrow \mathcal{X}^n$  that encodes each message  $m \in \mathcal{M}$  to a codeword  $\mathbf{x}^n(m) := x_1(m)x_2(m)\dots x_n(m) \in \mathcal{X}^n$ . The codeword  $\mathbf{x}^n(m)$  is then mapped to a state

$$W_{\mathbf{x}^n(m)}^{\otimes n} = W_{x_1(m)} \otimes W_{x_2(m)} \otimes \cdots \otimes W_{x_n(m)} \in \mathcal{S}(\mathcal{H}^{\otimes n}).$$

The *decoder* is described by a positive operator-valued measurement (POVM)  $\Pi_n = \{\Pi_{n,1}, \dots, \Pi_{n,M}\}$  on  $\mathcal{H}^{\otimes n}$ , where  $\Pi_{n,i} \geq 0$  and  $\sum_{i=1}^M \Pi_{n,i} = \mathbf{1}$ . The pair  $(f_n, \Pi_n) =: \mathcal{C}_n$  is called an  $(n, R)$ -code with rate  $R = \frac{1}{n} \log |\mathcal{C}_n| = \frac{1}{n} \log M$ . The error probability of sending a message  $m$  with the code  $\mathcal{C}_n$  is  $\epsilon_m(\mathcal{C}_n) := 1 - \text{Tr}(\Pi_{n,m} W_{\mathbf{x}^n(m)}^{\otimes n})$ . We use  $\epsilon_{\max}(\mathcal{C}_n) = \max_{m \in \mathcal{M}} \epsilon_m(\mathcal{C}_n)$  and  $\bar{\epsilon}(\mathcal{C}_n) = \frac{1}{M} \sum_{m \in \mathcal{M}} \epsilon_m(\mathcal{C}_n)$  to denote the *maximal* error probability and the *average* error probability, respectively. Denote by  $\epsilon^*(n, R)$  the smallest average probability of error among all the coding strategies with a blocklength  $n$  and coding rate  $R$ , i.e.

$$\epsilon^*(n, R) := \inf\{\bar{\epsilon}(\mathcal{C}_n) : \mathcal{C}_n \text{ is an } (n, R)\text{-code}\}.$$

The reliability function of the channel  $\mathcal{W}$  and the coding rate  $R$  is defined by<sup>7</sup>

$$E(R) := \limsup_{n \rightarrow +\infty} -\frac{1}{n} \log \epsilon^*(n, R).$$

Winter [16] and Dalai [17] showed that the reliability function of a c-q channel can be upper bounded by  $E(R) \leq \tilde{E}_{\text{sp}}(R)$  and  $E(R) \leq E_{\text{sp}}(R)$ , respectively. Given a sequence  $\mathbf{x}^n \in \mathcal{X}^n$ , we denote by

$$P_{\mathbf{x}^n}(x) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{x = x_i\}, \quad \forall x \in \mathcal{X}$$

the empirical distribution of  $\mathbf{x}^n$ , where  $x_i$  is the  $i$ -th position of  $\mathbf{x}^n$ ; and  $\mathbf{1}\{A\} = 1$  if the event  $A$  is true; otherwise  $\mathbf{1}\{A\} = 0$ . A constant composition code with a composition  $P_{\mathbf{x}^n}$  refers to a codebook whose codewords all have the same distribution  $P_{\mathbf{x}^n}$ .

**2.3. Nussbaum-Szkoła Distributions.** Assume the dimension of the Hilbert space  $\mathcal{H}$  is  $d$ . Given density operators  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  with spectral decompositions

$$\rho = \sum_{i \in [d]} \lambda_i |e_i\rangle \langle e_i|, \quad \text{and} \quad \sigma = \sum_{j \in [d]} \mu_j |f_j\rangle \langle f_j|,$$

we define the *Nussbaum-Szkoła distributions* [62]  $p^{\rho, \sigma}, q^{\rho, \sigma}$  as

$$p^{\rho, \sigma}(i, j) := \lambda_i |\langle e_i | f_j \rangle|^2, \quad q^{\rho, \sigma}(i, j) := \mu_j |\langle e_i | f_j \rangle|^2. \quad (27)$$

The distributions  $p^{\rho, \sigma}, q^{\rho, \sigma}$  have the same mathematical properties as the density operators  $\rho, \sigma$  in some cases, and thus are useful in the sequel. First, one can verify that [62, 46],

$$D_\alpha(\rho\|\sigma) = D_\alpha(p^{\rho, \sigma}\|q^{\rho, \sigma}), \quad \forall \alpha \in [0, 1].$$

<sup>7</sup>Throughout this paper, we skip the dependence of the channel  $\mathcal{W}$  in the reliability function and error-exponent functions.

Second, for product states  $\rho_1 \otimes \rho_2$  and  $\sigma_1 \otimes \sigma_2$ , we have

$$p^{\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2} = p^{\rho_1, \sigma_1} \otimes p^{\rho_2, \sigma_2}, \quad \text{and} \quad q^{\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2} = q^{\rho_1, \sigma_1} \otimes q^{\rho_2, \sigma_2}.$$

Third,  $\rho \ll \sigma$  if and only if  $p^{\rho, \sigma} \ll q^{\rho, \sigma}$ . Moreover, we will use  $\omega$  to represent the pair of indices  $(i, j)$  in Eq. (2.3), and view the distributions  $p^{\rho, \sigma}, q^{\rho, \sigma}$  as diagonal matrices, e.g.  $\text{Tr}[p^{\rho, \sigma}] = \sum_{\omega \in [d] \times [d]} p^{\rho, \sigma}(\omega)$ .

### 3. RELATION BETWEEN THE STRONG AND WEAK SPHERE-PACKING EXPONENTS

This section derives alternative formulations of the strong and weak sphere-packing exponents of Eqs. (1)-(1), and provides a relation between these two exponents. As we will show later, the derived formulations are essentially optimization problems in the primal domain, while the expressions in Eqs. (1) and (1) are corresponding dual representations.

We first consider the following convex optimization problem and then exploit it to establish variational formulations of the sphere-packing exponents. Let  $\rho, \tau \in \mathcal{S}(\mathcal{H})$  be two density operators. Consider the following convex optimization problem:

$$\begin{aligned} \text{(P)} \quad e(r) &:= \inf_{\sigma \in \mathcal{S}(\mathcal{H})} D(\sigma \| \rho), \\ &\text{subject to} \quad D(\sigma \| \tau) \leq r. \end{aligned} \tag{28}$$

The above primal problem is interpreted as finding the optimal operator  $\sigma^*$  that achieves the minimum relative entropy  $e(r)$  to  $\rho$ , within  $r$ -radius to  $\tau$ . The following result shows the dual representation of problem (P) via Lagrangian duality.

**Lemma 5** ([59, Section 3.7], [63], [50, Theorem 3.6]). *The dual problem of (P) is given by*

$$\text{(D)} \quad \sup_{s \geq 0} \left\{ -(1+s) \log Q_{\frac{1}{1+s}}^b(\rho \| \tau) - sr \right\}.$$

*Proof.* By the method of Lagrange multipliers, the primal problem in Eq. (3) can be rewritten as

$$\begin{aligned} &\sup_{s \geq 0} \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \{ D(\sigma \| \rho) + s(D(\sigma \| \tau) - r) \} \\ &= \sup_{s \geq 0} \left\{ (1+s) \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \left\{ \frac{1}{1+s} D(\sigma \| \rho) + \frac{s}{1+s} D(\sigma \| \tau) \right\} - sr \right\} \\ &= \sup_{s \geq 0} \left\{ -(1+s) \log Q_{\frac{1}{1+s}}^b(\rho \| \tau) - sr \right\}, \end{aligned}$$

where the last equality follows from [50, Theorem 3.6].  $\square$

**Theorem 6** (Variational Representations of the Sphere-Packing Exponents). *Let  $\mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$  be a classical-quantum channel. For any  $R > C_{0, \mathcal{W}}$ , we have*

$$\tilde{E}_{\text{sp}}(R, P) = \sup_{0 < \alpha \leq 1} \min_{\sigma \in \mathcal{S}(\mathcal{H})} \left\{ \frac{1-\alpha}{\alpha} \left( D_{\alpha}^b(\mathcal{W} \| \sigma | P) - R \right) \right\}, \quad \text{and} \tag{29}$$

$$E_{\text{sp}}(R, P) \leq \sup_{0 < \alpha \leq 1} \min_{\sigma \in \mathcal{S}(\mathcal{H})} \left\{ \frac{1-\alpha}{\alpha} \left( D_{\alpha}(\mathcal{W} \| \sigma | P) - R \right) \right\}, \tag{30}$$

where  $\tilde{E}_{\text{sp}}(R, P)$  and  $E_{\text{sp}}(R, P)$  are defined in Eqs. (2.1) and (2.1), respectively.

Moreover, equality in Eq. (6) is attained when maximizing over all prior distributions, i.e.,

$$E_{\text{sp}}(R) = \max_{P \in \mathcal{P}(\mathcal{X})} E_{\text{sp}}(R, P) = \max_{P \in \mathcal{P}(\mathcal{X})} \sup_{0 < \alpha \leq 1} \min_{\sigma \in \mathcal{S}(\mathcal{H})} \left\{ \frac{1-\alpha}{\alpha} \left( D_{\alpha}(\mathcal{W} \| \sigma | P) - R \right) \right\}. \tag{31}$$

*Proof.* We start with the proof of Eq. (6). Recall Eq. (2.1):

$$I(P, \mathcal{V}) = \min_{\sigma \in \mathcal{S}(\mathcal{H})} D(\mathcal{V} \| \sigma | P).$$

We find

$$\begin{aligned}
\tilde{E}_{\text{sp}}(R, P) &= \min_{\mathcal{V}: \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})} \{D(\mathcal{V} \parallel \mathcal{W} | P) : I(P, \mathcal{V}) \leq R\} \\
&= \min_{\mathcal{V}: \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})} \left\{ D(\mathcal{V} \parallel \mathcal{W} | P) : \min_{\sigma \in \mathcal{S}(\mathcal{H})} D(\mathcal{V} \parallel \sigma | P) \leq R \right\} \\
&= \sup_{s \geq 0} \min_{\mathcal{V}: \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})} \left\{ D(\mathcal{V} \parallel \mathcal{W} | P) + s \left( \min_{\sigma \in \mathcal{S}(\mathcal{H})} D(\mathcal{V} \parallel \sigma | P) - R \right) \right\} \\
&= \sup_{s \geq 0} \min_{\sigma \in \mathcal{S}(\mathcal{H})} \min_{\mathcal{V}: \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})} \left\{ -sR + \sum_{x \in \mathcal{X}} P(x) [D(V_x \parallel W_x) + s \cdot D(V_x \parallel \sigma)] \right\} \\
&= \sup_{s \geq 0} \min_{\sigma \in \mathcal{S}(\mathcal{H})} \left\{ \sum_{x \in \mathcal{X}} P(x) \min_{V_x \in \mathcal{S}(\mathcal{H})} [D(V_x \parallel W_x) + s \cdot D(V_x \parallel \sigma) - sR] \right\} \\
&= \sup_{s \geq 0} \min_{\sigma \in \mathcal{S}(\mathcal{H})} \left\{ \sum_{x \in \mathcal{X}} P(x) \left[ -(1+s) \log Q_{\frac{1}{1+s}}^b(W_x \parallel \sigma) - sR \right] \right\}.
\end{aligned}$$

In the third equality we introduced the constraint into the objective function via the Lagrange multiplier  $s \geq 0$ ; The fifth equality follows from the linearity of the convex combination; and the last line is due to Lemma 5. Further, we use the substitution  $\alpha = 1/(1+s)$  and recall the definition of the log-Euclidean  $\alpha$ -Rényi divergence given in Eq. (2.1) to obtain:

$$\tilde{E}_{\text{sp}}(R, P) = \sup_{0 < \alpha \leq 1} \min_{\sigma \in \mathcal{S}(\mathcal{H})} \left\{ \frac{1-\alpha}{\alpha} \left( D_{\alpha}^b(\mathcal{W} \parallel \sigma | P) - R \right) \right\}.$$

Hence, we prove the first claim in Eq. (6).

Next, we will prove Eq. (6). From Jensen's inequality and the concavity of the logarithm, the right-hand side of Eq. (6) implies that

$$\begin{aligned}
&\sup_{0 < \alpha \leq 1} \min_{\sigma \in \mathcal{S}(\mathcal{H})} \left\{ \frac{1-\alpha}{\alpha} \left( \sum_{x \in \mathcal{X}} P(x) D_{\alpha}(W_x \parallel \sigma) - R \right) \right\} \\
&= \sup_{0 < \alpha \leq 1} \min_{\sigma \in \mathcal{S}(\mathcal{H})} \left\{ -\frac{1}{\alpha} \sum_{x \in \mathcal{X}} P(x) \log \text{Tr} [W_x^{\alpha} \sigma^{1-\alpha}] - \frac{1-\alpha}{\alpha} R \right\} \\
&\geq \sup_{0 < \alpha \leq 1} \min_{\sigma \in \mathcal{S}(\mathcal{H})} \left\{ -\frac{1}{\alpha} \log \text{Tr} \left[ \sum_{x \in \mathcal{X}} P(x) [W_x^{\alpha} \sigma^{1-\alpha}] \right] - \frac{1-\alpha}{\alpha} R \right\} \\
&= E_{\text{sp}}(R, P),
\end{aligned}$$

where the last equality follows from Eq. (2.1).

Finally, we invoke the following identity proved by Mosonyi and Ogawa [50, Proposition 4.2]:

$$\max_{P \in \mathcal{P}(\mathcal{X})} \min_{\sigma \in \mathcal{S}(\mathcal{H})} D_{\alpha}(\mathcal{W} \parallel \sigma | P) = \max_{P \in \mathcal{P}(\mathcal{X})} \min_{\sigma \in \mathcal{S}(\mathcal{H})} D_{\alpha}(P \circ \mathcal{W} \parallel P \otimes \sigma).$$

Note that the above relation also holds for  $D_{\alpha}^b$ . Then, combining Eqs. (2.1), (2.1) and (2.1), Eq. (6) holds as follows:

$$\begin{aligned}
E_{\text{sp}}(R) &= \max_{P \in \mathcal{P}(\mathcal{X})} E_{\text{sp}}(R, P) \\
&= \max_{P \in \mathcal{P}(\mathcal{X})} E_{\text{sp}}^{(1)}(R, P) \\
&= \max_{P \in \mathcal{P}(\mathcal{X})} \min_{\sigma \in \mathcal{S}(\mathcal{H})} D_{\alpha}(P \circ \mathcal{W} \parallel P \otimes \sigma) \\
&= \max_{P \in \mathcal{P}(\mathcal{X})} \min_{\sigma \in \mathcal{S}(\mathcal{H})} D_{\alpha}(\mathcal{W} \parallel \sigma | P).
\end{aligned}$$

□

The following corollary is a simple consequence of the variational representations of the sphere-packing exponents in Theorem 6 and Eq. (2.1).

**Corollary 7.** *For any classical-quantum channel  $\mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$ ,  $R > C_{0,\mathcal{W}}$ , and  $P \in \mathcal{P}(\mathcal{X})$ , it holds that*

$$E_{\text{sp}}(R, P) \leq \tilde{E}_{\text{sp}}(R, P).$$

#### 4. FINITE BLOCKLENGTH SPHERE-PACKING BOUND

The main result in the section is a finite blocklength strong sphere-packing bound for c-q channels with a polynomial prefactor (Theorem 8), improving upon a subexponential prefactor obtained in [17]. To establish this result, the key is to employ a hypothesis testing reduction and a sharp concentration inequality [26, 35]. Our proof consists of three major steps: (i) reducing the channel coding problem to binary hypothesis testing (Proposition 10); (ii) bounding its type-I error from below (Propositions 12 and 13); (iii) employing Theorem 6 to relate the derived bound to the strong sphere-packing exponent. We discuss Propositions 10, 12, and 13 in Section 4.1, and prove the main result Theorem 8 in Section 4.2.

**Theorem 8** (Refined Strong Sphere-Packing Bound of Constant Composition Codes). *Consider a classical-quantum channel  $\mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$  and  $R \in (C_{0,\mathcal{W}}, C_{\mathcal{W}})$ . For every  $\gamma > 0$ , there exist an  $N_0 \in \mathbb{N}$  and a constant  $A > 0$  such that for all constant composition codes  $\mathcal{C}_n$  of length  $n \geq N_0$  with message size  $|\mathcal{C}_n| \geq \exp\{nR\}$ , we have*

$$\bar{\epsilon}(\mathcal{C}_n) \geq \frac{A}{n^{\frac{1}{2}(1+|E'_{\text{sp}}(R)|+\gamma)}} \exp\{-nE_{\text{sp}}(R)\}.$$

The following corollary generalizes the refined sphere-packing bound for constant composition codes to arbitrary codes via a standard argument [7, p. 95].

**Corollary 9** (Refined Strong Sphere-Packing Bound for General Codes). *Consider a classical-quantum channel  $\mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$  and  $R \in (C_{0,\mathcal{W}}, C_{\mathcal{W}})$ . There exist some  $t > 1/2$  and  $N_0 \in \mathbb{N}$  such that for all codes of length  $n \geq N_0$ , we have*

$$\epsilon^*(n, R) \geq n^{-t} \exp\{-nE_{\text{sp}}(R)\}.$$

Proofs for Theorem 8 and Corollary 9 are provided in Section 4.2.

Theorem 8 yields

$$\log \frac{1}{\bar{\epsilon}(\mathcal{C}_n)} \leq nE_{\text{sp}}(R) + \frac{1}{2} (1 + |E'_{\text{sp}}(R)|) \log n + o(\log n),$$

where the term  $\frac{1}{2} (1 + |E'_{\text{sp}}(R)|)$  can be viewed as a second-order term (see the discussions in [65, Section 4.4]). On the other hand, for the case of classical *non-singular* channels<sup>8</sup>, it was shown that [24, Theorem 3.6], for all constant composition codes  $\mathcal{C}_n$  and rate  $R \in (R_{\text{crit}}, C_{\mathcal{W}})$ ,

$$\log \frac{1}{\bar{\epsilon}(\mathcal{C}_n)} \geq nE_{\text{r}}(R) + \frac{1}{2} (1 + |E'_{\text{r}}(R)|) \log n + \Omega(1),$$

where  $E_{\text{r}}(R)$  is the *random coding exponent*, and  $R_{\text{crit}}$  is the critical rate such that  $E_{\text{r}}(R) = E_{\text{sp}}(R)$  for all  $R \geq R_{\text{crit}}$  [6, p. 160], [15]. Hence our result, Theorem 8, matches the achievability up to the logarithmic order. We note that whether the third order  $o(\log n)$  in Eq. (4) can be improved to  $O(1)$  is still unknown even for the classical case.

<sup>8</sup>For classical *singular* channels, one has  $\log \frac{1}{\bar{\epsilon}(\mathcal{C}_n)} \geq nE_{\text{r}}(R) + \frac{1}{2} \log n + \Omega(1)$  [24]. Further, it was conjectured that [66] that  $\log \frac{1}{\bar{\epsilon}(\mathcal{C}_n)} \leq nE_{\text{sp}}(R) + \frac{1}{2} \log n + o(\log n)$ , for all asymmetric classical singular channels and constant composition codes. However, such a result remains open.

**4.1. Converse Bounds for Quantum Hypothesis Testing.** This section contains the hypothesis testing reduction method (Proposition 10) and two converse bounds for the optimal type-I error of a composition (Propositions 12 and 13). First of all, Proposition 10 allows us to lower bound the optimal error probability of a code to the optimal type-I error in quantum hypothesis testing. To further lower bound the type-I error, we prove a converse Hoeffding bound for a binary hypothesis testing of product states in Theorem 11. Choosing the hypotheses in Theorem 11 as a channel output state of a codeword against an  $n$ -copy state and employing a uniform continuity (Proposition 18 in Appendix C), we establish a finite blocklength Chebyshev-type converse bound (Proposition 12), and a sharp converse bound (Proposition 13), respectively. In Section 4.2, we will combine these two bounds to prove the desired sphere-packing bound (Theorem 8).

We remark that the converse bounds (Propositions 12 and 13) for fixed composition of codewords hold for all sufficiently large  $n$ . Moreover, we call such bounds *finite blocklength bounds* because the blocklength  $n$  only depends on the channel  $\mathcal{W}$ , coding rate  $R$ , and  $|\mathcal{X}|$ . The independence of the composition of codes allows us to establish finite blocklength sphere-packing bounds. The key ingredient here is to prove a non-trivial uniform continuity property in the large deviation regime (Appendix C).

In the following, we show Proposition 10 and its proof. We note that Proposition 10 below is similar to the meta-converse in Ref. [67]. However, the idea dates back to Fano [4], Shannon-Gallager-Berlekamp [7], and Blahut [10, 64].

**Proposition 10.** *For any classical-quantum channel  $\mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$  and any code  $\mathcal{C}$  with message size  $M$ , it follows that*

$$\epsilon_{\max}(\mathcal{C}) \geq \max_{\sigma \in \mathcal{S}(\mathcal{H})} \min_{x \in \mathcal{C}} \hat{\alpha}_{\frac{1}{M}}(W_x \| \sigma).$$

*Proof.* Let  $x(m)$  be the codeword encoding the message  $m \in \{1, \dots, M\}$ . Define a binary hypothesis testing problem:

$$\begin{aligned} H_0 &: W_x, \\ H_1 &: \sigma, \end{aligned}$$

where  $\sigma \in \mathcal{S}(\mathcal{H})$  can be viewed as a dummy channel output. Since  $\sum_{m=1}^M \beta(\Pi_m; \sigma) = 1$  for any POVM  $\Pi = \{\Pi_1, \dots, \Pi_M\}$ , and  $\beta(\Pi_m; \sigma) \geq 0$  for every  $m \in \mathcal{M}$ , there must exist a message  $m \in \mathcal{M}$  for any code  $\mathcal{C}$  such that  $\beta(\Pi_m; \sigma) \leq \frac{1}{M}$ . Fix such  $x := x(m)$ . Then

$$\epsilon_{\max}(\mathcal{C}) \geq \epsilon_m(\mathcal{C}) = \alpha(\Pi_m; W_x) \geq \hat{\alpha}_{\frac{1}{M}}(W_x \| \sigma) \geq \min_{x \in \mathcal{C}} \hat{\alpha}_{\frac{1}{M}}(W_x \| \sigma). \quad (32)$$

Since the above inequality (4.1) holds for every  $\sigma \in \mathcal{S}(\mathcal{H})$ , it follows that

$$\epsilon_{\max}(\mathcal{C}) \geq \max_{\sigma \in \mathcal{S}(\mathcal{H})} \min_{x \in \mathcal{C}} \hat{\alpha}_{\frac{1}{M}}(W_x \| \sigma).$$

□

Before showing the following converse Hoeffding bound, Theorem 11, we first introduce some notation. Let

$$\begin{aligned} H_0 &: \rho^n = \rho_1 \otimes \dots \otimes \rho_n; \\ H_1 &: \sigma^n = \sigma_1 \otimes \dots \otimes \sigma_n, \end{aligned} \quad (33)$$

where  $\rho_x, \sigma_x \in \mathcal{S}(\mathcal{H})$  for  $x \in [n]$ . Further, denote by  $(p_i, q_i)$  be the Nussbaum-Szkoła distribution of  $(\rho_i, \sigma_i)$  given in Section 2.3. For  $\alpha \in [0, 1]$ , define

$$\begin{aligned} B_\alpha(\rho^n \| \sigma^n) &:= \frac{1}{n} \sum_{x \in [n]} \mathbb{E}_{v_{x,\alpha}} \left[ \log \frac{p_x}{q_x} \right]; \\ V_\alpha(\rho^n \| \sigma^n) &:= \frac{1}{n} \sum_{x \in [n]} \mathbb{E}_{v_{x,\alpha}} \left[ \left| \log \frac{p_x}{q_x} - \mathbb{E}_{v_{x,\alpha}} \left[ \log \frac{p_x}{q_x} \right] \right|^2 \right]; \\ T_\alpha(\rho^n \| \sigma^n) &:= \frac{1}{n} \sum_{x \in [n]} \mathbb{E}_{v_{x,\alpha}} \left[ \left| \log \frac{p_x}{q_x} - \mathbb{E}_{v_{x,\alpha}} \left[ \log \frac{p_x}{q_x} \right] \right|^3 \right], \end{aligned}$$

where  $(p_x, q_x)$  is the Nussbaum-Szkoła distribution of  $(\rho_x, \sigma_x)$  for  $x \in [n]$ , and the *tilted distribution* is

$$v_{x,\alpha}(i, j) := \frac{p_x^\alpha(i, j) q_x^{1-\alpha}(i, j)}{\sum_{i,j} p_x^\alpha(i, j) q_x^{1-\alpha}(i, j)}, \quad \alpha \in [0, 1].$$

With the above notation, we have the following converse bound.

**Theorem 11** (Sharp Converse Hoeffding Bounds for Quantum Hypothesis Testing). *Consider a binary hypothesis testing:  $H_0 : \rho^n = \bigotimes_{i=1}^n \rho_i$  and  $H_1 : \sigma^n = \bigotimes_{i=1}^n \sigma_i$  given in Eq. (4.1) with  $\rho^n \ll \sigma^n$ . Let  $r \in \mathbb{R}$  be such that there exists an  $\alpha^* \in (0, 1]$  such that*

$$\phi_n(r | \rho^n \| \sigma^n) = \frac{1 - \alpha^*}{\alpha^*} \left( \frac{1}{n} D_{\alpha^*}(\rho^n \| \sigma^n) - r \right).$$

Then, we have the following: (i) for any test  $Q_n$ , either

$$\alpha(Q_n; \rho^n) \geq \frac{1}{8} \exp \left\{ -n \phi_n(r | \rho^n \| \sigma^n) - \alpha^* \sqrt{2n V_{\alpha^*}(\rho^n \| \sigma^n)} \right\}, \quad (34)$$

or

$$\beta(Q_n; \sigma^n) \geq \frac{1}{8} \exp \left\{ -nr - (1 - \alpha^*) \sqrt{2n V_{\alpha^*}(\rho^n \| \sigma^n)} \right\} \quad (35)$$

holds; (ii) if  $\alpha^* \in (0, 1)$ , then for any test  $Q_n$ , either

$$\alpha(Q_n; \rho^n) \geq e^{-n \phi_n(r | \rho^n \| \sigma^n)} \frac{e^{-K_n(\alpha^*)}}{2\sqrt{2n\pi V_{\alpha^*}(\rho^n \| \sigma^n)}} \left( 1 - \frac{1 + (1 + K_n(\alpha^*)^2)}{2\sqrt{n V_{\alpha^*}(\rho^n \| \sigma^n)}} \right), \quad (36)$$

or

$$\beta(Q_n; \sigma^n) \geq e^{-nr} \frac{e^{-K_n(\alpha^*)}}{2\sqrt{2n\pi V_{\alpha^*}(\rho^n \| \sigma^n)}} \left( 1 - \frac{1 + (1 + K_n(\alpha^*)^2)}{2\sqrt{n V_{\alpha^*}(\rho^n \| \sigma^n)}} \right) \quad (37)$$

holds. Here,  $K_n(\alpha) := \frac{15\sqrt{2\pi} T_\alpha(\rho^n \| \sigma^n)}{V_\alpha(\rho^n \| \sigma^n)} \in \mathbb{R}_{>0}$  and  $V_{\alpha^*}(\rho^n \| \sigma^n) \in \mathbb{R}_{>0}$ .

*Proof.* The first claim directly follows from Dailai's result in [17, Theorem 4]. Before proceeding, we need to introduce some notation. Let  $\tilde{p}^n := \bigotimes_{i=1}^n \tilde{p}_i$  and  $\tilde{q}^n := \bigotimes_{i=1}^n \tilde{q}_i$ , where  $(\tilde{p}_i, \tilde{q}_i)$  are the Nussbaum-Szkoła distributions [62] of  $(\rho_i, \sigma_i)$  for  $i \in [n]$ . Since  $D_\alpha(\rho_i \| \sigma_i) = D_\alpha(\tilde{p}_i \| \tilde{q}_i)$ , for all  $\alpha \in (0, 1)$ , we shorthand

$$\phi_n(r) := \phi_n(r | \rho^n \| \sigma^n) = \phi_n(r | \tilde{p}^n \| \tilde{q}^n) \in \mathbb{R}_{>0}.$$

Applying Nagaoka's argument [63], for any  $0 \leq Q_n \leq 1$  with  $\delta = \exp\{nr - n\phi_n(r)\}$ , we have

$$\alpha(Q_n; \rho^n) + \delta \beta(Q_n; \sigma^n) \geq \frac{1}{2} \left( \alpha(\tilde{\mathcal{U}}; \tilde{p}^n) + e^{nr - n\phi_n(r)} \beta(\tilde{\mathcal{U}}; \tilde{q}^n) \right), \quad (38)$$

where  $\alpha(\tilde{\mathcal{U}}; \tilde{p}^n) := \sum_{\omega \notin \tilde{\mathcal{U}}} \tilde{p}^n(\omega)$ ,  $\beta(\tilde{\mathcal{U}}; \tilde{q}^n) := \sum_{\omega \in \tilde{\mathcal{U}}} \tilde{q}^n(\omega)$ , and

$$\tilde{\mathcal{U}} := \left\{ \omega : \tilde{p}^n(\omega) e^{n\phi_n(r)} > \tilde{q}^n(\omega) e^{nr} \right\}.$$



Now, we further define the non-normalized distributions  $p^n := \bigotimes_{i=1}^n p_i$  and  $q^n := \bigotimes_{i=1}^n q_i$ , where  $p_i := \tilde{p}_i q_i^0$ ,  $q_i := \tilde{q}_i p_i^0$ , for every  $i \in [n]$ . Namely, we restrict  $(p^n, q^n)$  to be in the joint support of  $\tilde{p}^n$  and  $\tilde{q}^n$ . Letting

$$\mathcal{U} := \left\{ \omega : p^n(\omega) e^{n\phi_n(r)} > q^n(\omega) e^{nr} \right\},$$

it is not hard to see that

$$\begin{aligned} \alpha(\mathcal{U}; p^n) &= \alpha(\tilde{\mathcal{U}}; \tilde{p}^n); \\ \beta(\mathcal{U}; q^n) &= \beta(\tilde{\mathcal{U}}; \tilde{q}^n) \\ \phi_n(r) &= \phi_n(r | p^n \| q^n). \end{aligned}$$

Hence, we focus on the pair  $(p^n, q^n)$  and the decision region  $\mathcal{U}$  onwards.

Let

$$\begin{aligned} \Lambda_{0,n}(\alpha) &:= \frac{1}{n} \sum_{i \in [n]} \Lambda_{0,i}(\alpha), \quad \Lambda_{0,i}(\alpha) := \log \mathbb{E}_{p_i} \left[ e^{(1-\alpha) \log \frac{q_i}{p_i}} \right]; \\ \Lambda_{1,n}(\alpha) &:= \frac{1}{n} \sum_{i \in [n]} \Lambda_{1,i}(\alpha), \quad \Lambda_{1,i}(\alpha) := \log \mathbb{E}_{q_i} \left[ e^{(1-\alpha) \log \frac{p_i}{q_i}} \right]. \end{aligned}$$

Since  $p^n$  and  $q^n$  have the same support, both  $\Lambda_{0,n}(\alpha)$  and  $\Lambda_{1,n}(\alpha)$  are smooth functions in  $\alpha \in \mathbb{R}$ . One can calculate their derivatives as follows:

$$\begin{aligned} \Lambda'_{0,n}(\alpha) &= \frac{1}{n} \sum_{i \in [n]} \mathbb{E}_{v_{i,\alpha}} \left[ \log \frac{p_i}{q_i} \right]; \quad \Lambda'_{1,n}(\alpha) = \frac{1}{n} \sum_{i \in [n]} \mathbb{E}_{v_{i,1-\alpha}} \left[ \log \frac{q_i}{p_i} \right] \\ \Lambda''_{0,n}(\alpha) &= \frac{1}{n} \sum_{i \in [n]} \text{Var}_{v_{i,\alpha}} \left[ \log \frac{p_i}{q_i} \right]; \quad \Lambda''_{1,n}(\alpha) = \frac{1}{n} \sum_{i \in [n]} \text{Var}_{v_{i,1-\alpha}} \left[ \log \frac{q_i}{p_i} \right], \\ T_{0,n}(\alpha) &:= \frac{1}{n} \sum_{i \in [n]} \mathbb{E}_{v_{i,\alpha}} \left[ \left| \log \frac{p_i}{q_i} - \Lambda'_{0,n}(\alpha) \right|^3 \right]; \\ T_{1,n}(\alpha) &:= \frac{1}{n} \sum_{i \in [n]} \mathbb{E}_{v_{i,1-\alpha}} \left[ \left| \log \frac{q_i}{p_i} - \Lambda'_{1,n}(\alpha) \right|^3 \right], \end{aligned}$$

where we denote the tilted distribution by

$$v_{i,\alpha}(\omega) := \frac{p_i^\alpha(\omega) q_i^{1-\alpha}(\omega)}{\sum_{\bar{\omega}} p_i^\alpha(\bar{\omega}) q_i^{1-\alpha}(\bar{\omega})}, \quad \alpha \in [0, 1].$$

Further, it is not hard to verify that for all  $\alpha \in [0, 1]$ .

$$\begin{aligned} \Lambda_{0,n}(\alpha) &= \Lambda_{1,n}(1-\alpha); \quad \Lambda'_{0,n}(\alpha) = -\Lambda'_{1,n}(1-\alpha); \\ \Lambda''_{0,n}(\alpha) &= \Lambda''_{1,n}(1-\alpha); \quad T_{0,n}(\alpha) = T_{1,n}(1-\alpha). \end{aligned} \tag{39}$$

Next, we define the *Lengendre-Fenchel transform*:

$$\Lambda_{j,n}^*(z) := \sup_{\alpha \in \mathbb{R}} \{ (1-\alpha)z - \Lambda_{j,n}(\alpha) \}, \quad j \in \{0, 1\}.$$

The quantities  $\Lambda_{j,n}^*(z)$  would appear in the lower bounds of  $\alpha(Q^n; \rho^n)$  and  $\beta(Q^n; \sigma^n)$  as shown later.

Now, we are ready to show the first claim. Ref. [17, Theorem 4] states that for any test  $Q^n$ , either

$$\alpha(Q^n; \rho^n) \geq \frac{1}{8} \exp \left\{ -n \left[ \alpha^* \Lambda'_{0,n}(\alpha^*) - \Lambda_{0,n}(\alpha^*) \right] - \alpha^* \sqrt{2n V_{\alpha^*}(\rho^n \| \sigma^n)} \right\},$$

or

$$\beta(Q^n; \sigma^n) \geq \frac{1}{8} \exp \left\{ -n \left[ -(1-\alpha^*) \Lambda'_{0,n}(\alpha^*) + \Lambda_{0,n}(\alpha^*) \right] - (1-\alpha^*) \sqrt{2n V_{\alpha^*}(\rho^n \| \sigma^n)} \right\},$$

holds. By Eqs. (4.1), and (A), (A) in Appendix A, we have

$$\begin{aligned}\phi_n(r) &= \alpha^* \Lambda'_{0,n}(\alpha^*) - \Lambda_{0,n}(\alpha^*); \\ r &= -(1 - \alpha^*) \Lambda'_{0,n}(\alpha^*) + \Lambda_{0,n}(\alpha^*),\end{aligned}$$

which proves the first claim.

To show the second claim, we will employ Bahadur-Ranga Rao's concentration inequality, Theorem 17, in Appendix B, to further lower bound  $\alpha(\mathcal{U}; p^n)$  and  $\beta(\mathcal{U}; q^n)$ . Letting  $Z_i = \log \frac{q_i}{p_i}$  with probability measure  $\mu_i = p_i$ , and  $z = r - \phi_n(r)$  in Theorem 17, the Bahadur-Ranga Rao's inequality gives

$$\begin{aligned}\alpha(\mathcal{U}; p^n) &:= \sum_{\omega \notin \mathcal{U}} p^n(\omega) \\ &= \Pr \left\{ \frac{1}{n} \sum_{i=1}^n Z_i \geq \phi_n(r) - r \right\} \\ &\geq \exp \left\{ -n \Lambda_{0,n}^*(\phi_n(r) - r) \right\} \frac{e^{-K_n(\alpha^*)}}{\sqrt{2\pi \Lambda''_{0,n}(\alpha^*)}} \left( 1 - \frac{1 + (1 + K_n(\alpha^*)^2)}{2\sqrt{\Lambda''_{0,n}(\alpha^*)}} \right),\end{aligned}$$

where

$$K_n(\alpha) := 15\sqrt{2\pi} \frac{T_\alpha(\rho^n \|\sigma^n)}{V_\alpha(\rho^n \|\sigma^n)} = 15\sqrt{2} \frac{T_{0,n}(\alpha)}{\Lambda''_{0,n}(\alpha)}.$$

Moreover, Lemma 16 in Appendix A relates the Legendre-Fenchel transform  $\Lambda_{j,P_{\mathbf{x}^n}}^*(z)$  to the desired error exponent function  $\phi_n(r)$ :

$$\begin{aligned}\Lambda''_{0,n}(\alpha^*) &> 0; \\ \Lambda_{0,n}^*(\phi_n(r) - r) &= \phi_n(r); \\ \Lambda_{1,n}^*(r - \phi_n(r)) &= r.\end{aligned}$$

Hence, we have

$$\alpha(\mathcal{U}; p^n) \geq \exp \left\{ -n \phi_n(r) \right\} \frac{e^{-K_n(\alpha^*)}}{\sqrt{2\pi \Lambda''_{0,n}(\alpha^*)}} \left( 1 - \frac{1 + (1 + K_n(\alpha^*)^2)}{2\sqrt{\Lambda''_{0,n}(\alpha^*)}} \right). \quad (40)$$

Similarly, applying Theorem 17 with  $Z_i = \log \frac{p_i}{q_i}$ ,  $\mu_i = q_i$ , and  $z = \phi_n(r) - r$  yields

$$\begin{aligned}\beta(\mathcal{U}; q^n) &:= \sum_{\omega \in \mathcal{U}} q^n(\omega) \\ &= \Pr \left\{ \frac{1}{n} \sum_{i=1}^n Z_i \geq \phi_n(r) - r \right\} \\ &\geq \exp \left\{ -n \Lambda_{1,n}^*(r - \phi_n(r)) \right\} \frac{e^{-K_n(1-\alpha^*)}}{\sqrt{2\pi \Lambda''_{1,n}(1-\alpha^*)}} \left( 1 - \frac{1 + (1 + K_n(1-\alpha^*)^2)}{2\sqrt{\Lambda''_{1,n}(1-\alpha^*)}} \right) \\ &= \exp \left\{ -nr \right\} \frac{e^{-K_n(1-\alpha^*)}}{\sqrt{2\pi \Lambda''_{1,n}(1-\alpha^*)}} \left( 1 - \frac{1 + (1 + K_n(1-\alpha^*)^2)}{2\sqrt{\Lambda''_{1,n}(1-\alpha^*)}} \right) \\ &= \exp \left\{ -nr \right\} \frac{e^{-K_n(\alpha^*)}}{\sqrt{2\pi \Lambda''_{0,n}(\alpha^*)}} \left( 1 - \frac{1 + (1 + K_n(\alpha^*)^2)}{2\sqrt{\Lambda''_{0,n}(\alpha^*)}} \right),\end{aligned} \quad (41)$$

where the last equality follows from Eq. (4.1). Hence, by Eqs. (4.1), (4.1), and (4.1), we conclude our claim.  $\square$

In the following Proposition, we will prove a finite blocklength bound of a composition via Eqs. (11) and (11) in the above Theorem 11. The difficulty of deriving a finite blocklength result is that one needs to obtain some universal coefficients independent of all possible compositions. Our core technique here is a uniform continuity property, Proposition 18, which will be presented in Appendix C.

The following result is essentially a Chebyshev-type bound with prefactor  $\exp\{O(\sqrt{n})\}$ . We will employ it to lower bound the error of “bad sequences” that yield a inferior error exponent in Section 4.2.

**Proposition 12** (Chebyshev-Type Bound for a Fixed Composition). *Let  $\mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$  be a classical-quantum channel. Fix  $R \in (C_{0,\mathcal{W}}, C_{\mathcal{W}})$ . Consider a sequence  $\mathbf{x}^n \in \mathcal{X}^n$ . Then, for every  $c > 0$ , there exist a state  $\sigma^* \in \mathcal{S}(\mathcal{H})$ , an integer  $N_0 \in \mathbb{N}$ , independent of the sequences  $\mathbf{x}^n$  and  $\sigma$ , such that for all  $n \geq N_0$  we have*

$$\hat{\alpha}_{c \exp\{-nR\}}(W_{\mathbf{x}^n}^{\otimes n} \| (\sigma^*)^{\otimes n}) \geq \exp \left\{ -A\sqrt{n} - nE_{\text{sp}}^{(2)}(R, P_{\mathbf{x}^n}) \right\},$$

where  $A \in \mathbb{R}_{>0}$  is a finite positive constant depending on  $R$  and  $\mathcal{W}$ .

*Proof.* Fix an arbitrary  $\underline{R} \in (C_{0,\mathcal{W}}, R)$ . Let  $\gamma_n := \frac{a\sqrt{n}}{2n} + \frac{\log 8 - \log c}{n}$  and  $R_n := R - \gamma_n$  for some  $a \in \mathbb{R}$ . The choice of  $a$  and the rate back-off term  $\gamma_n$  will become evident later. Let  $N_1 \in \mathbb{N}$  such that  $R_n \in [\underline{R}, R]$  for all  $n \geq N_1$ . Subsequently, we choose such  $n \geq N_1$  onwards.

We choose the optimal output state as

$$\sigma^* = \arg \min_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{0 < \alpha \leq 1} \frac{1 - \alpha}{\alpha} (D_{\alpha}(\mathcal{W} \| \sigma | P_{\mathbf{x}^n}) - R_n).$$

Let  $p^n := \bigotimes_{i=1}^n p_{x_i}$  and  $q^n := \bigotimes_{i=1}^n q_{x_i}$ , where  $(p_{x_i}, q_{x_i})$  are Nussbaum-Szkoła distributions [62] of  $(W_{x_i}, \sigma_{R,P}^*)$  for every  $i \in [n]$ . Since  $D_{\alpha}(W_{x_i} \| \sigma^*) = D_{\alpha}(p_{x_i} \| q_{x_i})$ , for  $\alpha \in (0, 1]$ , again we shorthand for all  $R_n \in [\underline{R}, R]$ ,

$$\phi_n(R_n) := \phi_n(R_n | W_{\mathbf{x}^n}^{\otimes n} \| (\sigma^*)^{\otimes n}) = \phi_n(R_n | p^n \| q^n) = E_{\text{sp}}^{(2)}(R_n, P_{\mathbf{x}^n}), \quad (42)$$

where the last equality in Eq. (4.1) follows from the saddle-point property, item (a) in Proposition 3. Moreover, item (b) in Proposition 3 implies that the state  $\sigma^*$  dominates all the states:  $\sigma^* \gg W_x$ , for all  $x \in \text{supp}(P_{\mathbf{x}^n})$ . Hence, we have  $p^n \ll q^n$ . This guarantees that  $V_{\alpha}(W_x \| \sigma^*)$  is finite for all  $\alpha \in [0, 1]$  and all  $x \in \text{supp}(P_{\mathbf{x}^n})$ .

Theorem 11 implies that for any test  $Q^n$  either

$$\alpha(Q^n; W_{\mathbf{x}^n}^{\otimes n}) \geq \frac{1}{8} \exp \left\{ -n\phi_n(R_n) - \alpha_{R_n, P_{\mathbf{x}^n}}^* \sqrt{2nV_{\alpha_{R_n, P_{\mathbf{x}^n}}^*}^*(P_{\mathbf{x}^n}, \mathcal{W})} \right\}, \quad (43)$$

or

$$\beta(Q^n; (\sigma^*)^{\otimes n}) \geq \frac{1}{8} \exp \left\{ -nR_n - (1 - \alpha_{R_n, P_{\mathbf{x}^n}}^*) \sqrt{2nV_{\alpha_{R_n, P_{\mathbf{x}^n}}^*}^*(P_{\mathbf{x}^n}, \mathcal{W})} \right\},$$

where  $\alpha_{r, P_{\mathbf{x}^n}}^* \in (0, 1)$  satisfies, for all  $r \in [\underline{R}, R]$ ,

$$\phi_n(r) = \frac{1 - \alpha_{r, P_{\mathbf{x}^n}}^*}{\alpha_{r, P_{\mathbf{x}^n}}^*} \left( I_{\alpha_{r, P_{\mathbf{x}^n}}^*}^{(2)}(P_{\mathbf{x}^n}, \mathcal{W}) - r \right).$$

and we define

$$V_{\alpha}(P, \mathcal{W}) := \sum_{x \in \mathcal{X}} P(x) \mathbb{E}_{v_{x, \alpha}} \left[ \left| \log \frac{p_x}{q_x} - \mathbb{E}_{v_{x, \alpha}} \left[ \log \frac{p_x}{q_x} \right] \right|^2 \right],$$

$$v_{x, \alpha}(\omega) := \frac{p_x^{\alpha}(\omega) q_x^{1-\alpha}(\omega)}{\sum_{\bar{\omega}} p_x^{\alpha}(\bar{\omega}) q_x^{1-\alpha}(\bar{\omega})}.$$

Now, we will introduce a constant independent of  $\mathbf{x}^n$  to obtain a finite blocklength lower bound for  $\hat{\alpha}_{c \exp\{-nR\}}(W_{\mathbf{x}^n}^{\otimes n} \| (\sigma^*)^{\otimes n})$ . Let

$$V_{\max}(R) := \max_{(r, P) \in [\underline{R}, R] \times \mathcal{P}(\mathcal{X})} V_{\alpha_{r, P}^*}(P, \mathcal{W}). \quad (44)$$

Recall that  $V_\alpha(P_{\mathbf{x}^n}, \mathcal{W})$  is finite for all  $\alpha \in [0, 1]$ . Proposition 18 in Appendix C implies that  $V_{\alpha_{r,P}^*}(P, \mathcal{W})$  is joint continuous on  $[\underline{R}, R] \times \mathcal{P}(\mathcal{X})$ . Further, since  $\mathcal{P}(\mathcal{X})$  is compact, the quantity in Eq. (4.1) is well-defined and finite. Therefore,

$$\begin{aligned}\beta(Q^n; (\sigma^*)^{\otimes n}) &\geq \frac{1}{8} \exp \left\{ -nR_n - (1 - \alpha_{R_n, P_{\mathbf{x}^n}}^*) \sqrt{2nV_{\alpha_{R_n, P_{\mathbf{x}^n}}^*}(P_{\mathbf{x}^n}, \mathcal{W})} \right\} \\ &\geq \frac{1}{8} \exp \left\{ -nR_n - \sqrt{2nV_{\max}(R)} \right\} \\ &= c \exp \{-nR\},\end{aligned}\tag{45}$$

where we choose  $a := \sqrt{2V_{\max}(R)}$  in the rate back-off term  $\gamma_n := \frac{a\sqrt{n}}{2n} + \frac{\log 8 - \log c}{n}$ .

Next, Eqs. (4.1) and (4.1) yield

$$\begin{aligned}\hat{\alpha}_{c \exp\{-nR\}}(W_{\mathbf{x}^n}^{\otimes n} \| (\sigma^*)^{\otimes n}) &\geq \frac{1}{8} \exp \left\{ -n\phi_n(R_n) - \alpha_{R_n, P_{\mathbf{x}^n}}^* \sqrt{2nV_{\alpha_{R_n, P_{\mathbf{x}^n}}^*}(P_{\mathbf{x}^n}, \mathcal{W})} \right\} \\ &\geq \frac{1}{8} \exp \left\{ -\sqrt{2nV_{\max}(R)} - nE_{\text{sp}}^{(2)}(R - \gamma_n, P_{\mathbf{x}^n}) \right\}.\end{aligned}\tag{46}$$

Further, the convexity and the monotone decreases of  $r \mapsto E_{\text{sp}}^{(2)}(r, P)$  given in Proposition 4-(a) shows that

$$\begin{aligned}E_{\text{sp}}^{(2)}(R - \gamma_n, P_{\mathbf{x}^n}) &\leq E_{\text{sp}}^{(2)}(R, P_{\mathbf{x}^n}) - \gamma_n \left. \frac{\partial E_{\text{sp}}^{(2)}(r, P_{\mathbf{x}^n})}{\partial r} \right|_{r=R}, \\ &\leq E_{\text{sp}}^{(2)}(R, P_{\mathbf{x}^n}) - \gamma_n \left. \frac{\partial E_{\text{sp}}^{(2)}(r, P_{\mathbf{x}^n})}{\partial r} \right|_{r=\underline{R}}.\end{aligned}\tag{47}$$

Next, we denote

$$\Upsilon := \max_{P \in \mathcal{P}(\mathcal{X})} \left| \left. \frac{\partial E_{\text{sp}}^{(2)}(r, P)}{\partial r} \right|_{r=\underline{R}} \right|.\tag{48}$$

Observe that  $\Upsilon \in \mathbb{R}_{\geq 0}$  due to  $\underline{R} > C_{0,\mathcal{W}}$  and item (d) of Proposition 3. Then, Eqs. (4.1), (4.1), and (4.1) lead to

$$\hat{\alpha}_{c \exp\{-nR\}}(W_{\mathbf{x}^n}^{\otimes n} \| (\sigma^*)^{\otimes n}) \geq \exp \left\{ -\log 8 - \sqrt{2nV_{\max}(R)} - \gamma_n \Upsilon - nE_{\text{sp}}^{(2)}(R, P_{\mathbf{x}^n}) \right\}.$$

Since  $\gamma_n = O(1/\sqrt{n})$ , for any  $A > \sqrt{2V_{\max}(R)}$ , there exists a sufficiently large  $N_2 \in \mathbb{N}$  such that for all  $n \geq N_2$ .

$$\log 8 + \sqrt{2nV_{\max}(R)} + \gamma_n \Upsilon \leq A\sqrt{n}.$$

By letting  $N_0 := \max\{N_1, N_2\}$  completes the proof.  $\square$

The following Proposition 13 is a sharp converse bound with polynomial prefactors obtained from Eqs. (11) and (11) in Theorem 11, which in turn were proved by employing Bahadur-Ranga Rao's inequality (see Appendix B). Similar to Proposition 12 presented before, we will employ the uniform continuity, Proposition 18, given in Appendix C to prove Proposition 13. In Section 4.2, we will exploit this result to bound the error of "good sequences" with a polynomial prefactor.

**Proposition 13** (Sharp Converse Bound for a Fixed Composition). *Let  $\mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$  be a classical-quantum channel Fix  $R \in (C_{0,\mathcal{W}}, C_{\mathcal{W}})$ . Consider a sequence  $\mathbf{x}^n \in \mathcal{X}^n$  satisfying*

$$E_{\text{sp}}^{(2)}(R, P_{\mathbf{x}^n}) \in [\nu, +\infty)\tag{49}$$

*for some constant  $\nu > 0$ . Then, for every  $c > 0$ , there exist a state  $\sigma^* \in \mathcal{S}(\mathcal{H})$ , an integer  $N_0 \in \mathbb{N}$ , independent of the sequences  $\mathbf{x}^n$  and  $\sigma$ , such that for all  $n \geq N_0$  we have*

$$\hat{\alpha}_{c \exp\{-nR\}}(W_{\mathbf{x}^n}^{\otimes n} \| (\sigma^*)^{\otimes n}) \geq \frac{A}{n^{\frac{1}{2}(1+s_{R, P_{\mathbf{x}^n}}^*)}} \exp \left\{ -nE_{\text{sp}}^{(2)}(R, P_{\mathbf{x}^n}) \right\},$$

where  $s_{R,P}^* := - \left. \frac{\partial E_{\text{sp}}^{(2)}(r,P)}{\partial r} \right|_{r=R}$ , and  $A \in \mathbb{R}_{>0}$  is a finite positive constant depending on  $R, \nu$  and  $\mathcal{W}$ .

*Proof.* Fix an arbitrary  $\underline{R} \in (C_{0,\mathcal{W}}, R)$ . Let  $\gamma_n := \frac{\log n}{2n} + \frac{x}{n}$  and  $R_n := R - \gamma_n$  for some  $x \in \mathbb{R}$ . The choice of  $x$  and the rate back-off term  $\gamma_n$  will become evident later. Let  $N_1 \in \mathbb{N}$  such that  $R_n \in [\underline{R}, R]$  for all  $n \geq N_1$ . Subsequently, we choose such  $n \geq N_1$  onwards.

We choose the optimal output state as

$$\sigma^* = \arg \min_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{0 < \alpha \leq 1} \frac{1 - \alpha}{\alpha} (D_\alpha(\mathcal{W} \| \sigma | P_{\mathbf{x}^n}) - R_n).$$

as in the proof of Proposition 12. Let  $p^n := \bigotimes_{i=1}^n p_{x_i}$  and  $q^n := \bigotimes_{i=1}^n q_{x_i}$ , where  $(p_{x_i}, q_{x_i})$  are Nussbaum-Szkoła distributions [62] of  $(W_{x_i}, \sigma^*)$  for every  $i \in [n]$ . Since  $D_\alpha(W_{x_i} \| \sigma^*) = D_\alpha(p_{x_i} \| q_{x_i})$ , for  $\alpha \in (0, 1]$ , again we shorthand for all  $R_n \in [\underline{R}, R]$ ,

$$\phi_n(R_n) := \phi_n(R_n | W_{\mathbf{x}^n}^{\otimes n} \| (\sigma^*)^{\otimes n}) = \phi_n(R_n | p^n \| q^n) = E_{\text{sp}}^{(2)}(R_n, P_{\mathbf{x}^n}), \quad (50)$$

where the last equality in Eq. (4.1) follows from the saddle-point property, item (a) in Proposition 3. Moreover, item (b) in Proposition 3 implies that the state  $\sigma^*$  dominates all the states:  $\sigma^* \gg W_x$ , for all  $x \in \text{supp}(P_{\mathbf{x}^n})$ . Hence, we have  $p^n \ll q^n$ . Without loss of generality, we set zero all elements of  $q_{x_i}$  that do not lie in the support of  $p_{x_i}$ , i.e.  $q_{x_i}(\omega) = 0$ ,  $\omega \notin \text{supp}(p_{x_i})$ ,  $i \in [n]$ , because those elements do not contribute in  $\phi_n(R_n)$ .

Next, we define

$$\begin{aligned} V_\alpha(P, \mathcal{W}) &:= \sum_{x \in \mathcal{X}} P(x) \mathbb{E}_{v_{x,\alpha}} \left[ \left| \log \frac{p_x}{q_x} - \mathbb{E}_{v_{x,\alpha}} \left[ \log \frac{p_x}{q_x} \right] \right|^2 \right]; \\ T_\alpha(P, \mathcal{W}) &:= \sum_{x \in \mathcal{X}} P(x) \mathbb{E}_{v_{x,\alpha}} \left[ \left| \log \frac{p_x}{q_x} - \mathbb{E}_{v_{x,\alpha}} \left[ \log \frac{p_x}{q_x} \right] \right|^3 \right], \\ v_{x,\alpha}(\omega) &:= \frac{p_x^\alpha(\omega) q_x^{1-\alpha}(\omega)}{\sum_{\bar{\omega}} p_x^\alpha(\bar{\omega}) q_x^{1-\alpha}(\bar{\omega})}, \end{aligned}$$

Applying Theorem 11, we have for any test  $Q^n$ ,

$$\alpha(Q^n; W_{\mathbf{x}^n}^{\otimes n}) \geq e^{-n\phi_n(R_n)} \frac{e^{-K_n(\alpha_{R_n, P_{\mathbf{x}^n}}^*)}}{2\sqrt{2n\pi V_{\alpha_{R_n, P_{\mathbf{x}^n}}^*}(P_{\mathbf{x}^n}, \mathcal{W})}} \left( 1 - \frac{1 + (1 + K_n(\alpha_{R_n, P_{\mathbf{x}^n}}^*)^2)}{2\sqrt{n V_{\alpha_{R_n, P_{\mathbf{x}^n}}^*}(P_{\mathbf{x}^n}, \mathcal{W})}} \right) \quad (51)$$

$$\beta(Q^n; (\sigma^*)^{\otimes n}) \geq e^{-nR_n} \frac{e^{-K_n(\alpha_{R_n, P_{\mathbf{x}^n}}^*)}}{2\sqrt{2n\pi V_{\alpha_{R_n, P_{\mathbf{x}^n}}^*}(P_{\mathbf{x}^n}, \mathcal{W})}} \left( 1 - \frac{1 + (1 + K_n(\alpha_{R_n, P_{\mathbf{x}^n}}^*)^2)}{2\sqrt{n V_{\alpha_{R_n, P_{\mathbf{x}^n}}^*}(P_{\mathbf{x}^n}, \mathcal{W})}} \right), \quad (52)$$

where  $K_n(\alpha) := 15\sqrt{2} \frac{T_\alpha(P_{\mathbf{x}^n}, \mathcal{W})}{V_\alpha(P_{\mathbf{x}^n}, \mathcal{W})}$ , and  $\alpha_{r, P_{\mathbf{x}^n}}^* \in (0, 1)$  satisfies, for all  $r \in [\underline{R}, R]$ ,

$$\phi_n(r) = \frac{1 - \alpha_{r, P_{\mathbf{x}^n}}^*}{\alpha_{r, P_{\mathbf{x}^n}}^*} \left( I_{\alpha_{r, P_{\mathbf{x}^n}}^*}^{(2)}(P_{\mathbf{x}^n}, \mathcal{W}) - r \right).$$

In the following, we will remove the dependency of  $R_n$  and  $P_{\mathbf{x}^n}$  in  $V_{(\cdot)}(\cdot)$  and  $K_n(\cdot)$ . Define the following quantities:

$$V_{\max}(R, \nu) := \max_{(r,P) \in [\underline{R}, R] \times \mathcal{P}_{R,\nu}(\mathcal{X})} V_{\alpha_{r,P}^*}(P, \mathcal{W}); \quad (53)$$

$$V_{\min}(R, \nu) := \min_{(r,P) \in [\underline{R}, R] \times \mathcal{P}_{R,\nu}(\mathcal{X})} V_{\alpha_{r,P}^*}(P, \mathcal{W}); \quad (54)$$

$$K_{\max}(R, \nu) := 15\sqrt{2\pi} \max_{(r,P) \in [\underline{R}, R] \times \mathcal{P}_{R,\nu}(\mathcal{X})} \frac{T_{\alpha_{r,P}^*}(P, \mathcal{W})}{V_{\alpha_{r,P}^*}(P, \mathcal{W})}, \quad (55)$$

where

$$\mathcal{P}_{R,\nu}(\mathcal{X}) := \left\{ P \in \mathcal{P}(\mathcal{X}) : \nu \leq E_{\text{sp}}^{(2)}(R, P_{\mathbf{x}^n}) \leq E_{\text{sp}}(R) < +\infty \right\} \quad (56)$$

is a compact set owing to the continuity of  $r \mapsto E_{\text{sp}}^{(2)}(r, P)$  given in Proposition 4. Also, Proposition 18 in Appendix C shows that the objective functions in Eqs. (4.1), (4.1), and (4.1) are continuous functions on  $\mathcal{P}(\mathcal{X})$ , which guarantees the maximization and minimization in the above definitions are well-defined and finite. Further, the quantity  $V_{\min}(R, \nu)$  is bounded away from zero because of the positivity given in Theorem 11.

Now, we are ready to derive the lower bounds for  $\hat{\alpha}_{c \exp\{-nR\}}(W_{\mathbf{x}^n}^{\otimes n} \| (\sigma^*)^{\otimes n})$ . Let  $N_2 \in \mathbb{N}$  be sufficiently large such that for all  $n \geq N_2$ ,

$$\sqrt{n} \geq \frac{1 + (1 + K_{\max}(R, \nu))^2}{\sqrt{V_{\min}(R, \nu)}}.$$

Then, Eqs. (4.1) and (4.1) give

$$\begin{aligned} \alpha(Q^n; W_{\mathbf{x}^n}^{\otimes n}) &\geq \frac{A(R, \nu)}{\sqrt{n}} \exp\{-n\phi_n(R_n)\}; \\ \beta(Q^n; (\sigma^*)^{\otimes n}) &\geq \frac{A(R, \nu)}{\sqrt{n}} \exp\{-nR_n\}, \end{aligned}$$

where

$$A(R, \nu) := \frac{e^{-K_{\max}(R, \nu)}}{4\sqrt{2\pi V_{\max}(R, \nu)}}.$$

Choosing  $x = -\log A(R, \nu) + \log c$  in the rate back-off term  $\gamma_n = \frac{\log n}{2n} + \frac{x}{n}$ , we have

$$\beta(Q^n; (\sigma^*)^{\otimes n}) \geq c \exp\{-nR\}. \quad (57)$$

Combining Eqs. (4.1) and (4.1) then yields

$$\hat{\alpha}_{c \exp\{-nR\}}(W_{\mathbf{x}^n}^{\otimes n} \| (\sigma^*)^{\otimes n}) \geq \frac{A(R, \nu)}{\sqrt{n}} \exp\{-n\phi_n(R_n)\} = \frac{A(R, \nu)}{\sqrt{n}} \exp\left\{-nE_{\text{sp}}^{(2)}(R - \gamma_n, P_{\mathbf{x}^n})\right\}. \quad (58)$$

It remains to remove the rate back-off term  $\gamma_n$  in Eq. (4.1). By Taylor's theorem, one has

$$E_{\text{sp}}^{(2)}(R - \gamma_n, P_{\mathbf{x}^n}) = E_{\text{sp}}^{(2)}(R, P_{\mathbf{x}^n}) - \gamma_n \left. \frac{\partial E_{\text{sp}}^{(2)}(r, P_{\mathbf{x}^n})}{\partial r} \right|_{r=R} + \frac{\gamma_n^2}{2} \left. \frac{\partial^2 E_{\text{sp}}^{(2)}(r, P_{\mathbf{x}^n})}{\partial r^2} \right|_{r=\bar{R}}, \quad (59)$$

for some  $\bar{R} \in (\underline{R}, R)$ . Recalling item (d) in Lemma 16, one can show that

$$\begin{aligned} - \left. \frac{\partial E_{\text{sp}}^{(2)}(r, P_{\mathbf{x}^n})}{\partial r} \right|_{r=R} &= s_{R, P_{\mathbf{x}^n}}^* = \frac{1 - \alpha_{R, P_{\mathbf{x}^n}}^*}{\alpha_{R, P_{\mathbf{x}^n}}^*} \in \mathbb{R}_{>0}, \\ \left. \frac{\partial^2 E_{\text{sp}}^{(2)}(r, P_{\mathbf{x}^n})}{\partial r^2} \right|_{r=\bar{R}} &= \frac{(1 + \bar{s})^3}{V_{\alpha_{\bar{R}, P_{\mathbf{x}^n}}^*}(P_{\mathbf{x}^n}, \mathcal{W})} \leq \frac{(1 + s_0)^3}{V_{\min}(R, \nu)} =: \Upsilon \in \mathbb{R}_{>0}, \end{aligned} \quad (60)$$

where

$$\bar{s} := - \left. \frac{\partial E_{\text{sp}}^{(2)}(r, P_{\mathbf{x}^n})}{\partial r} \right|_{r=\bar{R}} \leq - \left. \frac{\partial E_{\text{sp}}^{(2)}(r, P_{\mathbf{x}^n})}{\partial r} \right|_{r=\underline{R}} =: s_0 \in \mathbb{R}_{>0}$$

by the monotone decreases of  $r \mapsto E_{\text{sp}}^{(2)}(r, P)$ . Then, Eqs. (4.1), (4.1) and (4.1) lead to

$$\begin{aligned} \hat{\alpha}_{c \exp\{-nR\}}(W_{\mathbf{x}^n}^{\otimes n} \| (\sigma^*)^{\otimes n}) &\geq \frac{A(R, \nu)}{\sqrt{n}} \exp\left\{-nE_{\text{sp}}^{(2)}(R, P_{\mathbf{x}^n}) - n\left[\gamma_n \left(s_{R, P_{\mathbf{x}^n}}^* + \frac{\gamma_n}{2}\Upsilon\right)\right]\right\} \\ &= \frac{A(R, \nu)}{n^{\frac{1}{2}(1+s_{R, P_{\mathbf{x}^n}}^*)}} \exp\left\{-nE_{\text{sp}}^{(2)}(R, P_{\mathbf{x}^n}) - \ell_n\right\}, \end{aligned} \quad (61)$$

where we denote by

$$\ell_n := - \left( s_{R, P_{\mathbf{x}^n}}^* + \frac{\gamma_n}{2} \Upsilon \right) \log A(R, \nu) + \frac{\gamma_n \Upsilon}{4} \log n.$$

Since  $s_{R, P_{\mathbf{x}^n}}^* \in \mathbb{R}_{>0}$  and  $\gamma_n \log n = o(1)$ , we choose a constant  $L \in \mathbb{R}_{>0}$  and  $N_3 \in \mathbb{N}$  such that

$$\ell_n \leq L, \quad \forall N \geq N_3. \quad (62)$$

Hence, Eqs. (4.1) and (4.1) lead to

$$\hat{\alpha}_{c \exp\{-nR\}} (W_{\mathbf{x}^n}^{\otimes n} \| (\sigma^*)^{\otimes n}) = \frac{A(R, \nu) \exp\{-L\}}{n^{\frac{1}{2}(1+s_{R, P_{\mathbf{x}^n}}^*)}} \exp \left\{ -n E_{\text{sp}}^{(2)}(R, P_{\mathbf{x}^n}) \right\}.$$

By letting  $N_0 := \max\{N_1, N_2, N_3\}$  and  $A' := A(R, \nu) \exp\{-L\}$ , we conclude the proof.  $\square$

**4.2. Proofs of Theorem 8 and Corollary 9.** We are ready to prove our main result—the refined strong sphere-packing bound in Theorem 8 for constant composition codes and Corollary 9 for general codes.

*Proof of Theorem 8.* Fix any rate  $C_{0, \mathcal{W}} < R < C_{\mathcal{W}}$ . First note that by Ref. [15, Proposition 10], we find

$$E_{\text{sp}}(R) \in \mathbb{R}_{>0}.$$

By Proposition 10 and the standard expurgation method (see e.g. [7, p. 96], [10, Theorem 20], [64, p. 395]), it holds for every constant composition code  $\mathcal{C}_n$  with a common composition  $P_{\mathbf{x}^n}$  that

$$\begin{aligned} \bar{\epsilon}(\mathcal{C}_n) &\geq \frac{1}{2} \epsilon_{\max}(\mathcal{C}'_n) \geq \max_{\sigma \in \mathcal{S}(\mathcal{H})} \frac{1}{2} \hat{\alpha}_{1/|\mathcal{C}'_n|} (W_{\mathbf{x}^n}^{\otimes n} \| \sigma^{\otimes n}) \\ &\geq \max_{\sigma \in \mathcal{S}(\mathcal{H})} \frac{1}{2} \hat{\alpha}_{2 \exp\{-nR\}} (W_{\mathbf{x}^n}^{\otimes n} \| \sigma^{\otimes n}) \end{aligned} \quad (63)$$

$$\geq \frac{1}{2} \hat{\alpha}_{2 \exp\{-nR\}} (W_{\mathbf{x}^n}^{\otimes n} \| (\sigma^*)^{\otimes n}), \quad (64)$$

where  $\mathcal{C}'_n$  is an expurgated code with message size  $|\mathcal{C}'_n| = \lceil |\mathcal{C}_n|/2 \rceil \geq \frac{1}{2} \exp\{nR\}$ . Inequality (4.2) holds because the map  $\mu \mapsto \hat{\alpha}_\mu$  is monotone decreasing. In the last line (4.2) we denote by  $\sigma^*$  a channel output state that depends on the coding rate  $R$  and the composition  $P_{\mathbf{x}^n}$ , and  $\sigma^*$  will be chosen later.

In the following, we deal with sequences of inputs that will yield different lower bounds. Fix an arbitrary  $\delta \in (0, E_{\text{sp}}(R))$ . Let  $\nu := E_{\text{sp}}(R) - \delta > 0$ , and recall the definition in Eq. (4.1):

$$\mathcal{P}_{R, \nu}(\mathcal{X}) := \left\{ P_{\mathbf{x}^n} \in \mathcal{P}(\mathcal{X}) : \nu \leq E_{\text{sp}}^{(2)}(R, P_{\mathbf{x}^n}) \leq E_{\text{sp}}(R) < +\infty \right\}.$$

The set  $\mathcal{P}_{R, \nu}(\mathcal{X})$  ensures that the error exponents of the input sequences  $\mathbf{x}^n$  with composition  $P_{\mathbf{x}^n} \in \mathcal{P}_{R, \nu}(\mathcal{X})$  are close to the sphere-packing exponent  $E_{\text{sp}}(R)$ .

For sequences  $\mathbf{x}^n$  with  $P_{\mathbf{x}^n} \notin \mathcal{P}_{R, \nu}(\mathcal{X})$ , we infer that

$$E_{\text{sp}}(R) - E_{\text{sp}}^{(2)}(R, P_{\mathbf{x}^n}) = \delta > 0.$$

We then apply the Chebyshev-type bound, Proposition 12, with  $c = 2$  to obtain, for some  $\kappa \in \mathbb{R}_{>0}$  and  $\forall P_{\mathbf{x}^n} \notin \mathcal{P}_{R, \nu}(\mathcal{X})$ ,

$$\hat{\alpha}_{2 \exp\{-nR\}} (W_{\mathbf{x}^n}^{\otimes n} \| (\sigma^*)^{\otimes n}) \geq \exp \left\{ -\kappa \sqrt{n} - n E_{\text{sp}}^{(2)}(R, P_{\mathbf{x}^n}) \right\}, \quad (65)$$

$$\geq \exp \left\{ -\kappa \sqrt{n} - n [E_{\text{sp}}(R) - \delta] \right\}, \quad (66)$$

for all sufficiently large  $n$ , say  $n \geq N_1 \in \mathbb{N}$ . The equality in Eq. (4.2) follows from the saddle-point property, item (a) in Proposition 3, and the constants  $\kappa_1, \kappa_2$  are positive and finite constants.

Next, we consider sequences  $\mathbf{x}^n$  with  $P_{\mathbf{x}^n} \in \mathcal{P}_{R, \nu}(\mathcal{X})$ . Since such sequences satisfy Eq. (13), we apply the sharp lower bound, Proposition 13, with  $c = 2$  to obtain,  $\forall P_{\mathbf{x}^n} \in \mathcal{P}_{R, \nu}(\mathcal{X})$ ,

$$\hat{\alpha}_{2 \exp\{-nR\}} (W_{\mathbf{x}^n}^{\otimes n} \| (\sigma^*)^{\otimes n}) \geq \frac{2A}{n^{\frac{1}{2}(1+s_{R, P_{\mathbf{x}^n}}^*)}} \exp \left\{ -n E_{\text{sp}}^{(2)}(R, P_{\mathbf{x}^n}) \right\}, \quad (67)$$



for all sufficiently large  $n$ , say  $n \geq N_2 \in \mathbb{N}$ , and some  $A \in \mathbb{R}_{>0}$ . In the following, we will relate the term  $s_{R, P_{\mathbf{x}^n}}^*$  in Eq. (4.2) to  $|E'_{\text{sp}}(R)|$ . The idea follows similar from [35, Eqs. (111)–(114)]. Let

$$\begin{aligned}\mathcal{P}_R^*(\mathcal{X}) &:= \left\{ P \in \mathcal{P}(\mathcal{X}) : E_{\text{sp}}^{(2)}(R, P) = E_{\text{sp}}(R) \right\}, \\ \mathcal{P}_\theta(\mathcal{X}) &:= \left\{ P \in \mathcal{P}_{R, \nu}(\mathcal{X}) : \min_{Q \in \mathcal{P}_R^*(\mathcal{X})} \|P - Q\|_1 \geq \theta \right\}.\end{aligned}$$

Since  $s_{R, (\cdot)}^*$  is uniformly continuous on the compact set  $P \in \mathcal{P}_{R, \nu}(\mathcal{X})$  (see item (d) of Proposition 3), one has

$$\forall \gamma \in \mathbb{R}_{>0}, \exists f(\gamma) \in \mathbb{R}_{>0}, \text{ such that } \forall P, Q \in \mathcal{P}_{R, \nu}(\mathcal{X}), \|P - Q\|_1 < f(\gamma) \Rightarrow |s_{R, P}^* - s_{R, Q}^*| < \gamma. \quad (68)$$

By choosing  $\gamma \in \mathbb{R}_{>0}$  that satisfies Eq. (4.2), it follows that

$$s_{R, P_{\mathbf{x}^n}}^* \leq |E'_{\text{sp}}(R)| + \gamma, \quad \forall P_{\mathbf{x}^n} \in \mathcal{P}_{R, \nu}(\mathcal{X}) \setminus \mathcal{P}_{f(\gamma)}(\mathcal{X}). \quad (69)$$

Hence, Eqs. (4.2) and (4.2) further lead to,  $\forall P_{\mathbf{x}^n} \in \mathcal{P}_{R, \nu}(\mathcal{X}) \setminus \mathcal{P}_{f(\gamma)}(\mathcal{X})$ ,

$$\hat{\alpha}_{2 \exp\{-nR\}} (W_{\mathbf{x}^n}^{\otimes n} \|(\sigma^*)^{\otimes n}) \geq \frac{2A}{n^{\frac{1}{2}}(1+|E'_{\text{sp}}(R)|+\gamma)} \exp\{-nE_{\text{sp}}(R)\}. \quad (70)$$

For the case  $P_{\mathbf{x}^n} \in \mathcal{P}_{R, \nu}(\mathcal{X}) \cap \mathcal{P}_{f(\gamma)}(\mathcal{X})$ , we have

$$E_{\text{sp}}(R) - \max_{P \in \mathcal{P}_{f(\gamma)}(\mathcal{X})} E_{\text{sp}}^{(2)}(R, P_{\mathbf{x}^n}) =: \delta' > 0. \quad (71)$$

Then, Eqs. (4.2) and (4.2) give,  $\forall P_{\mathbf{x}^n} \in \mathcal{P}_{R, \nu}(\mathcal{X}) \cap \mathcal{P}_{f(\gamma)}(\mathcal{X})$ ,

$$\hat{\alpha}_{2 \exp\{-nR\}} (W_{\mathbf{x}^n}^{\otimes n} \|(\sigma^*)^{\otimes n}) \geq \frac{2A}{n^{\frac{1}{2}}(1+s_{R, P_{\mathbf{x}^n}}^*)} \exp\{-n[E_{\text{sp}}(R) - \delta']\}. \quad (72)$$

Finally, by comparing the bounds in Eqs. (4.2), (4.2) and (4.2), the first-order leading term in the right-hand side of Eq. (4.2) decays faster than that of Eqs. (4.2) and (4.2). Thus, for sufficiently large  $n$ , say  $n \geq N_3 \in \mathbb{N}$ , we combine the bounds to obtain, for all compositions  $P_{\mathbf{x}^n} \in \mathcal{P}(\mathcal{X})$ ,

$$\hat{\alpha}_{2 \exp\{-nR\}} (W_{\mathbf{x}^n}^{\otimes n} \|(\sigma^*)^{\otimes n}) \geq \frac{2A}{n^{\frac{1}{2}}(1+|E'_{\text{sp}}(R)|+\gamma)} \exp\{-nE_{\text{sp}}(R)\}. \quad (73)$$

By combining Eqs. (4.2), (4.2), we conclude our result: for any  $\gamma > 0$  and every  $n$ -blocklength constant composition code  $\mathcal{C}_n$ ,

$$\bar{\epsilon}(\mathcal{C}_n) \geq \frac{A}{n^{\frac{1}{2}}(1+|E'_{\text{sp}}(R)|+\gamma)} \exp\{-nE_{\text{sp}}(R)\},$$

for all sufficiently large  $n \geq N_0 := \max\{N_1, N_2, N_3\}$ .  $\square$

*Proof of Corollary 9.* For an  $n$ -blocklength code, there are at most  $\binom{n+|\mathcal{X}|-1}{|\mathcal{X}|-1} < n^{|\mathcal{X}|}$  different compositions. Hence, for any code with  $M = \exp\{nR\}$  codewords, there exists some codewords  $M'$  of the same composition such that  $M' \geq M/n^{|\mathcal{X}|}$ . Denote by  $\mathcal{C}'_n$  such constant composition codes with composition  $P_{\mathbf{x}^n}$ .

Fix an arbitrary  $\underline{R} \in (C_{0, \mathcal{W}}, R)$ , and choose  $N_1$  be an integer such that  $R - \frac{|\mathcal{X}|}{n} \log n \geq \underline{R}$  for all  $n \geq N_1$ . Consider such  $n \geq N_1$  onwards. By following the similar steps in Theorem 8, we obtain

$$\epsilon^*(n, R) \geq \bar{\epsilon}(\mathcal{C}'_n) \geq \frac{A}{n^{\frac{1}{2}}(1+s_{R, P_{\mathbf{x}^n}}^*)} \exp\left\{-nE_{\text{sp}}^{(2)}\left(R - \frac{|\mathcal{X}|}{n} \log n, P_{\mathbf{x}^n}\right)\right\}, \quad (74)$$

for all sufficiently large  $n$ , say  $n \geq N_2 \in \mathbb{N}$ , and some  $s_{R, P_{\mathbf{x}^n}}^* \in \mathbb{R}_{>0}$ . Let

$$\Upsilon := \max_{P \in \mathcal{P}(\mathcal{X}) : E_{\text{sp}}^{(2)}(\bar{R}, P) = E_{\text{sp}}(\bar{R})} \left| \frac{\partial E_{\text{sp}}^{(2)}(r, P)}{\partial r} \right|_{r=\underline{R}}.$$

Then, item (a) in Proposition 4 implies that

$$\begin{aligned} E_{\text{sp}}^{(2)}\left(R - \frac{|\mathcal{X}|}{n} \log n, P_{\mathbf{x}^n}\right) &\leq E_{\text{sp}}^{(2)}(R, P_{\mathbf{x}^n}) + \Upsilon \cdot \frac{|\mathcal{X}|}{n} \log n \\ &\leq E_{\text{sp}}(R) + \Upsilon \cdot \frac{|\mathcal{X}|}{n} \log n, \quad \forall n \geq N_2 \end{aligned} \quad (75)$$

Combining Eqs. (4.2) and (4.2) gives

$$\epsilon^*(n, R) \geq \frac{A}{n^{\frac{1}{2}(1+s_{R, P_{\mathbf{x}^n}}^*) + \Upsilon|\mathcal{X}|}} \exp\{-nE_{\text{sp}}(R)\}, \quad \forall n \geq \max\{N_1, N_2\}.$$

By choosing  $t \in \mathbb{R}_{>0}$  such that  $n^{-t} \leq A n^{-\frac{1}{2}(1+s_{R, P_{\mathbf{x}^n}}^*) - \Upsilon|\mathcal{X}|}$ , and letting  $N_0 := \max\{N_1, N_2\}$ , we conclude our claim.  $\square$

## 5. SYMMETRIC CLASSICAL-QUANTUM CHANNELS

In this section, we consider a symmetric c-q channels. By using the symmetric property of the channels, we show that the uniform distribution, denoted by  $U_{\mathcal{X}}$ , achieves the maximum of  $E_{\text{sp}}^{(1)}(R, \cdot)$  and  $E_{\text{sp}}^{(2)}(R, \cdot)$ . Then, by choosing the optimal output state  $\sigma_R^* = \sigma_{R, U_{\mathcal{X}}}^*$ , every input sequence in the codebook is a good codeword and attains the sphere-packing exponent  $E_{\text{sp}}(R)$ . Hence, we can remove the assumption of constant composition codes and apply Theorem 8 in Section 4 to obtain the optimal prefactor for the sphere-packing bound (Theorem 14).

A c-q channel  $\mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$  is *symmetric* if it satisfies

$$W_x := V^{x-1} W_1 (V^\dagger)^{x-1}, \quad \forall x \in \mathcal{X}, \quad (76)$$

where  $W_1 \in \mathcal{S}(\mathcal{H})$  is an arbitrary density operator, and  $V$  is a unitary operator on  $\mathcal{S}(\mathcal{H})$  that satisfies  $V^\dagger V = V V^\dagger = V^{|\mathcal{X}|} = \mathbb{1}_{\mathcal{H}}$ .

**Theorem 14** (Exact Sphere-packing Bound for Symmetric Classical-Quantum Channels). *For any rate  $R \in (C_{0, \mathcal{W}}, C_{\mathcal{W}})$ , there exist  $A > 0$  and  $N_0 \in \mathbb{N}$  such that for all codes  $\mathcal{C}_n$  of length  $n \geq N_0$  with message size  $|\mathcal{C}_n| \geq \exp\{nR\}$ , we have*

$$\epsilon_{\max}(\mathcal{C}_n) \geq \frac{A}{n^{\frac{1}{2}(1+|E'_{\text{sp}}(R)|)}} \exp\{-nE_{\text{sp}}(R)\}.$$

*Proof.* The proof consists of the following steps. First, we show that the distribution  $U_{\mathcal{X}}$  satisfies  $E_{\text{sp}}^{(1)}(R, U_{\mathcal{X}}) = E_{\text{sp}}^{(2)}(R, U_{\mathcal{X}}) = E_{\text{sp}}(R)$ . Second, we show that  $E_{\text{sp}}^{(2)}(R, P) = E_{\text{sp}}(R)$  for all  $P \in \mathcal{P}(\mathcal{X})$ , which means that any codeword attains the sphere-packing exponent. Finally, we follow Theorem 8 to complete the proof.

Fix any  $R \in (C_{0, \mathcal{W}}, C_{\mathcal{W}})$ . From the definition of the symmetric channels in Eq. (5), it is not hard to verify that  $U_{\mathcal{X}} \mathcal{W}^\alpha = V U_{\mathcal{X}} \mathcal{W}^\alpha V^\dagger$  for all  $\alpha \in (0, 1]$ , where we denote by  $P \mathcal{W}^\alpha := \sum_{x \in \mathcal{X}} P(x) W_x^\alpha$  for all  $\alpha \in (0, 1]$ . Hence, it follows that

$$\begin{aligned} \text{Tr}[W_x^\alpha (U_{\mathcal{X}} \mathcal{W}^\alpha)^{\frac{1-\alpha}{\alpha}}] &= \text{Tr}[V^{x-1} W_1^\alpha V^{\dagger x-1} (U_{\mathcal{X}} \mathcal{W}^\alpha)^{\frac{1-\alpha}{\alpha}}] \\ &= \text{Tr}[W_1^\alpha (U_{\mathcal{X}} \mathcal{W}^\alpha)^{\frac{1-\alpha}{\alpha}}] \end{aligned} \quad (77)$$

for all  $x \in \mathcal{X}$  and  $\alpha \in (0, 1]$ . Summing Eq. (5) over all  $x \in \mathcal{X}$  and dividing by  $M$  yields that

$$\text{Tr}[W_x^\alpha (U_{\mathcal{X}} \mathcal{W}^\alpha)^{\frac{1-\alpha}{\alpha}}] = \text{Tr}[(U_{\mathcal{X}} \mathcal{W}^\alpha)^{\frac{1}{\alpha}}], \quad (78)$$

for all  $x \in \mathcal{X}$  and  $\alpha \in (0, 1]$ . Recalling Proposition 15 below, the above equation shows that the distribution  $U_{\mathcal{X}}$  indeed maximizes  $E_0(s, P)$ ,  $\forall s \in \mathbb{R}_{\geq 0}$ . Then we have

$$E_{\text{sp}}^{(1)}(R, U_{\mathcal{X}}) = \sup_{s \geq 0} \left\{ \max_{P \in \mathcal{P}(\mathcal{X})} E_0(s, P) - sR \right\} = E_{\text{sp}}(R).$$

Further, Jensen's inequality shows that  $E_{\text{sp}}^{(2)}(R, U_{\mathcal{X}}) \geq E_{\text{sp}}^{(1)}(R, U_{\mathcal{X}}) = E_{\text{sp}}(R)$ , and thus,  $E_{\text{sp}}^{(2)}(R, U_{\mathcal{X}}) = E_{\text{sp}}(R)$ .

Next, let  $(\alpha_R^*, \sigma_R^*)$  be the saddle-point of  $F_{R, U_{\mathcal{X}}}(\cdot, \cdot)$  (see Eq. (3)). One can observe from the definition of  $E_{\text{sp}}^{(2)}$  and Eq. (5) that all the quantities  $D_{\alpha_R^*}(W_x \| \sigma_R^*)$ ,  $x \in \mathcal{X}$ , are equal. By Proposition 2-(b), we obtain

$$\sigma_R^* = \frac{(U_{\mathcal{X}} \mathcal{W}^{\alpha_R^*})^{1/\alpha_R^*}}{\text{Tr} \left[ (U_{\mathcal{X}} \mathcal{W}^{\alpha_R^*})^{1/\alpha_R^*} \right]},$$

which, in turn, implies that

$$E_{\text{sp}}^{(2)}(R, P) = \sup_{\alpha \in (0, 1]} F_{R, P}(\alpha, \sigma_R^*) = \sup_{s \geq 0} \{E_0(s, U_{\mathcal{X}}) - sR\} = E_{\text{sp}}(R), \quad \forall P \in \mathcal{P}(\mathcal{X}). \quad (79)$$

Further, we have

$$|E'_{\text{sp}}(R)| = \frac{1 - \alpha_R^*}{\alpha_R^*} = \left| \frac{\partial E_{\text{sp}}^{(2)}(R, P)}{\partial R} \right|, \quad \forall P \in \mathcal{P}(\mathcal{X}). \quad (80)$$

Since Eqs. (5) and (80) indicates that every input sequence attains the sphere-packing exponent, we apply the same arguments in the proof of Theorem 8 to conclude this theorem.

**Proposition 15** ([14, Eq. (38)]). *Let  $s \in \mathbb{R}_{\geq 0}$  be arbitrary. The necessary and sufficient condition for the distribution  $P^*$  to maximize  $E_0(s, P)$  is*

$$\text{Tr} \left[ W_x^{1/(1+s)} \cdot \left( \sum_{\bar{x} \in \mathcal{X}} P^*(\bar{x}) W_{\bar{x}}^{1/(1+s)} \right)^s \right] \geq \text{Tr} \left[ \left( \sum_{\bar{x} \in \mathcal{X}} P^*(\bar{x}) W_{\bar{x}}^{1/(1+s)} \right)^{1+s} \right], \quad \forall x \in \mathcal{X}$$

with equality if  $P^*(x) \neq 0$ .

□

## 6. CONCLUSIONS

In this paper, we provided an exposition of sphere-packing bounds in classical and quantum channel coding. Unlike classical results, there are two different quantum sphere-packing exponents, one being stronger than the other. We provided variational representations for these two exponents, and showed that they are ordered by the Golden-Thompson inequality. Our proof strategy was inspired by Blahut's approach of hypothesis testing reduction [10] and Altuğ-Wagner's technique in strong large deviation theory [35]. Specifically, the prefactor of the bound, that is akin to the converse Hoeffding bound in quantum hypothesis testing, can be improved by Bahadur-Ranga Rao's sharp concentration inequality [26, 27]. Consequently, we obtained a refined strong sphere-packing bound for c-q channels and constant composition codes with a polynomial prefactor  $f(n) = n^{-\frac{1}{2}(1+|E'_{\text{sp}}(R)|+o(1))}$ . Moreover, the established result matches the best known random coding bound (i.e. achievability) up to the logarithmic order [35, 23, 24, 25]. For the case of general codes, the derived prefactor is of the polynomial order, i.e.  $f(n) = O(n^{-t})$  for some  $t > 1/2$ . We are able to obtain the exact prefactor without the assumption of constant composition codes for a class of symmetric c-q channels. We note that the exact prefactor for general codes is still open even in the classical case. Finally, our refinement enables a moderate deviation analysis in c-q channels [29] (see also [30]).

## ACKNOWLEDGEMENTS

M.-H. Hsieh was supported by an ARC Future Fellowship under Grant FT140100574 and by US Army Research Office for Basic Scientific Research Grant W911NF-17-1-0401. H.-C. Cheng was supported by Ministry of Science and Technology Overseas Project for Post Graduate Research (Taiwan) under Grant 105-2917-I-002-028, 108-2917-I-564-042, and 104-2221-E-002-072. The author would like to thank two reviewers and the editor for their useful comments. The author would like to thank Barış Nakiboğlu for the suggestion on the manuscript.

## APPENDIX A. LENGENDRE-FENCHEL TRANSFORM AND ERROR-EXPONENT FUNCTIONS

In this section, we will see that the Lengendre-Fenchel transform is closely related to the error-exponent function of hypothesis testing and channel coding. Let  $Z$  be a random variable, and denote by  $\Lambda(\alpha) := \log \mathbb{E}[e^{(1-\alpha)Z}]$  the cumulant-generating function. We call the following quantity the *Lengendre-Fenchel transform* of  $\Lambda(\alpha)$ :

$$\Lambda^*(z) := \sup_{\alpha \in \mathbb{R}} \{z(1-\alpha) - \Lambda(\alpha)\}.$$

This quantity arises in the exponent of various large deviation concentration inequalities [27]. By choosing  $Z$  to be the log-likelihood of two distributions as will be shown later, the Lengendre-Fenchel transform will determine how fast the errors in a hypothesis testing exponentially decay.

Consider the following binary hypotheses:

$$\begin{aligned} H_0 : p^n &:= p_{x_1} \otimes p_{x_2} \otimes \cdots p_{x_n}, \\ H_1 : q^n &:= q_{x_1} \otimes q_{x_2} \otimes \cdots q_{x_n}, \end{aligned} \tag{81}$$

where  $p_{x_i}, q_{x_i}$  for  $x_i \in \mathcal{X}$ ,  $i \in [n]$  are probability mass functions with  $p^n \ll q^n$ . Given any  $r \geq 0$ , recall the definition of the error-exponent function in Eq. (2.2):

$$\phi_n(r) = \phi_n(r|p^n||q^n) = \sup_{\alpha \in (0,1]} \left\{ \frac{1-\alpha}{\alpha} \left( \frac{1}{n} D_\alpha(p^n||q^n) - r \right) \right\}. \tag{82}$$

Without loss of generality, we set zero all elements of  $q_{x_i}$  that do not lie in the support of  $p_{x_i}$ , i.e.  $q_{x_i}(\omega) = 0$ ,  $\omega \notin \text{supp}(p_{x_i})$ ,  $i \in [n]$ , because those elements do not contribute in  $\phi_n(r)$ .

Let random variable  $Z_0 = \log \frac{q^n}{p^n}$  with probability distribution  $p^n$ , and let  $Z_1 = \log \frac{p^n}{q^n}$  with probability distribution  $q^n$ . We denote their cumulant generating functions by

$$\begin{aligned} \Lambda_{0,n}(\alpha) &:= \frac{1}{n} \log \mathbb{E}_{p^n} [e^{(1-\alpha)Z_0}] = \frac{1}{n} \sum_{i \in [n]} \Lambda_{0,i}(\alpha), \\ \Lambda_{1,n}(\alpha) &:= \frac{1}{n} \log \mathbb{E}_{q^n} [e^{(1-\alpha)Z_1}] = \frac{1}{n} \sum_{i \in [i]} \Lambda_{1,i}(\alpha); \end{aligned}$$

where

$$\Lambda_{0,i}(\alpha) := \log \mathbb{E}_{p_i} \left[ e^{(1-\alpha) \log \frac{q_i}{p_i}} \right], \quad \Lambda_{1,i}(\alpha) := \log \mathbb{E}_{q_i} \left[ e^{(1-\alpha) \log \frac{p_i}{q_i}} \right].$$

Rewrite the right-hand side of Eq. (A) with  $\alpha = \frac{1}{1+s}$ , and observe that

$$\begin{aligned} \sum_{x \in \mathcal{X}} P_{\mathbf{x}^n}(x) s D_{\frac{1}{1+s}}(p_x||q_x) &= -(1+s) \Lambda_{0,P_{\mathbf{x}^n}} \left( \frac{1}{1+s} \right) \\ &=: E_0^{(2)}(s, P_{\mathbf{x}^n}). \end{aligned} \tag{83}$$

Then the error-exponent function in Eq. (A) can also be viewed as a Lengendre-Fenchel transform of  $E_0^{(2)}(s, P_{\mathbf{x}^n})$ :

$$\phi_n(r) = \sup_{s \geq 0} \left\{ E_0^{(2)}(s, P_{\mathbf{x}^n}) - sr \right\}. \tag{84}$$

The following lemma relates  $\phi_n(r)$  to  $\Lambda_{j,n}^*(z)$ , the Lengendre-Fenchel transform of Eq. (A):

$$\Lambda_j^*(z) := \sup_{\alpha \in \mathbb{R}} \{z(1-\alpha) - \Lambda_{j,n}(\alpha)\}, \quad j \in \{0, 1\}. \tag{85}$$

Such a transform plays a significant role in concentration inequalities, convex analysis, and large deviation theory [27].

**Lemma 16.** *Let  $p^n$  and  $q^n$  be described as above. Assume  $r > \frac{1}{n} D_0(p^n||q^n)$  and  $\phi_n(r) > 0$ . The following hold:*

$$(a) \quad \Lambda_{0,n}''(\alpha) > 0 \text{ for all } \alpha \in [0, 1].$$

- (b)  $\Lambda_{0,n}^*(\phi_n(r) - r) = \phi_n(r)$ .  
(c)  $\Lambda_{1,n}^*(r - \phi_n(r)) = r$ .  
(d) The optimizer  $\alpha^* \in (0, 1)$  of  $\Lambda_0^*(z)$  in Eq. (A) is unique, and satisfies  $\Lambda'_{0,P_{\mathbf{x}^n}}(\alpha^*) = \phi_n(r) - r$ . In particular, one has  $\alpha^* = \frac{s^*}{1+s^*}$  and

$$s^* = -\frac{\partial \phi_n(r)}{\partial r};$$

$$\frac{\partial^2 \phi_n(r)}{\partial r^2} = -\left(\frac{\partial^2 E_0^{(2)}(s, P_{\mathbf{x}^n})}{\partial s^2}\bigg|_{s=s^*}\right)^{-1} = \frac{(1+s^*)^3}{\Lambda''_{0,P_{\mathbf{x}^n}}(\alpha^*)} > 0.$$

Before proving Lemma 16, we will need the following partial derivatives with respect to  $t$ :

$$\begin{aligned}\Lambda'_{0,i}(\alpha) &= \mathbb{E}_{v_{i,\alpha}} \left[ \log \frac{p_i}{q_i} \right], & \Lambda'_{1,i}(\alpha) &= \mathbb{E}_{v_{i,1-\alpha}} \left[ \log \frac{q_i}{p_i} \right]; \\ \Lambda''_{0,i}(\alpha) &= \text{Var}_{v_{i,\alpha}} \left[ \log \frac{p_i}{q_i} \right], & \Lambda''_{1,i}(\alpha) &= \text{Var}_{v_{i,\alpha}} \left[ \log \frac{q_i}{p_i} \right],\end{aligned}\tag{86}$$

where we denote the *tilted distributions* for every  $i \in [n]$  and  $t \in [0, 1]$  by

$$v_{i,\alpha}(\omega) := \frac{p_i^\alpha(\omega) q_i^{1-\alpha}(\omega)}{\sum_{\bar{\omega}} p_i^\alpha(\bar{\omega}) q_i^{1-\alpha}(\bar{\omega})}, \quad \bar{\omega} \in \text{supp}(p_i).$$

Since  $p^n$  and  $q^n$  share the same support, the above derivatives are well-defined. Further, it is not hard to verify that

$$\Lambda_{0,i}(\alpha) = \Lambda_{1,i}(1-\alpha), \quad \Lambda'_{0,i}(\alpha) = -\Lambda'_{1,i}(1-\alpha), \quad \Lambda''_{0,i}(\alpha) = \Lambda''_{1,i}(1-\alpha).\tag{87}$$

This lemma closely follows Ref. [35, Lemma 9]; however, the major difference is that we prove the claim using  $\phi_n(r|\rho^n||\sigma^n)$  in Eq. (2.2) instead of the discrimination function:  $\min \{D(\tau||\rho) : D(\tau||\sigma) \leq r\}$  in Eq. (3). This expression is crucial to obtaining the sphere-packing bound in Theorem 8 in the strong form, cf. Eq. (1), instead of the weak form, cf. Eq. (1).

*Proof of Lemma 16-(a).* We will prove this statement by contradiction. Let  $\alpha \in [0, 1]$ , Assuming that  $\Lambda''_{0,n}(\alpha) = 0$ , implies  $\Lambda''_{0,i}(\alpha) = 0$ ,  $\forall i \in [n]$ . Recall from Eq. (A)

$$0 = \Lambda''_{0,i}(\alpha) = \text{Var}_{v_{i,\alpha}} \left[ \log \frac{p_i}{q_i} \right],$$

which is equivalent to

$$p_i(\omega) = q_i(\omega) \cdot e^{\Lambda'_{0,i}(\alpha)}, \quad \forall \omega \in \text{supp}(p_i).\tag{88}$$

Summing both sides of Eq. (A) over  $\omega \in \text{supp}(p_i)$  gives

$$1 = \text{Tr} [p_i^0 q_i] e^{\Lambda'_{0,i}(\alpha)}.\tag{89}$$

Then, Eqs. (A) and (A) imply that

$$\begin{aligned}\phi_n(r) &= \sup_{0 < \alpha \leq 1} \frac{\alpha - 1}{\alpha} \left( r - \frac{1}{n} D_\alpha(p^n || q^n) \right) \\ &= \sup_{0 < \alpha \leq 1} \frac{\alpha - 1}{\alpha} \left( r + \frac{1}{n} \sum_{i \in [n]} \log \text{Tr} [p_i^0 q_i] \right) \\ &= 0,\end{aligned}\tag{90}$$

where Eq. (A) follows since  $r > \frac{1}{n} D_0(p^n || q^n)$  by assumption. However, this contradicts with the assumption  $\phi_n(r) > 0$ . Hence, we conclude item (a).  $\square$

*Proof of Lemma 16-(b).* In the following, we use the substitution  $s = \frac{1-\alpha}{\alpha}$  for convenience.

Observe that  $E_0^{(2)}(s, P_{\mathbf{x}^n}) - sr$  in Eq. (A) is strictly concave in  $s \in \mathbb{R}_{\geq 0}$  since

$$\frac{\partial^2 E_0^{(2)}(s, P_{\mathbf{x}^n})}{\partial s^2} = -\frac{1}{(1+s)^3} \Lambda_{0, P_{\mathbf{x}^n}}'' \left( \frac{s}{1+s} \right) < 0, \quad (91)$$

owing to Eqs. (A), (A), and item (a). Moreover,  $s = 0$  cannot be an optimum in Eq. (A); otherwise, it will violate the assumption  $\phi_n(r) \geq 0$ . Thus a unique maximizer  $s^* \in \mathbb{R}_{> 0}$  exists such that

$$\phi_n(r) = -s^*r + E_0^{(2)}(s^*, P_{\mathbf{x}^n}) \quad (92)$$

$$= \frac{1}{1+s^*} \Lambda_{0, P_{\mathbf{x}^n}}' \left( \frac{1}{1+s^*} \right) - \Lambda_{0, P_{\mathbf{x}^n}} \left( \frac{1}{1+s^*} \right). \quad (93)$$

where in the second equality we use Eq. (A) and

$$r = \left. \frac{\partial E_0^{(2)}(s, P_{\mathbf{x}^n})}{\partial s} \right|_{s=s^*} \quad (94)$$

$$= -\frac{1}{1+s^*} \Lambda_{0, P_{\mathbf{x}^n}}' \left( \frac{1}{1+s^*} \right) - \Lambda_{0, P_{\mathbf{x}^n}} \left( \frac{1}{1+s^*} \right). \quad (95)$$

Comparing Eq. (A) with (A) gives

$$\Lambda_{0, P_{\mathbf{x}^n}}' \left( \frac{1}{1+s^*} \right) = \phi_n(r) - r, \quad (96)$$

which is exactly the optimum solution to  $\Lambda_{0, P_{\mathbf{x}^n}}^*(z)$  in Eq. (A) with

$$\begin{aligned} \alpha^* &= \frac{1}{1+s^*} \in (0, 1), \\ z &= \phi_n(r) - r. \end{aligned} \quad (97)$$

Hence, we obtain

$$\begin{aligned} \Lambda_{0, P_{\mathbf{x}^n}}^*(\phi_n(r) - r) &= \frac{1}{1+s^*} z - \Lambda_{0, P_{\mathbf{x}^n}} \left( \frac{1}{1+s^*} \right) \\ &= \frac{1}{1+s^*} (\phi_n(r) - r) - \Lambda_{0, P_{\mathbf{x}^n}} \left( \frac{1}{1+s^*} \right) \\ &= \frac{s^*}{1+s^*} \Lambda_{0, P_{\mathbf{x}^n}}' \left( \frac{1}{1+s^*} \right) - \Lambda_{0, P_{\mathbf{x}^n}} \left( \frac{s^*}{1+s^*} \right) \\ &= \phi_n(r), \end{aligned}$$

where Eqs. (A) and (A) are used in the third and last equalities.  $\square$

*Proof of Lemma 16-(c).* This proof follows from similar arguments in item (b) and Eq. (A). Eqs. (A) and (A) lead to

$$\Lambda_{1, P_{\mathbf{x}^n}}' \left( \frac{s^*}{1+s^*} \right) = r - \phi_n(r), \quad (98)$$

which satisfies the optimum solution to  $\Lambda_{1, P_{\mathbf{x}^n}}(z)$  in Eq. (A) with  $\alpha^* = \frac{s^*}{1+s^*} \in (0, 1)$  and  $z = r - \phi_n(r)$ . Then,

$$\begin{aligned} \Lambda_{1, P_{\mathbf{x}^n}}^*(r - \phi_n(r)) &= \alpha^* z - \Lambda_{1, P_{\mathbf{x}^n}}(\alpha^*) \\ &= \frac{s^*}{1+s^*} (r - \phi_n(r)) - \Lambda_{1, P_{\mathbf{x}^n}} \left( \frac{s^*}{1+s^*} \right) \\ &= \frac{1}{1+s^*} \Lambda_{1, P_{\mathbf{x}^n}}' \left( \frac{s^*}{1+s^*} \right) - \Lambda_{1, P_{\mathbf{x}^n}} \left( \frac{s^*}{1+s^*} \right) \\ &= r, \end{aligned}$$

where the third equality is due to Eq. (A), and the last equality follows from Eqs. (A) and (A).  $\square$

*Proof of Lemma 16-(d).* The fact that a unique optimizer  $\alpha^* \in (0, 1)$  exists such that  $\Lambda'_{0, P_{\mathbf{x}^n}}(\alpha^*) = \phi_n(r) - r$  follows directly from Eqs. (A), (A) and  $\Lambda''_{0, P_{\mathbf{x}^n}}(\alpha) > 0$ , for  $\alpha \in [0, 1]$ .

Moreover, Eqs. (A), (A), and (A) yield

$$-\frac{\partial \phi_n(r)}{\partial r} = s^*,$$

$$\frac{\partial^2 \phi_n(r)}{\partial r^2} = -\frac{\partial s^*}{\partial r} = -\left(\frac{\partial^2 E_0^{(2)}(s, P_{\mathbf{x}^n})}{\partial s^2}\right)^{-1}\bigg|_{s=s^*} = \frac{(1+s^*)^3}{\Lambda_{0, P_{\mathbf{x}^n}}\left(\frac{1}{1+s^*}\right)},$$

which completes the claim in item (d).  $\square$

## APPENDIX B. A TIGHT LARGE DEVIATION INEQUALITY

Let  $(Z_i)_{i=1}^n$  be a sequence of independent, real-valued random variables with probability measures  $(\mu_i)_{i=1}^n$ . Let  $\Lambda_i(\alpha) := \log \mathbb{E}[e^{(1-\alpha)Z_i}]$  and define the Legendre-Fenchel transform of  $\frac{1}{n} \sum_{i=1}^n \Lambda_i(\cdot)$  to be:

$$\Lambda_n^*(z) := \sup_{\alpha \in \mathbb{R}} \left\{ z(1-\alpha) - \frac{1}{n} \sum_{i=1}^n \Lambda_i(\alpha) \right\}, \quad \forall z \in \mathbb{R}.$$

The so-called *large deviation inequality* means estimate the probability of sum of independent random variables that deviate from the mean by a linear amount, i.e.  $\Pr\left\{\frac{1}{n} \sum_{i=1}^n Z_i \geq z\right\}$ . Cramér's theorem [27, Chapter 2] states that

$$\Pr\left\{\frac{1}{n} \sum_{i=1}^n Z_i \geq z\right\} = e^{-n\Lambda_n^*(z)+o(n)}, \quad (99)$$

where  $e^{o(n)}$  is some subexponential prefactor. The upper bound of Eq. (B) can be simply shown by the exponent Chebyshev inequality:

$$\Pr\left\{\frac{1}{n} \sum_{i=1}^n Z_i \geq z\right\} \leq e^{-n\Lambda_n^*(z)},$$

while the lower bound is more involved. The Bahadur-Ranga Rao's concentration inequality below then provides a sharp lower bound and show that the subexponential prefactor can be improved to  $O(\frac{1}{\sqrt{n}})$ .

Let  $z \in \mathbb{R}$  be such that  $\exists \alpha^* \in (0, 1)$  and

$$\Lambda_n^*(z) = z(1-\alpha^*) - \frac{1}{n} \sum_{i=1}^n \Lambda_i(\alpha^*). \quad (100)$$

Define the probability measure  $\tilde{\mu}_i$  via

$$\frac{d\tilde{\mu}_i}{d\mu_i}(z_i) := e^{z_i(1-\alpha^*) - \Lambda_i(\alpha^*)},$$

and let  $\bar{Z}_i := Z_i - \mathbb{E}_{\tilde{\mu}_i}[Z_i]$ . Furthermore, define  $m_{2,n} := \frac{1}{n} \sum_{i=1}^n \text{Var}_{\tilde{\mu}_i}[\bar{Z}_i]$ ,  $m_{3,n} := \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\tilde{\mu}_i}[|\bar{Z}_i|^3]$ , and  $K_n(\alpha^*) := \frac{15\sqrt{2\pi}m_{3,n}}{m_{2,n}}$ . With these definitions, we can now state the following sharp concentration inequality for  $\frac{1}{n} \sum_{i=1}^n Z_i$ :

**Theorem 17** (Bahadur-Ranga Rao's Concentration Inequality [35, Proposition 5], [26]). *Provided Eq. (B) holds, then*

$$\Pr\left\{\frac{1}{n} \sum_{i=1}^n Z_i \geq z\right\} \geq e^{-n\Lambda_n^*(z)} \frac{e^{-K_n(\alpha^*)}}{2\sqrt{2n\pi m_{2,n}}} \left(1 - \frac{1 + (1 + K_n(\alpha^*))^2}{2\sqrt{2nm_{2,n}}}\right).$$



## APPENDIX C. UNIFORM CONTINUITY

The goal of this section is to present various uniform continuity properties, which play a significant role in proving finite blocklength converse bounds (see Propositions 12 and 13). Let  $r \in (C_{0,W}, C_{1,W})$  throughout this section. For any  $P \in \mathcal{P}_r(\mathcal{X}) := \left\{ P \in \mathcal{P}(\mathcal{X}) : E_{\text{sp}}^{(2)}(r, P) > 0 \right\}$ , we denote by  $(\alpha_{r,P}^*, \sigma_{r,P}^*) \in (0, 1) \times \mathcal{S}(\mathcal{H})$  the unique saddle-point of  $F(r, P)$  (see Proposition 3-(c)). For  $P \notin \mathcal{P}_r(\mathcal{X})$ , note that  $(1, \sigma)$  is a saddle-point of  $F(r, P)$  for all  $\sigma \in \mathcal{S}(\mathcal{H})$ . We thus choose  $(1, PW)$  to be the saddle-point of  $F(r, P)$  for  $P \notin \mathcal{P}_r(\mathcal{X})$ , subsequently. Define

$$B_r(P, W) := \sum_{x \in \mathcal{X}} P(x) \mathbb{E}_{v_{x, \alpha_{r,P}^*}} \left[ \log \frac{p_x}{q_x} \right]; \quad (101)$$

$$V_r(P, W) := \sum_{x \in \mathcal{X}} P(x) \mathbb{E}_{v_{x, \alpha_{r,P}^*}} \left[ \left| \log \frac{p_x}{q_x} - \mathbb{E}_{v_{x, \alpha_{r,P}^*}} \left[ \log \frac{p_x}{q_x} \right] \right|^2 \right]; \quad (102)$$

$$T_r(P, W) := \sum_{x \in \mathcal{X}} P(x) \mathbb{E}_{v_{x, \alpha_{r,P}^*}} \left[ \left| \log \frac{p_x}{q_x} - \mathbb{E}_{v_{x, \alpha_{r,P}^*}} \left[ \log \frac{p_x}{q_x} \right] \right|^3 \right], \quad (103)$$

where  $(p_x, q_x)$  is the Nussbaum-Szkoła distribution [62] of  $(W_x, \sigma_{r,P}^*)$ , and the *tilted distribution* is

$$v_{x,\alpha}(i, j) := \frac{p_x^\alpha(i, j) q_x^{1-\alpha}(i, j)}{\sum_{i,j} p_x^\alpha(i, j) q_x^{1-\alpha}(i, j)}, \quad \alpha \in [0, 1].$$

Inspired by Ref. [67, Lemma 62], we show the following continuity property, which are crucial for establishing the large deviation bounds in finite blocklength regime.

**Proposition 18** (Uniform Continuity). *Fix  $R \in (C_{0,W}, C_{1,W})$ . For every  $\underline{R} \in (C_{0,W}, R]$ ,  $B_r(P, W)$ ,  $V_r(P, W)$ , and  $T_r(P, W)$  are jointly continuous functions of  $(r, P)$  on  $[\underline{R}, R] \times \mathcal{P}(\mathcal{X})$ .*

*Remark C.1.* When  $R \geq I_1^{(2)}(P, W) = I(P, W)$ , Proposition 4-(a) implies that  $(\alpha_{R,P}^*, \sigma_{R,P}^*) = (1, PW)$ . In this case,  $V_R(P, W)$  equals to the *information variance*  $V(P)$  introduced by Tomamichel and Tan [48, Appendix B.3], and so does  $T_R(P, W) = T(P)$ . The established Proposition 18 covers the special case of the continuities of  $V(P)$  and  $T(P)$  in  $P$ , and provides a rigorous proof for [48, Lemma 29]. We emphasize that such a continuity property is a critical step to ensure that the third-order term in the asymptotic expansion of coding rate (see e.g. [67], [48]) independent of all codeword sequences.

*Proof of Proposition 18.* Inspecting the definitions given in Eqs. (C), (C), and (C). it is not hard to see that the quantities  $B_r(P, W)$ ,  $V_r(P, W)$ , and  $T_r(P, W)$  are sums of finitely many terms. We thus prove that each term is a continuous function in  $(r, P)$ . In the following, we first show the continuity of  $B_r(P, W)$ . The proof for  $V_r(P, W)$  and  $T_r(P, W)$  follow similarly.

Fix an arbitrary  $x \in \mathcal{X}$  onwards. Let  $(R_k, P_k)_{k \in \mathbb{N}}$  be an arbitrary sequence such that  $(R_k, P_k) \in [\underline{R}, R] \times \mathcal{P}(\mathcal{X})$ , and  $\lim_{k \rightarrow +\infty} (R_k, P_k) = (R_0, P_0) \in [\underline{R}, R] \times \mathcal{P}(\mathcal{X})$ . To ease the burden of notation, we let

$$\alpha_k := \alpha_{R_k, P_k}^*, \text{ and } \sigma_k := \sigma_{R_k, P_k}^*, \quad \forall k \in \mathbb{N}.$$

Note that the joint continuity proved in Proposition 3-(d) guarantees that

$$\begin{aligned} \lim_{k \rightarrow +\infty} \alpha_k &= \alpha_{R_0, P_0}^* =: \alpha_0, \\ \lim_{k \rightarrow +\infty} \sigma_k &= \sigma_{R_0, P_0}^* =: \sigma_0. \end{aligned} \quad (104)$$

Given the eigenvalue decompositions  $W_x = \sum_i \lambda_i |e_i\rangle\langle e_i|$  and  $\sigma_k = \sum_j \mu_j(\sigma_k) |f_j^k\rangle\langle f_j^k|$ , Nussbaum-Szkoła distributions are  $p_x(i, j) = \lambda_i |\langle e_i | f_j^k \rangle|^2$  and  $q_x(i, j) = \mu_j(\sigma_{R_k, P_k}^*) |\langle e_i | f_j^k \rangle|^2$ . Here, we write  $f_j^k$  and  $\mu_j(\sigma_k)$

to emphasize the dependence on  $P_k$ . To prove the continuity of  $B_r(P, \mathcal{W})$ , it suffices to show

$$\begin{aligned} & P_k(x) \frac{\lambda_i^{\alpha_k} \mu_j^{1-\alpha_k}(\sigma_k) |\langle e_i | f_j^k \rangle|^2}{\sum_{i,j} \lambda_i^{\alpha_k} \mu_j^{1-\alpha_k}(\sigma_k) |\langle e_i | f_j^k \rangle|^2} \log \frac{\lambda_i}{\mu_j(\sigma_k)} \\ & \longrightarrow P_0(x) \frac{\lambda_i^{\alpha_0} \mu_j^{1-\alpha_0}(\sigma_0) |\langle e_i | f_j^0 \rangle|^2}{\sum_{i,j} \lambda_i^{\alpha_0} \mu_j^{1-\alpha_0}(\sigma_0) |\langle e_i | f_j^0 \rangle|^2} \log \frac{\lambda_i}{\mu_j(\sigma_0)}. \end{aligned} \quad (105)$$

In the following, we first exclude some trivial cases. If  $\lambda_i = 0$ , then the convergence is obvious (recalling that the power function is only acting on the support). We assume  $\lambda_i > 0$  onwards. If  $P_k(x) > 0$  only for finite number of  $k$ , then the convergence in Eq. (C) is also trivial. We may assume  $P_k(x) > 0$  for all  $k \in \mathbb{N}$  (switching to a subsequence if necessary). Further, Proposition 3-(b) implies that  $W_x \ll \sigma_k$  for all  $k \in \mathbb{N}$ . We have  $\lambda_i |\langle e_i | f_j^k \rangle|^2 = 0$  whenever  $\mu_j(\sigma_k) |\langle e_i | f_j^k \rangle|^2 = 0$  by the absolute continuity, which in turn implies the convergence of Eq. (C). We may assume  $\mu_j(\sigma_k) |\langle e_i | f_j^k \rangle|^2 > 0$  for all  $k \in \mathbb{N}$ .

With the above assumptions, we study two cases:  $P_0(x) = 0$  or not, separately. If  $P_0(x) > 0$ , then  $W_x \ll \sigma_0$ . The absolute continuity again implies that  $\mu_j(\sigma_0) > 0$  by the previous argument. Hence, we can deduce that  $\mu_j(\sigma_k)$  is bounded away from zero. Using the continuity given in Eqs. (C) and the continuity of logarithm,  $\log \lambda_i / \mu_j(\sigma_k)$  tends to  $\log \lambda_i / \mu_j(\sigma_0)$ , which shows the convergence in Eq. (C).

It remains to show the case of  $P_0(x) = 0$ . Observe that the convergence in Eq. (C) holds when  $\mu_j(\sigma_k) \not\rightarrow 0$ . We thus consider the circumstance that  $\mu_j(\sigma_k) \rightarrow 0$ . To achieve our goal, we will show that the log-likelihood ratio  $\log \frac{\lambda_i}{\mu_j(\sigma_k)}$  does not diverge too fast.

In what follows we inspect the eigenvalue  $\mu_j(\sigma_k)$ . The saddle-point property given in Proposition 3-(a) and Proposition 2-(b) indicate that  $\sigma_k$  must satisfy

$$\sigma_k = \left( \sum_{\bar{x} \in \mathcal{X}} P_k(\bar{x}) \frac{W_{\bar{x}}^{\alpha_k}}{\text{Tr}[W_{\bar{x}}^{\alpha_k}(\sigma_k)^{1-\alpha_k}]} \right)^{\frac{1}{\alpha_k}}.$$

Further, since  $\alpha_{r,P}^* \in (0, 1]$  for all  $(r, P) \in (C_{0,W}, \infty] \times \mathcal{P}(\mathcal{X})$ , the continuity of  $P \mapsto \alpha_{r,P}^*$  given in Proposition 3-(d) and the compactness of  $\mathcal{P}(\mathcal{X})$  imply that

$$\alpha_{\underline{R}} := \min_{P \in \mathcal{P}(\mathcal{X})} \alpha_{\underline{R}, P}^* > 0 \quad (106)$$

By the convexity of  $r \mapsto E_{\text{sp}}^{(2)}(r, P)$  and Proposition 4-(c), we have  $\alpha_k \in [\alpha_{\underline{R}}, 1]$  for all  $k \in \{0\} \cup \mathbb{N}$ .

Therefore,

$$\begin{aligned} \mu_j(\sigma_k) &= \langle f_j^k | \sigma_k | f_j^k \rangle \\ &\geq \left( \sum_{\bar{x}} P_k(\bar{x}) \frac{\langle f_j^k | W_{\bar{x}}^{\alpha_k} | f_j^k \rangle}{\text{Tr}[W_{\bar{x}}^{\alpha_k}(\sigma_k)^{1-\alpha_k}]} \right)^{\frac{1}{\alpha_k}} \end{aligned} \quad (107)$$

$$\begin{aligned} &\geq \left( P_k(x) \frac{\langle f_j^k | W_x^{\alpha_k} | f_j^k \rangle}{\text{Tr}[W_x^{\alpha_k}(\sigma_k)^{1-\alpha_k}]} \right)^{\frac{1}{\alpha_k}} \\ &= \left( P_k(x) \frac{\sum_i \lambda_i^{\alpha_k} |\langle e_i | f_j^k \rangle|^2}{\sum_{i,j} \lambda_i^{\alpha_k} \mu_j^{1-\alpha_k}(\sigma_k) |\langle e_i | f_j^k \rangle|^2} \right)^{\frac{1}{\alpha_k}} \\ &\geq \left( P_k(x) \frac{\lambda_i^{\alpha_k} |\langle e_i | f_j^k \rangle|^2}{\sum_{i,j} \lambda_i^{\alpha_k} \mu_j^{1-\alpha_k}(\sigma_k) |\langle e_i | f_j^k \rangle|^2} \right)^{\frac{1}{\alpha_k}} \\ &\geq \lambda_i c_k^{\frac{1}{\alpha_k}}, \end{aligned} \quad (108)$$

where inequality (C) follows from Jensen's inequality<sup>9</sup> for the convex function  $(\cdot)^{\frac{1}{\alpha_k}}$  with  $\frac{1}{\alpha_k} \in (0, 1]$ , and we denote by  $c_k := P_k(x) \frac{|e_i|f_j^k|^2}{\sum_{i,j} \lambda_i^{\alpha_k} \mu_j^{1-\alpha_k}(\sigma_k) |\langle e_i|f_j^k \rangle|^2}$  in the last line. Note that  $\mu_j(\sigma_k) \leq 1$ . Eq. (C) then implies

$$\begin{aligned} \left| \log \frac{\lambda_i}{\mu_j(\sigma_k)} \right| &\leq \log \frac{1}{\lambda_i} - \log \left( \lambda_i c_k^{\frac{1}{\alpha_R}} \right) \\ &= 2 \log \frac{1}{\lambda_i} - \frac{1}{\alpha_R} \log c_k. \end{aligned} \quad (109)$$

Since we assume  $\mu_j(\sigma_k) \rightarrow 0$  and  $\lambda_i > 0$ , Eq. (C) guarantees that

$$c_k \rightarrow 0. \quad (110)$$

Using Eqs. (C), (C), (C), and the fact that  $\lambda_i^{\alpha_k} \mu_j^{1-\alpha_k}(\sigma_k) \in [0, 1]$  for all  $k \in \mathbb{N}$ , we are able show that the left-hand side of Eq. (C) converges to 0:

$$\begin{aligned} &P_k(x) \frac{\lambda_i^{\alpha_k} \mu_j^{1-\alpha_k}(\sigma_k) |\langle e_i|f_j^k \rangle|^2}{\sum_{i,j} \lambda_i^{\alpha_k} \mu_j^{1-\alpha_k}(\sigma_k) |\langle e_i|f_j^k \rangle|^2} \left| \log \frac{\lambda_i}{\mu_j(\sigma_k)} \right| \\ &\leq c_k \left| \log \frac{\lambda_i}{\mu_j(\sigma_k)} \right| \\ &\leq 2c_k \log \frac{1}{\lambda_i} - \frac{1}{\alpha_R} c_k \log c_k \\ &\rightarrow 0, \end{aligned}$$

which proves the joint continuity of  $(r, P) \mapsto B_r(P, \mathcal{W})$ .

Next, we show the continuity of  $V_r(P, \mathcal{W})$  and  $T_r(P, \mathcal{W})$ . Denote by  $B_r(W_x \| \sigma_k) := \mathbb{E}_{v_x, \alpha_k} [\log p_x / q_x]$  for convenience. For  $P_0(x) > 0$ ,  $\mu_j(\sigma_k)$  is bounded away from zero. Then,  $\log \lambda_i / \mu_j(\sigma_k)$  tends to  $\log \lambda_i / \mu_j(\sigma_0)$ , and it is not hard to see that  $B_{R_k}(W_x \| \sigma_k) \rightarrow B_{R_k}(W_x \| \sigma_0)$ . It suffices to prove the convergence when  $P_k(x) \rightarrow 0$  and  $\mu_j(\sigma_k) \rightarrow 0$  as mentioned before. Eq. (C) immediately implies that

$$\begin{aligned} B_{R_k}(W_x \| \sigma_k) &= \sum_{i,j} \frac{\lambda_i^{\alpha_k} \mu_j^{1-\alpha_k}(\sigma_k) |\langle e_i|f_j^k \rangle|^2}{\sum_{i,j} \lambda_i^{\alpha_k} \mu_j^{1-\alpha_k}(\sigma_k) |\langle e_i|f_j^k \rangle|^2} \log \frac{\lambda_i}{\mu_j(\sigma_k)} \\ &\leq 2 \log \frac{1}{\lambda_i} - \frac{1}{\alpha_R} \log c_k. \end{aligned} \quad (111)$$

Using the inequality  $|a + b|^2 \leq 2(|a|^2 + |b|^2)$ , we obtain

$$\begin{aligned} &P_k(x) \frac{\lambda_i^{\alpha_k} \mu_j^{1-\alpha_k}(\sigma_k) |\langle e_i|f_j^k \rangle|^2}{\sum_{i,j} \lambda_i^{\alpha_k} \mu_j^{1-\alpha_k}(\sigma_k) |\langle e_i|f_j^k \rangle|^2} \left| \log \frac{\lambda_i}{\mu_j(\sigma_k)} - B_{R_k}(W_x \| \sigma_k) \right|^2 \\ &\leq 2c_k \left| \log \frac{\lambda_i}{\mu_j(\sigma_k)} \right|^2 + 2c_k B_{R_k}^2(W_x \| \sigma_k). \end{aligned} \quad (112)$$

Combining Eqs. (C), (C), (C), and (C), we prove the continuity of  $V_r(P, \mathcal{W})$ .

Similarly, using the inequality  $|a + b|^3 \leq 4(|a|^3 + |b|^3)$  gives

$$\begin{aligned} &P_k(x) \frac{\lambda_i^{\alpha_k} \mu_j^{1-\alpha_k}(\sigma_k) |\langle e_i|f_j^k \rangle|^2}{\sum_{i,j} \lambda_i^{\alpha_k} \mu_j^{1-\alpha_k}(\sigma_k) |\langle e_i|f_j^k \rangle|^2} \left| \log \frac{\lambda_i}{\mu_j(\sigma_k)} - B_{\alpha_k}(W_x \| \sigma_k) \right|^3 \\ &\leq 4c_k \left| \log \frac{\lambda_i}{\mu_j(\sigma_k)} \right|^3 + 4c_k B_{R_k}^3(W_x \| \sigma_k). \end{aligned} \quad (113)$$

<sup>9</sup>For every Hermitian matrix  $B$  with eigenvalue decomposition  $B = \sum_i \lambda_i |u_i\rangle\langle u_i|$ , it follows that for every unit vector  $|v\rangle$  and positive convex function  $f$ ,  $f(\langle v|B|v\rangle) \leq f(\sum_i \lambda_i |\langle v|u_i\rangle|^2) \leq \sum_i f(\lambda_i) |\langle v|u_i\rangle|^2 = \langle v|f(B)|v\rangle$ .

Further, Eq. (C) implies

$$\left| \log \frac{\lambda_i}{\mu_j(\sigma_k)} \right|^3 \leq -4 \log^3 \lambda_i - \frac{4}{\alpha_R} \log^3 c_k. \quad (114)$$

Combining Eqs. (C), (C), (C), and (C), proves the continuity of  $T_r(P, \mathcal{W})$ .  $\square$

#### APPENDIX D. PROOF OF PROPOSITION 2

**Proposition 2** (Properties of order  $\alpha$  Augustin Information and Radius). *Given any classical-quantum channel  $\mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$  with  $|\mathcal{X}| < \infty$ , the following hold:*

- (a) *For every  $P \in \mathcal{P}(\mathcal{X})$ ,  $\alpha \mapsto I_\alpha^{(2)}(P, \mathcal{W})$  is monotone increasing on  $[0, 1]$ , and  $I_\alpha^{(2)}(P, \mathcal{W}) \leq \log |\mathcal{X}|$  for all  $\alpha \in [0, 1]$ .*
- (b) *For every  $(\alpha, P) \in (0, 1] \times \mathcal{P}(\mathcal{X})$ , there exists a unique  $\sigma_{\alpha, P} \in \mathcal{S}(\mathcal{H})$ , termed Augustin mean, such that*

$$I_\alpha^{(2)}(P, \mathcal{W}) = D_\alpha(\mathcal{W} \| \sigma_{\alpha, P} | P),$$

and

$$\mathsf{T}_{\alpha, P}(\sigma) = \sigma \text{ and } \sigma \gg P\mathcal{W} \text{ if and only if } \sigma = \sigma_{\alpha, P},$$

where the map  $\mathsf{T}_{\alpha, P} : \mathcal{S}_{P, \mathcal{W}}(\mathcal{H}) \rightarrow \mathcal{S}(\mathcal{H})$  is defined as

$$\mathsf{T}_{\alpha, P}(\sigma) = \sum_{x \in \mathcal{X}} P(x) \frac{\sigma^{\frac{1-\alpha}{2}} W_x^\alpha \sigma^{\frac{1-\alpha}{2}}}{\text{Tr}[W_x^\alpha \sigma^{1-\alpha}]}.$$

- (c) *For every  $\alpha \in [0, 1]$ , the map  $P \mapsto I_\alpha^{(2)}(P, \mathcal{W})$  is concave on  $\mathcal{P}(\mathcal{X})$ .*
- (d) *For every  $P \in \mathcal{P}(\mathcal{X})$ ,  $\alpha \mapsto \frac{1-\alpha}{\alpha} I_\alpha^{(2)}(P, \mathcal{W})$  is concave on  $(0, 1]$ .*
- (e) *For every  $P \in \mathcal{P}(\mathcal{X})$ ,  $\alpha \mapsto I_\alpha^{(2)}(P, \mathcal{W})$  is continuous on  $[0, 1]$ .*
- (f) *The family of functions  $\{I_\alpha^{(2)}(P, \mathcal{W})\}_{\alpha \in [0, 1]}$  is uniformly equicontinuous in  $P \in \mathcal{P}(\mathcal{X})$ . Moreover, The map  $(\alpha, P) \mapsto I_\alpha^{(2)}(P, \mathcal{W})$  is jointly continuous on  $[0, 1] \times \mathcal{P}(\mathcal{X})$ .*
- (g) *The map  $(\alpha, P) \mapsto \sigma_{\alpha, P}$  is jointly continuous on  $(0, 1] \times \mathcal{P}(\mathcal{X})$ .*
- (h) *The map  $\alpha \mapsto C_{\alpha, \mathcal{W}}$  is continuous and monotone increasing on  $[0, 1]$ .*

*Proof of Proposition 2-(a).* Recalling the definition of  $I_\alpha^{(2)}$  given in Eq. (2.1). The statement immediately follows from Lemma 1-(a) (see also [50, Lemma 4.6]) because the minimization over  $\sigma \in \mathcal{S}(\mathcal{H})$  preserves the monotonicity. Hence, we have  $I_\alpha^{(2)}(P, \mathcal{W}) \leq I_1(P, \mathcal{W}) \leq \log |\mathcal{X}|$ , where the last inequality follows from the well-known upper bound for the Holevo quantity (see e.g. [69, Chapter 12]).  $\square$

*Proof of Proposition 2-(b).* We first note that the infimum in Eq. (2.1) can be attained. This can be verified by the following argument. Lemma 1-(c) shows that  $D_\alpha$  is lower semi-continuous in its second argument. Hence, the linear combination, i.e.  $D_\alpha(\mathcal{W} \| \sigma | P)$ , is also lower semi-continuous on  $\mathcal{S}(\mathcal{H})$ . Further,  $\mathcal{S}(\mathcal{H})$  is compact owing to the assumption of the finite-dimensional Hilbert space  $\mathcal{H}$ . Thus, the extreme value theorem [84, Chapter 30§12.2] guarantees that the infimum can be attained

For  $\alpha = 1$ , it is well-known that (see e.g. [83])  $\sigma_{1, P} = P\mathcal{W}$ . Using the fact  $P\mathcal{W} \gg W_x$  for all  $x \in \text{supp}(P)$ , the statements immediately follow.

We fix an arbitrary  $(\alpha, P) \in (0, 1) \times \mathcal{P}(\mathcal{X})$  subsequently. Without loss of generality, we may further assume

$$\bigcup_{x \in \text{supp}(P)} \text{supp}(W_x) = \mathbb{1}_{\mathcal{H}},$$

and hence  $P\mathcal{W}$  has full support. We first show that the minimizer  $\sigma_{\alpha, P}$  has full support too. Second, we prove the fixed-point property Eq. ((b)). Finally, we establish the uniqueness of  $\sigma_{\alpha, P}$ . We remark that the uniqueness has been proven by Dalai and Winter [18, Appendix D]. Here, we provide an alternative proof for the completeness. Our approach follows closely from Hayashi and Tomamichel [56, Appendix C].

Define

$$\mathcal{M}_\alpha(\mathcal{H}) := \arg \min_{\sigma \in \mathcal{S}(\mathcal{H})} D_\alpha(\mathcal{W} \parallel \sigma | P) = \arg \max_{\sigma \in \mathcal{S}(\mathcal{H})} g_\alpha(\sigma) = \arg \max_{\sigma \in \mathcal{S}_{P, \mathcal{W}}(\mathcal{H})} g_\alpha(\sigma)$$

where

$$g_\alpha(\sigma) := \sum_{x \in \mathcal{X}} P(x) \log \text{Tr} [W_x^\alpha \sigma^{1-\alpha}].$$

To show that the optimizer of  $g_\alpha(\cdot)$  has full support, we observe that the directional derivative on the boundary of  $\mathcal{S}(\mathcal{H})$  where at least one eigenvalue is zero in a direction that increases its rank diverges to positive infinite. Namely, it suffices to show

$$\lim_{t \rightarrow 0} \frac{g_\alpha((1-t)\sigma + t\sigma^\perp) - g_\alpha(\sigma)}{t} = +\infty, \quad (115)$$

where  $\sigma \in \mathcal{S}_{P, \mathcal{W}}(\mathcal{H})$  is some singular density operator, and  $\sigma^\perp := \frac{(\mathbb{1}_{\mathcal{H}} - \sigma)}{\text{Tr}[\mathbb{1}_{\mathcal{H}} - \sigma]}$ . For  $x \in \text{supp}(P)$  with  $W_x \ll \sigma$ , we have  $W_x \perp \sigma^\perp$ . It is not hard to see that

$$\begin{aligned} & \lim_{t \rightarrow 0} P(x) \frac{\log \text{Tr} [W_x^\alpha ((1-t)\sigma + t\sigma^\perp)^{1-\alpha}] - \log \text{Tr} [W_x^\alpha \sigma^{1-\alpha}]}{t} \\ &= \lim_{t \rightarrow 0} P(x) \frac{\log \text{Tr} [W_x^\alpha ((1-t)^{1-\alpha} \sigma^{1-\alpha} + t^{1-\alpha} (\sigma^\perp)^{1-\alpha})] - \log \text{Tr} [W_x^\alpha \sigma^{1-\alpha}]}{t} \end{aligned} \quad (116)$$

$$= \lim_{t \rightarrow 0} P(x) \frac{(1-\alpha) \log(1-t)}{t} \quad (117)$$

$$= \lim_{t \rightarrow 0} P(x) \frac{-(1-\alpha)}{1-t} \quad (118)$$

$$= -P(x)(1-\alpha) > -\infty \quad (119)$$

where Eq. (D) holds because  $\sigma \perp \sigma^\perp$ ; Eq. (D) is due to  $W_x \perp \sigma^\perp$ ; and Eq. (D) is owing to L'Hôpital's rule.

On the other hand, since  $\sigma$  is singular, there must be some  $x \in \text{supp}(P)$  such that  $W_x \not\ll \sigma$ . Hence, by denoting  $c := \frac{\text{Tr}[W_x^\alpha (\sigma^\perp)^{1-\alpha}]}{\text{Tr}[W_x^\alpha \sigma^{1-\alpha}]} > 0$ , Eq. (D) leads to

$$\begin{aligned} & \lim_{t \rightarrow 0} P(x) \frac{\log \{(1-t)^{1-\alpha} + t^{1-\alpha} c\}}{t} \\ &= \lim_{t \rightarrow 0} P(x) \frac{-(1-\alpha)(1-t)^{-\alpha} + (1-\alpha)t^{-\alpha} c}{(1-t)^{1-\alpha} + t^{1-\alpha} c} \end{aligned} \quad (120)$$

$$= +\infty, \quad (121)$$

where Eq. (D) is by L'Hôpital's rule again. Combining Eqs. (D) and (D) concludes Eq. (D).

Next, we show the fixed-point property:  $\mathcal{M}_\alpha(\mathcal{H}) = \mathcal{F}_\alpha(\mathcal{H})$ , where  $\mathcal{F}_\alpha(\mathcal{H}) := \{\sigma \in \mathcal{S}_{>0}(\mathcal{H})\}$  denotes the fixed-points of the map:  $\mathsf{T}_{\alpha, P} : \mathcal{S}_{P, \mathcal{W}}(\mathcal{H}) \rightarrow \mathcal{S}(\mathcal{H})$ . A necessary and sufficient condition for  $\sigma$  to be an optimizer is

$$\partial_\omega g_\alpha(\sigma) := \text{D}g_\alpha(\sigma)[\omega - \sigma] = 0,$$

for all  $\omega \in \mathcal{S}(\mathcal{H})$ , where  $\text{D}g_\alpha(\sigma)$  denotes the Fréchet derivative of the map  $g_\alpha$  (see e.g. [56, Appendix C]). Using the chain rule of Fréchet derivatives, it follows

$$\begin{aligned} \partial_\omega g_\alpha(\sigma) &= \text{Tr} \left[ \sum_{x \in \mathcal{X}} P(x) \frac{W_x^\alpha}{\text{Tr} [W_x^\alpha \sigma^{1-\alpha}]} \partial_\omega \sigma^{1-\alpha} \right] \\ &= \text{Tr} \left[ \sum_{x \in \mathcal{X}} P(x) \frac{\sigma^{-\frac{\alpha}{2}} W_x^\alpha \sigma^{-\frac{\alpha}{2}}}{\text{Tr} [W_x^\alpha \sigma^{1-\alpha}]} \sigma^{\frac{\alpha}{2}} \partial_\omega \sigma^{1-\alpha} \sigma^{\frac{\alpha}{2}} \right]. \end{aligned} \quad (122)$$

We claim that the operators

$$\left\{ \Delta_\omega = \sigma^{\frac{\alpha}{2}} \partial_\omega \sigma^{1-\alpha} \sigma^{\frac{\alpha}{2}} : \omega \in \mathcal{S}(\mathcal{H}) \right\}$$

span the space of traceless Hermitian operators on  $\mathcal{S}(\mathcal{H})$ . Let  $\sigma = \sum_i \lambda_i |i\rangle\langle i|$  with  $\lambda_i > 0$  be the eigenvalue decomposition. One can verify [77, Theorem 3.25] that

$$\langle i | \Delta_\omega | j \rangle = \begin{cases} (\lambda_i \lambda_j)^{\frac{\alpha}{2}} \frac{\lambda_i^{1-\alpha} - \lambda_j^{1-\alpha}}{\lambda_i - \lambda_j} \langle i | \omega - \sigma | j \rangle, & \text{if } \lambda_i \neq \lambda_j \\ (1 - \alpha) \langle i | \omega - \sigma | j \rangle, & \text{if } \lambda_i = \lambda_j \end{cases}.$$

Therefore,  $\Delta_\omega$  is Hermitian and  $\text{Tr}[\Delta_\omega] = 0$  for all  $\omega \in \mathcal{S}(\mathcal{H})$ . Moreover, the basis of the traceless Hermitian operators is given by the operators

$$\{\Gamma_{ij} = |i\rangle\langle j| + |j\rangle\langle i|, \Gamma'_{ij} = |i\rangle\langle j| - |j\rangle\langle i|, \Gamma''_{ij} = |i\rangle\langle i| - |j\rangle\langle j|\}_{i \neq j}.$$

For every tuple  $(i, j)$  with  $i \neq j$  there exists an  $\varepsilon > 0$  such that the state  $\omega = \sigma + \varepsilon \Gamma_{ij}$  is still in  $\mathcal{S}(\mathcal{H})$ . For this state, we find that  $\Delta_\omega = \eta \Gamma_{ij}$  for some real  $\eta > 0$ . The similar argument applies to  $\Gamma'_{ij}$  and  $\Gamma''_{ij}$ . Hence, we have verified that the operators  $\{\Delta_\omega\}_{\omega \in \mathcal{S}(\mathcal{H})}$  span the space of traceless Hermitian operators.

Armed with the above discussion, the condition that  $\partial_\omega g_\alpha(\sigma) = 0$  for all  $\omega \in \mathcal{S}(\mathcal{H})$  is equivalent to the condition that the operators

$$\sum_{x \in \mathcal{X}} P(x) \frac{\sigma^{-\frac{\alpha}{2}} W_x^\alpha \sigma^{-\frac{\alpha}{2}}}{\text{Tr}[W_x^\alpha \sigma^{1-\alpha}]}$$

must be proportional to the identity. Thus, the optimum must be a fixed point of the map  $\mathsf{T}_{\alpha, P}(\cdot)$ .

Lastly, to prove the uniqueness of the optimizer, it remains to show  $\partial_\omega^2 g_\alpha(\sigma) : \mathsf{D}^2 g_\alpha(\sigma)[\omega - \sigma, \omega - \sigma] < 0$  for all  $\omega \neq \sigma$  and  $\sigma > 0$ . Continuing on Eq. (D), we have

$$\begin{aligned} \partial_\omega^2 g_\alpha(\sigma) &= -\text{Tr} \left[ \sum_{x \in \mathcal{X}} P(x) \frac{W_x^\alpha}{\text{Tr}^2[W_x^\alpha \sigma^{1-\alpha}]} \partial_\omega \sigma^{1-\alpha} \right] + \text{Tr} \left[ \sum_{x \in \mathcal{X}} P(x) \frac{W_x^\alpha}{\text{Tr}[W_x^\alpha \sigma^{1-\alpha}]} \partial_\omega^2 \sigma^{1-\alpha} \right] \\ &< \text{Tr} \left[ \sum_{x \in \mathcal{X}} P(x) \frac{W_x^\alpha}{\text{Tr}[W_x^\alpha \sigma^{1-\alpha}]} \partial_\omega^2 \sigma^{1-\alpha} \right], \end{aligned} \quad (123)$$

where Eq. (D) holds by noting that  $\partial_\omega \sigma^{1-\alpha} \neq 0$  for all  $\omega \neq \sigma$ . Further,  $\partial_\omega^2 \sigma^{1-\alpha} \leq 0$  since  $u \mapsto u^{1-\alpha}$  is operator concave. Thus,  $\partial_\omega^2 g_\alpha(\sigma) < 0$ , item (b) is proved.  $\square$

*Proof of Proposition 2(c).* Recall the definition given in Eq. (2.1). The assertion follows because the pointwise infimum of linear functions is concave.  $\square$

*Proof of Proposition 2(d).* This assertion was proved by Mosonyi and Ogawa [50, Corollary B.2].  $\square$

*Proof of Proposition 2(e).* The idea of the proof originate from Nakiboğlu [41, Lemmas 16, 17].

Recalling item (d), the map  $\alpha \mapsto \frac{1-\alpha}{\alpha} I_\alpha^{(2)}(P, \mathcal{W})$  is concave on  $(0, 1]$ . Since any concave function is continuous in its interior [81, Corollary 6.3.3], the map  $\alpha \mapsto I_\alpha^{(2)}(P, \mathcal{W})$  is continuous on  $(0, 1)$ . The continuity at  $\alpha = 0$  can be verified as follows. Let  $\sigma_{0,P} \in \mathcal{S}(\mathcal{H})$  be any state such that  $I_0^{(2)}(P, \mathcal{W}) = D_0(\mathcal{W} \| \sigma_{0,P} | P)$ . Then, the monotonicity in item (a) and the definition of  $I_\alpha^{(2)}$  given in Eq. (2.1) imply that

$$I_0^{(2)}(P, \mathcal{W}) \leq I_\alpha^{(2)}(P, \mathcal{W}) \leq D_\alpha(\mathcal{W} \| \sigma_{0,P} | P), \quad \forall \alpha \in (0, 1].$$

The continuity of  $\alpha \mapsto D_\alpha$  on  $[0, 1]$ , Lemma 1(a), thus implies  $\lim_{\alpha \downarrow 0} I_\alpha(P, \mathcal{W}) = I_0(P, \mathcal{W})$ .

It remains to show the continuity at  $\alpha = 1$ . We claim the following fact about the Augustin mean.

**Lemma 19.** *Given  $\alpha \in (0, 1]$  and  $P \in \mathcal{P}(\mathcal{X})$ , we let  $\sigma_{\alpha,P} \in \mathcal{S}(\mathcal{H})$  be the Augustin mean, i.e.  $I_\alpha^{(2)}(P, \mathcal{W}) = D_\alpha(\mathcal{W} \| \sigma_{\alpha,P} | P)$ . The following hold.*

- For any  $\alpha \in [1/2, 1]$ ,

$$D_\alpha(W_x \| \sigma_{\alpha,P}) \leq \log \frac{1}{P(x)}.$$

- For any  $\alpha \in [1/2, 1]$ ,

$$\sigma_{\alpha,P} \leq \left( \min_{x:P(x)>0} P(x) \right)^{\frac{\alpha-1}{\alpha}} \sigma_{1,P}. \quad (124)$$

Then, Eq. (19) and Lemma 1-(b) imply that

$$D_\alpha(\mathcal{W} \| \sigma_{\alpha,P} | P) \geq D_\alpha(\mathcal{W} \| \sigma_{1,P} | P) + \frac{1-\alpha}{\alpha} \log \left( \min_{x:P(x)>0} P(x) \right), \quad \forall \alpha \in [1/2, 1].$$

Moreover, the fact  $D_\alpha(\mathcal{W} \| \sigma_{\alpha,P} | P) = I_\alpha^{(2)}(P, \mathcal{W})$  in item (b) and the monotonicity in item (a) show that

$$D_\alpha(\mathcal{W} \| \sigma_{1,P} | P) + \frac{1-\alpha}{\alpha} \log \left( \min_{x:P(x)>0} P(x) \right) \leq I_\alpha^{(2)}(P, \mathcal{W}) \leq I_1^{(2)}(P, \mathcal{W}), \quad \forall \alpha \in [1/2, 1].$$

It is clearly that  $D_\alpha(\mathcal{W} \| \sigma_{1,P} | P) < +\infty$  by the fact that  $\sigma_{1,P} \gg W_x$  for all  $x \in \text{supp}(P)$ . Then, the continuity of  $\alpha \mapsto D_\alpha$  on  $[0, 1]$ , Lemma 1-(a), together with the fact  $D_1(\mathcal{W} \| \sigma_{1,P} | P) = I_1^{(2)}(P, \mathcal{W})$  imply the continuity of  $I_\alpha^{(2)}$  at  $\alpha = 1$ .

In the following, we prove Lemma 19 to complete the proof of item (e).

*Proof of Lemma 19.* Proposition 2 (b) implies that the Augustin mean satisfies

$$\sigma_{\alpha,P} = \left( \sum_{x \in \mathcal{X}} P(x) W_x^\alpha e^{(1-\alpha)D_\alpha(W_x \| \sigma_{\alpha,P})} \right)^{\frac{1}{\alpha}}. \quad (125)$$

Using the operator monotonicity of  $(\cdot)^{\frac{1-\alpha}{\alpha}}$  for  $\alpha \in [1/2, 1]$  (see e.g. [71, 77]), we have

$$\sigma_{\alpha,P}^{1-\alpha} \geq P(x)^{\frac{1-\alpha}{\alpha}} W_x^{1-\alpha} e^{\frac{(1-\alpha)^2}{\alpha} D_\alpha(W_x \| \sigma_{\alpha,P})}. \quad (126)$$

Then, Eq. (D) and Lemma 1-(b) imply that

$$\begin{aligned} D_\alpha(W_x \| \sigma_{\alpha,P}) &= \frac{1}{\alpha-1} \log \text{Tr} \left[ W_x^\alpha \sigma_{\alpha,P}^{1-\alpha} \right] \\ &\leq -\log P(x)^{\frac{1}{\alpha}} - \frac{1-\alpha}{\alpha} D_\alpha(W_x \| \sigma_{\alpha,P}), \end{aligned}$$

which proves the first claim.

For any  $\alpha \in [1/2, 1]$ , the map  $(\cdot)^{\frac{1}{\alpha}}$  is operator convex. Then, Eq. (D) yields

$$\sigma_{\alpha,P} \leq \sum_{x \in \mathcal{X}} P(x) W_x e^{\frac{1-\alpha}{\alpha} D_\alpha(W_x \| \sigma_{\alpha,P})}. \quad (127)$$

Note that  $P\mathcal{W} = \sigma_{1,P}$ . Applying Eq. (D) on (D) gives the desired result in Eq. (19).  $\square$

$\square$

*Proof of Proposition 2-(f).* To prove the equicontinuity, we need the following inequality:

$$I_\alpha^{(2)}(\alpha, P_\beta) \leq \beta I_\alpha^{(2)}(\alpha, P_1) + (1-\beta) I_\alpha^{(2)}(\alpha, P_0) + H(\beta) \quad (128)$$

for any  $P_1, P_0 \in \mathcal{P}(\mathcal{X})$ ,  $P_\beta = \beta P_1 + (1-\beta)P_0$ ,  $\beta \in (0, 1)$ ,  $\alpha \in [0, 1]$ ; and we shorthand  $H(\beta) := -\beta \log \beta - (1-\beta) \log(1-\beta)$  the binary entropy function.



Let  $\sigma_{\alpha,P} \in \mathcal{S}(\mathcal{H})$  be the Augustin mean<sup>10</sup> as in item (b) for  $\alpha \in [0, 1]$ . Lemma 1-(b) implies that, for every  $\alpha \in [0, 1]$ ,

$$\begin{aligned}
& \sum_{x \in \mathcal{X}} P_\beta(x) D_\alpha(W_x \| \beta \sigma_{\alpha, P_1} + (1 - \beta) \sigma_{\alpha, P_0}) \\
&= \beta \sum_{x \in \mathcal{X}} P_1(x) D_\alpha(W_x \| \beta \sigma_{\alpha, P_1} + (1 - \beta) \sigma_{\alpha, P_0}) + (1 - \beta) \sum_{x \in \mathcal{X}} P_0(x) D_\alpha(W_x \| \beta \sigma_{\alpha, P_1} + (1 - \beta) \sigma_{\alpha, P_0}) \\
&\leq \beta \sum_{x \in \mathcal{X}} P_1(x) D_\alpha(W_x \| \sigma_{\alpha, P_1}) - \beta \log \beta + (1 - \beta) \sum_{x \in \mathcal{X}} P_0(x) D_\alpha(W_x \| \sigma_{\alpha, P_0}) - (1 - \beta) \log(1 - \beta) \\
&= \beta I_\alpha^{(2)}(P_1, \mathcal{W}) + (1 - \beta) I_\alpha(P_0, \mathcal{W}) + H(\beta).
\end{aligned}$$

Let  $s_\wedge, s_1, s_0$  be

$$\begin{aligned}
s_\wedge &= \frac{P_1 \wedge P_0}{\|P_1 \wedge P_0\|_1}, \\
s_1 &= \frac{P_1 - P_1 \wedge P_0}{1 - \|P_1 \wedge P_0\|_1}, \\
s_0 &= \frac{P_0 - P_1 \wedge P_0}{1 - \|P_1 \wedge P_0\|_1}.
\end{aligned}$$

Here,  $\|\cdot\|_1$  denotes the  $\ell_1$ -norm. One can verify that (see e.g. [41])

$$\begin{aligned}
P_1 &= \left(1 - \frac{\|P_1 - P_0\|_1}{2}\right) s_\wedge + \frac{\|P_1 - P_0\|_1}{2} s_1, \\
P_0 &= \left(1 - \frac{\|P_1 - P_0\|_1}{2}\right) s_\wedge + \frac{\|P_1 - P_0\|_1}{2} s_0.
\end{aligned}$$

Then, the concavity of  $P \mapsto I_\alpha^{(2)}(P, \mathcal{W})$  given in item (c) together with Eq. (D) yield

$$\begin{aligned}
I_\alpha^{(2)}(P_0, \mathcal{W}) - I_\alpha^{(2)}(P_1, \mathcal{W}) &\leq H\left(\frac{\|P_1 - P_0\|_1}{2}\right) + \frac{\|P_1 - P_0\|_1}{2} \left(I_\alpha^{(2)}(s_0, \mathcal{W}) - I_\alpha^{(2)}(s_1, \mathcal{W})\right) \\
&\leq H\left(\frac{\|P_1 - P_0\|_1}{2}\right) + \frac{\|P_1 - P_0\|_1}{2} I_\alpha^{(2)}(s_0, \mathcal{W})
\end{aligned}$$

for all  $\alpha \in [0, 1]$ . Thus,

$$\left| I_\alpha^{(2)}(P_0, \mathcal{W}) - I_\alpha^{(2)}(P_1, \mathcal{W}) \right| \leq H\left(\frac{\|P_1 - P_0\|_1}{2}\right) + \frac{\|P_1 - P_0\|_1}{2} \log |\mathcal{X}|$$

since  $\alpha \mapsto I_\alpha^{(2)}$  is monotone increasing by item (a). The above inequality implies the desired equicontinuity.

The joint continuity of  $(\alpha, P) \mapsto I_\alpha^{(2)}(P, \mathcal{W})$  follows from the continuity of  $\alpha \mapsto I_\alpha^{(2)}(P, \mathcal{W})$  given in (e) and uniform equicontinuity.  $\square$

*Proof of Proposition 2-(g).* Let  $(\alpha_k, P_k)_{k \in \mathbb{N}}$  be an arbitrary sequence such that  $\alpha_k \in (0, 1]$ ,  $P_k \in \mathcal{P}(\mathcal{X})$ , and  $\lim_{k \rightarrow +\infty} (\alpha_k, P_k) = (\alpha_0, P_0) \in (0, 1] \times \mathcal{P}(\mathcal{X})$ . Further, let  $(\sigma_{\alpha_k, P_k})_{k \in \mathbb{N}}$  be the sequence of the Augustin mean corresponding to  $(\alpha_k, P_k)$ . Since  $\mathcal{S}(\mathcal{H})$  is compact, there exists a convergent subsequence  $\{k_l\}_{l \in \mathbb{N}}$  such that  $\lim_{l \rightarrow +\infty} \sigma_{\alpha_{k_l}, P_{k_l}} = \sigma_0$  for some  $\sigma_0 \in \mathcal{S}(\mathcal{H})$ .

The joint continuity of  $(\alpha, P) \mapsto I_\alpha^{(2)}(P, \mathcal{W})$  in item (f) thus implies

$$\lim_{k \rightarrow +\infty} I_{\alpha_k}^{(2)}(P_k, \mathcal{W}) = D_{\alpha_0}(\mathcal{W} \| \sigma_0 | P_0) = I_{\alpha_0}^{(2)}(P_0, \mathcal{W}) = D_{\alpha_0}(\mathcal{W} \| \sigma_{\alpha_0, P_0} | P_0).$$

Then, the uniqueness of the minimizer  $\sigma_{\alpha, P}$  in item (b) guarantees that  $\sigma_0 = \sigma_{\alpha_0, P_0}$ . Hence,

$$\lim_{k \rightarrow +\infty} \sigma_{\alpha_k, P_k} = \sigma_0 = \sigma_{\alpha_0, P_0},$$

<sup>10</sup>For  $\alpha = 1$ , the Augustin mean is not unique. We note that the proof of item (f) does not require the uniqueness.

which proves item (g).  $\square$

*Proof of Proposition 2-(h).* Berge's maximum theorem [72, Section IV.3], [73, Lemma 3.1] shows that the continuous map  $(\alpha, P) \mapsto I_\alpha^{(2)}(P, \mathcal{W})$  maximized over the compact set  $P \in \mathcal{P}(\mathcal{X})$  is still continuous for  $\alpha \in [0, 1]$ .  $\square$

#### APPENDIX E. PROOF OF PROPOSITION 3

**Proposition 3** (Saddle-Point). *Consider a classical-quantum channel  $\mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$ , any  $R \in (C_{0,\mathcal{W}}, C_{\mathcal{W}})$ , and  $P \in \mathcal{P}(\mathcal{X})$ . Let*

$$\mathcal{S}_{P,\mathcal{W}}(\mathcal{H}) := \{\sigma \in \mathcal{S}(\mathcal{H}) : \forall x \in \text{supp}(P), W_x \not\perp \sigma\}.$$

Define

$$F_{R,P}(\alpha, \sigma) := \begin{cases} \frac{1-\alpha}{\alpha} (D_\alpha(\mathcal{W} \parallel \sigma | P) - R), & \alpha \in (0, 1) \\ 0, & \alpha = 1 \end{cases}, \quad (129)$$

on  $(0, 1] \times \mathcal{S}(\mathcal{H})$ , and denote by

$$\mathcal{P}_R(\mathcal{X}) := \left\{ P \in \mathcal{P}(\mathcal{X}) : \sup_{0 < \alpha \leq 1} \inf_{\sigma \in \mathcal{S}(\mathcal{H})} F_{R,P}(\alpha, \sigma) \in \mathbb{R}_{>0} \right\}. \quad (130)$$

The following holds

(a) For any  $P \in \mathcal{P}(\mathcal{X})$ ,  $F_{R,P}(\cdot, \cdot)$  has a saddle-point on  $(0, 1] \times \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})$  with the saddle-value:

$$\min_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{0 < \alpha \leq 1} F_{R,P}(\alpha, \sigma) = \sup_{0 < \alpha \leq 1} \min_{\sigma \in \mathcal{S}(\mathcal{H})} F_{R,P}(\alpha, \sigma) = E_{\text{sp}}^{(2)}(R, P).$$

(b) Fix  $P \in \mathcal{P}_R(\mathcal{X})$ . Any saddle-point  $(\alpha_{R,P}^*, \sigma_{R,P}^*)$  of  $F_{R,P}(\cdot, \cdot)$  satisfies  $\alpha_{R,P}^* \in (0, 1)$  and

$$\sigma_{R,P}^* \gg W_x, \quad \forall x \in \text{supp}(P).$$

(c) For  $P \in \mathcal{P}_R(\mathcal{X})$ , the saddle-point is unique.

(d) For any  $\underline{R} \in (C_{0,\mathcal{W}}, R]$ , both  $\alpha_{r,P}^*$  and  $\sigma_{r,P}^*$  are jointly continuous functions of  $(r, P)$  on  $[\underline{R}, R] \times \mathcal{P}(\mathcal{X})$ .

*Proof of Proposition 3-(a).* Fix arbitrary  $R > C_{0,\mathcal{W}}$  and  $P \in \mathcal{P}(\mathcal{X})$ . In the following, we prove the existence of a saddle-point of  $F_{R,P}(\cdot, \cdot)$  on  $(0, 1] \times \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})$ . Ref. [75, Lemma 36.2] states that  $(\alpha^*, \sigma^*)$  is a saddle point of  $F_{R,P}(\cdot, \cdot)$  if and only if the supremum in

$$\sup_{\alpha \in (0,1]} \inf_{\sigma \in \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})} F_{R,P}(\alpha, \sigma) \quad (131)$$

is attained at  $\alpha^* \in (0, 1]$ , the infimum in

$$\inf_{\sigma \in \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})} \sup_{\alpha \in (0,1]} F_{R,P}(\alpha, \sigma) \quad (132)$$

is attained at  $\sigma^* \in \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})$ , and the two extrema in Eqs. (E), (E) are equal and finite. We first claim that,  $\forall \alpha \in (0, 1]$ ,

$$\inf_{\sigma \in \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})} F_{R,P}(\alpha, \sigma) = \inf_{\sigma \in \mathcal{S}(\mathcal{H})} F_{R,P}(\alpha, \sigma). \quad (133)$$

To see this, observe that for any  $\alpha \in (0, 1)$ , Eqs. (2.1) and (2.1) yield

$$\forall \sigma \in \mathcal{S}(\mathcal{H}) \setminus \mathcal{S}_{P,\mathcal{W}}(\mathcal{H}), \quad D_\alpha(\mathcal{W} \parallel \sigma | P) = +\infty, \quad (134)$$

which, in turn, implies

$$\forall \sigma \in \mathcal{S}(\mathcal{H}) \setminus \mathcal{S}_{P,\mathcal{W}}(\mathcal{H}), \quad F_{R,P}(\alpha, \sigma) = +\infty. \quad (135)$$

Further, Eq. (E) holds trivially when  $\alpha = 1$ . Hence, Eq. (E) yields

$$\sup_{\alpha \in (0,1]} \inf_{\sigma \in \mathcal{S}_{P,\mathcal{W}}(\mathcal{H})} F_{R,P}(\alpha, \sigma) = \sup_{\alpha \in (0,1]} \inf_{\sigma \in \mathcal{S}(\mathcal{H})} F_{R,P}(\alpha, \sigma)$$

Owing to the fact  $R > C_{0,W}$  and Eq. (2.1), we have

$$E_{\text{sp}}^{(2)}(R, P) = \sup_{\alpha \in (0,1]} \inf_{\sigma \in \mathcal{S}(\mathcal{H})} F_{R,P}(\alpha, \sigma) < +\infty, \quad (136)$$

which guarantees the supremum in the right-hand side of Eq. (E) is attained at some  $\alpha \in (0, 1]$ . Namely, there exists some  $\bar{\alpha}_{R,P} \in (0, 1]$  such that

$$\sup_{\alpha \in (0,1]} \inf_{\sigma \in \mathcal{S}_{P,W}(\mathcal{H})} F_{R,P}(\alpha, \sigma) = \max_{\alpha \in [\bar{\alpha}_{R,P}, 1]} \inf_{\sigma \in \mathcal{S}(\mathcal{H})} F_{R,P}(\alpha, \sigma) < +\infty. \quad (137)$$

Thus, we complete our claim in Eq. (E). It remains to show that the infimum in Eq.(E) is attained at some  $\sigma^* \in \mathcal{S}_{P,W}(\mathcal{H})$  and the supremum and infimum are exchangeable. To achieve this, we will show that  $([\bar{\alpha}_{R,P}, 1], \mathcal{S}_{P,W}(\mathcal{H}), F_{R,P})$  is a closed saddle-element (see Definition 20 below) and employ the boundness of  $[\bar{\alpha}_{R,P}, 1] \times \mathcal{S}_{P,W}(\mathcal{H})$  to conclude our claim.

**Definition 20** (Closed Saddle-Element [74]). We denote by  $\text{ri}$  and  $\text{cl}$  the relative interior and the closure of a set, respectively. Let  $\mathcal{A}, \mathcal{B}$  be subsets of a real vector space, and  $F : \mathcal{A} \times \mathcal{B} \rightarrow \mathbb{R} \cup \{\pm\infty\}$ . The triple  $(\mathcal{A}, \mathcal{B}, F)$  is called a closed saddle-element if for any  $x \in \text{ri}(\mathcal{A})$  (resp.  $y \in \text{ri}(\mathcal{B})$ ),

- (i)  $\mathcal{B}$  (resp.  $\mathcal{A}$ ) is convex.
- (ii)  $F(x, \cdot)$  (resp.  $F(\cdot, y)$ ) is convex (resp. concave) and lower (resp. upper) semi-continuous.
- (iii) Any accumulation point of  $\mathcal{B}$  (resp.  $\mathcal{A}$ ) that does not belong to  $\mathcal{B}$  (resp.  $\mathcal{A}$ ), say  $y_o$  (resp.  $x_o$ ) satisfies  $\lim_{y \rightarrow y_o} F(x, y) = +\infty$  (resp.  $\lim_{x \rightarrow x_o} F(x, y) = -\infty$ ).

Fix an arbitrary  $\alpha \in \text{ri}([\bar{\alpha}_{R,P}, 1]) = (\bar{\alpha}_{R,P}, 1)$ . We check that  $(\mathcal{S}_{P,W}(\mathcal{H}), F_{R,P}(\alpha, \cdot))$  fulfills the three items in Definition 20. (i) The set  $\mathcal{S}_{P,W}(\mathcal{H})$  is clearly convex. (ii) Recall Lemma 1-(c) that  $\sigma \mapsto D_\alpha(W_x \parallel \sigma)$  is convex and lower semi-continuous. Since convex combination preserves the convexity and the lower semi-continuity, Eq. (E) yields that  $\sigma \mapsto F_{R,P}(\alpha, \sigma)$  is convex and lower semi-continuous on  $\mathcal{S}_{P,W}(\mathcal{H})$ . (iii) Due to the compactness of  $\mathcal{S}(\mathcal{H})$ , any accumulation point of  $\mathcal{S}_{P,W}(\mathcal{H})$  that does not belong to  $\mathcal{S}_{P,W}(\mathcal{H})$ , say  $\sigma_o$ , satisfies  $\sigma_o \in \mathcal{S}(\mathcal{H}) \setminus \mathcal{S}_{P,W}(\mathcal{H})$ . Eqs. (E) and (E) then show that  $F_{R,P}(\alpha, \sigma_o) = +\infty$ .

Next, fix an arbitrary  $\sigma \in \text{ri}(\mathcal{S}_{P,W}(\mathcal{H}))$ . Owing to the convexity of  $\mathcal{S}_{P,W}(\mathcal{H})$ , it follows that  $\text{ri}(\mathcal{S}_{P,W}(\mathcal{H})) = \text{ri}(\text{cl}(\mathcal{S}_{P,W}(\mathcal{H})))$  (see e.g. [75, Theorem 6.3]). We first claim  $\text{cl}(\mathcal{S}_{P,W}(\mathcal{H})) = \mathcal{S}(\mathcal{H})$ . To see this, observe that  $\mathcal{S}_{>0}(\mathcal{H}) \subseteq \mathcal{S}_{P,W}(\mathcal{H})$  since a full-rank density operator is not orthogonal with every  $W_x$ ,  $x \in \mathcal{X}$ . Hence,

$$\mathcal{S}(\mathcal{H}) = \text{cl}(\mathcal{S}_{>0}(\mathcal{H})) \subseteq \text{cl}(\mathcal{S}_{P,W}(\mathcal{H})). \quad (138)$$

On the other hand, the fact  $\mathcal{S}_{P,W}(\mathcal{H}) \subseteq \mathcal{S}(\mathcal{H})$  leads to

$$\text{cl}(\mathcal{S}_{P,W}(\mathcal{H})) \subseteq \text{cl}(\mathcal{S}(\mathcal{H})) = \mathcal{S}(\mathcal{H}). \quad (139)$$

By Eqs. (E) and (E), we deduce that

$$\text{ri}(\mathcal{S}_{P,W}(\mathcal{H})) = \text{ri}(\text{cl}(\mathcal{S}_{P,W}(\mathcal{H}))) = \text{ri}(\mathcal{S}(\mathcal{H})) = \mathcal{S}_{>0}(\mathcal{H}), \quad (140)$$

where the last equality in Eq. (E) follows from [76, Proposition 2.9]. Hence, we obtain

$$\forall \sigma \in \text{ri}(\mathcal{S}_{P,W}(\mathcal{H})) \quad \text{and} \quad \forall x \in \mathcal{X}, \quad \sigma \gg W_x. \quad (141)$$

Now we verify that  $([\bar{\alpha}_{R,P}, 1], F_{R,P}(\cdot, \sigma))$  satisfies the three items in Definition 20. Fix an arbitrary  $\sigma \in \text{ri}(\mathcal{S}_{P,W}(\mathcal{H}))$ . (i) The set  $(0, 1]$  is obviously convex. (ii) From Lemma 1-(a), the map  $\alpha \mapsto F_{R,P}(\alpha, \sigma)$  is continuous on  $(0, 1)$ . Further, it is not hard to verify that  $F_{R,P}(1, \sigma) = 0 = \lim_{\alpha \uparrow 1} F_{R,P}(\alpha, \sigma)$  from Eqs. (E), (E), and (2.1). Item (d) in Proposition 2 and [50, Corollary B.2] implies that  $\alpha \mapsto F_{R,P}(\alpha, \sigma)$  on  $[\bar{\alpha}_{R,P}, 1]$  is concave. Moreover, the continuity of  $\alpha \mapsto F_{R,P}(\alpha, \sigma)$  on  $[\bar{\alpha}_{R,P}, 1]$  guarantees the concavity of  $\alpha \mapsto F_{R,P}(\alpha, \sigma)$  on  $[\bar{\alpha}_{R,P}, 1]$ . (iii) Since  $[\bar{\alpha}_{R,P}, 1]$  is closed, there is no accumulation point of  $[\bar{\alpha}_{R,P}, 1]$  that does not belong to  $[\bar{\alpha}_{R,P}, 1]$ .

We are at the position to prove item (a) of Proposition 3. The closed saddle-element, along with the boundness of  $\mathcal{S}_{P,W}(\mathcal{H})$  and Rockafellar's saddle-point result [74, Theorem 8], [75, Theorem 37.3] imply that

$$-\infty < \sup_{\alpha \in [\bar{\alpha}_{R,P}, 1]} \inf_{\sigma \in \mathcal{S}_{P,W}(\mathcal{H})} F_{R,P}(\alpha, \sigma) = \min_{\sigma \in \mathcal{S}_{P,W}(\mathcal{H})} \sup_{\alpha \in [\bar{\alpha}_{R,P}, 1]} F_{R,P}(\alpha, \sigma). \quad (142)$$

Then Eqs. (E) and (E) lead to the existence of a saddle-point of  $F_{R,P}(\cdot, \cdot)$  on  $(0, 1] \times \mathcal{S}_{P,W}(\mathcal{H})$ . Hence, item (a) is proved.  $\square$

*Proof of Proposition 3-(b).* Fix arbitrary  $R \in (C_{0,W}, C_W)$  and  $P \in \mathcal{P}_R(\mathcal{X})$ . We have

$$\sup_{0 < \alpha \leq 1} \min_{\sigma \in \mathcal{S}(\mathcal{H})} F_{R,P}(\alpha, \sigma) = \min_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{0 < \alpha \leq 1} F_{R,P}(\alpha, \sigma) \in \mathbb{R}_{>0} \quad (143)$$

by the saddle-point property in item (a) and the definition of  $\mathcal{P}_R(\mathcal{X})$  given in Eq. (E). First note that  $(1, \sigma)$  for any  $\sigma \in \mathcal{S}(\mathcal{H})$  will not be a saddle point of  $F_{R,P}(\cdot, \cdot)$  because  $F_{R,P}(1, \sigma) = 0$ ,  $\forall \sigma \in \mathcal{S}(\mathcal{H})$ , contradicting Eq. (E).

Next, we assume  $(\alpha^*, \sigma^*)$  is a saddle-point of  $F_{R,P}(\cdot, \cdot)$  with  $\alpha^* \in (0, 1)$ , it holds that

$$F_{R,P}(\alpha^*, \sigma^*) = \min_{\sigma \in \mathcal{S}(\mathcal{H})} F_{R,P}(\alpha^*, \sigma) = \frac{\alpha^* - 1}{\alpha^*} R + \frac{1 - \alpha^*}{\alpha^*} \min_{\sigma \in \mathcal{S}(\mathcal{H})} D_{\alpha^*}(\mathcal{W} \parallel \sigma | P). \quad (144)$$

Since  $\sigma^*$  is the minimizer of  $\min_{\sigma \in \mathcal{S}(\mathcal{H})} D_{\alpha^*}(\mathcal{W} \parallel \sigma | P)$ , it is clear from Proposition 2-(b) that

$$\sigma^* \gg W_x, \quad \forall x \in \text{supp}(P),$$

and thus item (b) is proved.  $\square$

*Proof of Proposition 3-(c).* Continuing from item (b), we show the uniqueness of the saddle-point. Since  $(1, \sigma)$  for  $\sigma \in \mathcal{S}(\mathcal{H})$  will not be a saddle-point of  $F_{R,P}(\cdot, \cdot)$  as shown in item (b), we let  $\alpha^* \in (0, 1)$  attain the supremum in the left-hand side of Eq. (E). Proposition 2-(b) implies that the minimizer to the map  $\sigma \mapsto D_{\alpha^*}(\mathcal{W} \parallel \sigma | P)$  is unique, and thus it follows that the minimizer of Eq. (E) is unique as well.

Next, we will invoke Lemma 16 in Appendix A to show the uniqueness of the maximizer. Let  $\sigma^* \in \mathcal{S}(\mathcal{H})$  be the minimizer of right-hand side of the equality in Eq. (E), and let  $\mathbf{x}^n \in \mathcal{X}^n$  be an arbitrary sequence with an empirical distribution  $P$ . Denote by  $p^n, q^n$  be two distributions with  $(p_i, q_i)$  being the Nussbaum-Szkoła mapping of  $(W_{x_i}, \sigma^*)$ , where  $x_i$  is the  $i$ -th symbol of  $\mathbf{x}^n$  for  $i \in [n]$ . Further, item (b) guarantees that  $p^n \ll q^n$ .

Now, we let  $p^n$  and  $q^n$  to be the hypotheses described in Eq. (A). It is not hard to observe that  $\sup_{0 < \alpha \leq 1} F_{R,P}(\alpha, \sigma^*) = \phi_n(R)$  given in Eq. (A). Items (b) and (d) in Lemma 16 then show that the optimizer  $\alpha^* \in (0, 1)$  of  $\sup_{0 < \alpha \leq 1} F_{R,P}(\alpha, \sigma^*)$  is unique, which completes the proof of item (c).  $\square$

*Proof of Proposition 3-(d).* We first prove the joint continuity of  $(r, P) \mapsto \alpha_{r,P}^*$  on  $[\underline{R}, R] \times \mathcal{P}(\mathcal{X})$ . To that end, it suffices to show that  $P \mapsto \alpha_{r,P}^*$  is continuous on  $\mathcal{P}(\mathcal{X})$  for every  $r \in [\underline{R}, R]$ , and the family  $\{\alpha_{r,P}^*\}_{P \in \mathcal{P}(\mathcal{X})}$  is uniformly equicontinuous in  $r$  on  $[\underline{R}, R]$ . Moreover, it is equivalent to prove the joint continuity of  $(r, P) \mapsto s_{r,P}^*$  on  $[\underline{R}, R] \times \mathcal{P}(\mathcal{X})$  by using the substitution  $s_{r,P}^* := (1 - \alpha_{r,P}^*)/\alpha_{r,P}^*$ . This will ease the burden of notation.

In the following, we show the continuity of  $P \mapsto s_{r,P}^*$ . The proof idea of such continuity is similar to [35, Proposition 3.4]. Fix  $r \in [\underline{R}, R]$ , any  $P_0 \in \mathcal{P}(\mathcal{X})$  and consider arbitrary  $\{P_k\}_{k \in \mathbb{N}}$  such that  $P_k \in \mathcal{P}(\mathcal{X})$  for all  $k \in \mathbb{N}$ , and  $\lim_{n \rightarrow +\infty} P_k = P_0$ . Following from Proposition 4-(c) that will be proved later in Appendix F, we have<sup>11</sup>

$$s_{r,P_k}^* = -\frac{\partial E_{\text{sp}}^{(2)}(r, P_k)}{\partial r} \in \mathbb{R}_{\geq 0}. \quad (145)$$

Since  $r \geq \underline{R} > C_{0,W}$ , the continuity of  $E_{\text{sp}}^{(2)}(r, \cdot)$  given in Proposition 4-(a) that will be proved later shows that

$$\lim_{k \rightarrow +\infty} E_{\text{sp}}^{(2)}(r, P_k) = E_{\text{sp}}^{(2)}(r, P_0).$$

Viewing  $(E_{\text{sp}}^{(2)}(r, P_k))_{k \in \mathbb{N}}$  as a sequence of functions that converges to  $E_{\text{sp}}^{(2)}(r, P_0)$ , Ref. [82, Corollary VI.6.2.8] proved that the sequence of first-order derivatives of differentiable convex functions converges

<sup>11</sup>Here, for  $E_{\text{sp}}^{(2)}(r, P) = 0$ , we adopt  $(1, PW)$  as the saddle-point in Eq. (3), which means  $s_{r,P}^* = 0$ .

to the first-order derivative of the limit. Indeed, Proposition 4-(a) guarantees that  $E_{\text{sp}}^{(2)}(\cdot, P)$  is convex. Therefore,

$$\lim_{k \rightarrow +\infty} s_{R, P_k}^* = \lim_{k \rightarrow +\infty} - \left. \frac{\partial E_{\text{sp}}^{(2)}(r, P_k)}{\partial r} \right|_{r=R} = - \left. \frac{\partial E_{\text{sp}}^{(2)}(r, P_0)}{\partial r} \right|_{r=R} = s_{R, P_0}^*,$$

which shows the continuity of  $P \mapsto s_{r, P}^*$  for every  $r \in [\underline{R}, R]$ .

Next, we prove the equicontinuity. Let  $R_1, R_2 \in [\underline{R}, R]$  be arbitrary. As will be shown later in Proposition 4-(a), for every  $P \in \mathcal{P}(\mathcal{X})$ ,  $E_{\text{sp}}^{(2)}(\cdot, P)$  is convex and non-increasing on  $[0, +\infty]$ . Using Eq. (E), the absolute value of the difference between the first-order derivative of  $E_{\text{sp}}^{(2)}(\cdot, P)$  at  $R_1$  and  $R_2$  can be calculated as follows

$$|s_{R_1, P}^* - s_{R_2, P}^*| \leq s_{R_1, P}^* \vee s_{R_2, P}^* = s_{R_1 \wedge R_2, P}^* \leq s_{\underline{R}, P}^*, \quad (146)$$

where  $s_{\underline{R}, P}^* = (1 - \alpha_{\underline{R}, P}^*)/\alpha_{\underline{R}, P}^*$  and  $\alpha_{\underline{R}, P}^*$  is the optimizer of  $E_{\text{sp}}^{(2)}(\underline{R}, P)$  given in Eq. (2.1).

For all  $P \in \mathcal{P}(\mathcal{X})$  such that  $\underline{R} \geq I_1(P, \mathcal{W})$ , the right-hand side of Eq. (E) is zero since  $E_{\text{sp}}^{(2)}(\underline{R}, P) = 0$  (see Proposition 4-(a) again). On the other hand, for all  $P \in \mathcal{P}(\mathcal{X})$  such that  $\underline{R} < I_1(P, \mathcal{W})$ , Proposition 4-(c) shows that

$$I_{\alpha_{\underline{R}, P}^*}^{(2)}(P, \mathcal{W}) > \underline{R}. \quad (147)$$

Further, since  $\underline{R} \in (C_{0, \mathcal{W}}, C_{1, \mathcal{W}})$ , the continuous monotone increase of the map  $\alpha \mapsto C_{\alpha, \mathcal{W}}$  proved in Proposition 2-(h) guarantees that there exists a  $\alpha_{\underline{R}} \in (0, 1)$  such that

$$C_{\alpha_{\underline{R}}} = \underline{R}. \quad (148)$$

Then, from Eqs. (E), (E), and the definition of the Rényi information radius given in Eq. (2.1), we have

$$I_{\alpha_{\underline{R}}}^{(2)}(P, \mathcal{W}) \leq C_{\underline{R}, \mathcal{W}} = \underline{R} < I_{\alpha_{\underline{R}, P}^*}^{(2)}(P, \mathcal{W}), \quad (149)$$

The above inequality (E) and the monotone increases of the map  $\alpha \mapsto I_{\alpha}^{(2)}(P, \mathcal{W})$  further imply that

$$\alpha_{\underline{R}} < \alpha_{\underline{R}, P}^*. \quad (150)$$

Both Eqs. (E) and (E) then yield

$$|s_{R_1, P}^* - s_{R_2, P}^*| \leq \frac{1 - \alpha_{\underline{R}}}{\alpha_{\underline{R}}} < \infty$$

for all  $P \in \mathcal{P}(\mathcal{X})$  such that  $\underline{R} < I_1(P, \mathcal{W})$ . This shows the equicontinuity of the family  $\{\alpha_{r, P}^*\}_{P \in \mathcal{P}(\mathcal{X})}$  on  $[\underline{R}, R]$ . Together with the continuity of  $P \mapsto s_{r, P}^*$  for all  $r \in [\underline{R}, R]$ , the joint continuity of  $(r, P) \mapsto s_{r, P}^*$  on  $[\underline{R}, R] \times \mathcal{P}(\mathcal{X})$  is proved.

Lastly, we move on to prove the continuity of  $(r, P) \mapsto \sigma_{r, P}^*$  on  $[\underline{R}, R] \times \mathcal{P}(\mathcal{X})$ . Let  $(R_k, P_k) \in [\underline{R}, R] \times \mathcal{P}(\mathcal{X})$  for all  $k \in \mathbb{N}$  be arbitrary such that  $\lim_{k \in \mathbb{N}} (R_k, P_k) = (R_0, P_0) \in [\underline{R}, R] \times \mathcal{P}(\mathcal{X})$ . From Eq. (E), the saddle-point property yields that

$$\sigma_{R_k, P_k}^* = \sigma_{\alpha_{R_k, P_k}^*, P_k}^*, \quad (151)$$

where in the right-hand side of the above equality we denote the Augustin mean by  $\sigma_{\alpha, P} := \min_{\sigma \in \mathcal{S}(\mathcal{H})} D_{\alpha}(\mathcal{W} \| \sigma | P)$ . Moreover, Proposition 2-(g) states that  $(\alpha, P) \mapsto \sigma_{\alpha, P}$  is jointly continuous on  $(0, 1] \times \mathcal{P}(\mathcal{X})$ . Hence, the joint continuity of  $(r, P) \mapsto \alpha_{r, P}^*$  proved above together with Eq. (E) show that

$$\begin{aligned} \lim_{k \rightarrow +\infty} \sigma_{R_k, P_k}^* &= \lim_{k \rightarrow +\infty} \sigma_{\alpha_{R_k, P_k}^*, P_k}^* \\ &= \sigma_{\lim_{k \rightarrow +\infty} \alpha_{R_k, P_k}^*, \lim_{k \rightarrow +\infty} P_k}^* \\ &= \sigma_{\alpha_{\lim_{k \rightarrow +\infty} (R_k, P_k)}, \lim_{k \rightarrow +\infty} P_k}^* \\ &= \sigma_{\alpha_{R_0, P_0}^*, P_0}^* \\ &= \sigma_{R_0, P_0}^*, \end{aligned}$$

which completes the proof of item (d).  $\square$

#### APPENDIX F. PROOF OF PROPOSITION 4

**Proposition 4** (Properties of Error-Exponent Functions). *Consider a classical-quantum channel  $\mathcal{W} : \mathcal{X} \rightarrow \mathcal{S}(\mathcal{H})$  with  $C_{0,\mathcal{W}} < C_{\mathcal{W}}$ . We have*

- (a) *Given every  $P \in \mathcal{P}(\mathcal{X})$ ,  $E_{\text{sp}}^{(2)}(\cdot, P)$  is convex and non-increasing on  $[0, +\infty]$ , and continuous on  $[I_0^{(2)}(P, \mathcal{W}), +\infty]$ . For every  $R > C_{0,\mathcal{W}}$ ,  $E_{\text{sp}}^{(2)}(R, \cdot)$  is continuous on  $\mathcal{P}(\mathcal{X})$ . Further,*

$$E_{\text{sp}}^{(2)}(R, P) = \begin{cases} +\infty, & R < I_0^{(2)}(P, \mathcal{W}) \\ 0, & R \geq I_1^{(2)}(P, \mathcal{W}) \end{cases}.$$

- (b)  *$E_{\text{sp}}(\cdot)$  is convex and non-increasing on  $[0, +\infty]$ , and continuous on  $[C_{0,\mathcal{W}}, +\infty]$ . Further,*

$$E_{\text{sp}}(R) = \begin{cases} +\infty, & R < C_{0,\mathcal{W}} \\ 0, & R \geq C_{1,\mathcal{W}} \end{cases}.$$

- (c) *Consider any  $R \in (C_{0,\mathcal{W}}, C_{\mathcal{W}})$  and  $P \in \mathcal{P}_R(\mathcal{X})$  (see Eq. (3)). The function  $E_{\text{sp}}^{(2)}(\cdot, P)$  is differentiable with*

$$s_{R,P}^* := \frac{1 - \alpha_{R,P}^*}{\alpha_{R,P}^*} = - \left. \frac{\partial E_{\text{sp}}^{(2)}(r, P)}{\partial r} \right|_{r=R} \in \mathbb{R}_{>0}, \quad (152)$$

where  $\alpha_{R,P}^*$  is the optimizer in Eq. (2.1). Moreover,

$$I_{\alpha_{R,P}^*}(P, \mathcal{W}) > R. \quad (153)$$

*Proof of Proposition 4-(a).* Fix any arbitrary  $P \in \mathcal{P}(\mathcal{X})$ . Item (a) in Proposition 2 shows that the map  $\alpha \mapsto I_{\alpha}^{(2)}(P, \mathcal{W})$  is monotone increasing on  $[0, 1]$ . Hence, from the definition in Eq. (2.1), it is not hard to verify that  $E_{\text{sp}}^{(2)}(R, P) = +\infty$  for all  $R \in (0, I_0^{(2)}(P, \mathcal{W}))$ ; finite for all  $R > I_0^{(2)}(P, \mathcal{W})$ ; and  $E_{\text{sp}}^{(2)}(R, P) = 0$ , for all  $R \geq I_1^{(2)}(P, \mathcal{W})$ .

For every  $\alpha \in (0, 1]$ , the function  $\frac{1-\alpha}{\alpha}(I_{\alpha}^{(2)}(P, \mathcal{W}) - R)$  in Eq. (2.1) is an non-increasing, convex, and continuous function in  $R \in \mathbb{R}_{>0}$ . Since  $E_{\text{sp}}^{(2)}(R, P)$  is the pointwise supremum of the above function,  $E_{\text{sp}}^{(2)}(R, P)$  is non-increasing, convex, and lower semi-continuous function for all  $R \geq 0$ . Furthermore, since a convex function is continuous on the interior of the interval if it is finite [81, Corollary 6.3.3], thus  $E_{\text{sp}}^{(2)}(R, P)$  is continuous for all  $R > I_0^{(2)}(P, \mathcal{W})$ , and continuous from the right at  $R = I_0^{(2)}(P, \mathcal{W})$ .

To establish the continuity of  $E_{\text{sp}}^{(2)}(R, P)$  in  $P \in \mathcal{P}(\mathcal{X})$ , we first claim that there exists some  $\bar{\alpha}_R \in (0, 1]$  such that for every  $P \in \mathcal{P}(\mathcal{X})$ ,

$$\sup_{\alpha \in (0, 1]} \frac{1-\alpha}{\alpha} (I_{\alpha}^{(2)}(P, \mathcal{W}) - R) = \sup_{\alpha \in [\bar{\alpha}_R, 1]} \frac{1-\alpha}{\alpha} (I_{\alpha}^{(2)}(P, \mathcal{W}) - R). \quad (154)$$

Recall that  $R > C_{0,\mathcal{W}} = \max_{P \in \mathcal{P}(\mathcal{X})} I_0^{(2)}(P, \mathcal{W})$ . The continuity, item (h) in Proposition 2, implies that there exists an  $\bar{\alpha}_R > 0$  such that

$$R \geq I_{\bar{\alpha}_R}^{(2)}(P, \mathcal{W}), \quad \forall P \in \mathcal{P}(\mathcal{X}). \quad (155)$$

Then, Eq. (F) and the monotone increases of the map  $\alpha \mapsto I_{\alpha}^{(2)}(P, \mathcal{W})$  yield that,

$$\frac{1-\alpha}{\alpha} (I_{\alpha}^{(2)}(P, \mathcal{W}) - R) < 0, \quad \forall P \in \mathcal{P}(\mathcal{X}), \text{ and } \alpha \in (0, \bar{\alpha}_R).$$

The non-negativity of  $E_{\text{sp}}^{(2)}(R, P) \geq 0$  ensures that the maximizer  $\alpha^*$  will not happen in the region  $(0, \bar{\alpha}_R)$ , and thus Eq. (F) is evident. Finally, Berge's maximum theorem [72, Section IV.3], [73, Lemma 3.1], the



compactness of  $[\bar{\alpha}_R, 1]$ , and item (f) in Proposition 2 complete our claim:

$$P \mapsto E_{\text{sp}}^{(2)}(R, P) = \sup_{\alpha \in [\bar{\alpha}_R, 1]} \frac{1 - \alpha}{\alpha} \left( I_{\alpha}^{(2)}(P, \mathcal{W}) - R \right) \text{ is continuous on } \mathcal{P}(\mathcal{X}).$$

□

*Proof of Proposition 4-(b).* The statement follows since item (a) holds for any  $P \in \mathcal{P}(\mathcal{X})$  and we invoke the definition of  $C_{\alpha, \mathcal{W}}$  in Eq. (2.1). □

*Proof of Proposition 4-(c).* For any  $R \in (C_{0, \mathcal{W}}, C_{\mathcal{W}})$  and  $P \in \mathcal{P}_R(\mathcal{X})$ , item (c) in Proposition 3 shows that the optimizer  $\alpha_{R, P}^*$  is unique. Eq. ((c)) directly follows from item (d) in Lemma 16.

The saddle-point property in Proposition 3-(a) shows that

$$E_{\text{sp}}^{(2)}(R, P) = \frac{1 - \alpha_{R, P}^*}{\alpha_{R, P}^*} \left( I_{\alpha_{R, P}^*}^{(2)}(P, \mathcal{W}) - R \right).$$

Further, since  $E_{\text{sp}}^{(2)}(R, P) > 0$  and  $\alpha_{R, P}^* \in (0, 1)$  for  $P \in \mathcal{P}_R(\mathcal{X})$ , the above equality implies Eq. ((c)). □

## REFERENCES

- [1] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.
- [2] —, “Probability of error for optimal codes in a Gaussian channel,” *Bell System Technical Journal*, vol. 38, no. 3, pp. 611–656, may 1959.
- [3] A. Feinstein, “Error bounds in noisy channels without memory,” *IEEE Transactions on Information Theory*, vol. 1, no. 2, pp. 13–14, sep 1955.
- [4] R. M. Fano, *Transmission of Information, A Statistical Theory of Communications*. The MIT Press, 1961.
- [5] R. Gallager, “A simple derivation of the coding theorem and some applications,” *IEEE Transaction on Information Theory*, vol. 11, no. 1, pp. 3–18, jan 1965.
- [6] —, *Information Theory and Reliable Communication*. Wiley, 1968. [Online]. Available: <http://as.wiley.com/WileyCDA/WileyTitle/productCd-0471290483.html>
- [7] C. Shannon, R. Gallager, and E. Berlekamp, “Lower bounds to error probability for coding on discrete memoryless channels. I,” *Information and Control*, vol. 10, no. 1, pp. 65–103, jan 1967.
- [8] E. A. Haroutunian, “Estimates of the error exponents for the semicontinuous memoryless channel,” *Problemy Peredachi Informatsii*, vol. 4, no. 4, pp. 37–48, 1968, (in Russian). [Online]. Available: <http://mi.mathnet.ru/eng/ppi1871>
- [9] E. A. Haroutunian, M. E. Haroutunian, and A. N. Harutyunyan, “Reliability criteria in information theory and in statistical hypothesis testing,” *Foundations and Trends® in Communications and Information Theory*, vol. 4, no. 2–3, pp. 97–263, 2007.
- [10] R. E. Blahut, “Hypothesis testing and information theory,” *IEEE Transaction on Information Theory*, vol. 20, no. 4, pp. 405–417, jul 1974.
- [11] A. Ben-Tal, M. Teboulle, and A. Charnes, “The role of duality in optimization problems involving entropy functionals with applications to information theory,” *Journal of Optimization Theory and Applications*, vol. 58, no. 2, pp. 209–223, aug 1988.
- [12] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press (CUP), 2011.
- [13] M. V. Burnashev and A. S. Holevo, “On the reliability function for a quantum communication channel,” *Problems of information transmission*, vol. 34, no. 2, pp. 97–107, 1998.
- [14] A. Holevo, “Reliability function of general classical-quantum channel,” *IEEE Transaction on Information Theory*, vol. 46, no. 6, pp. 2256–2261, 2000.
- [15] H.-C. Cheng and M.-H. Hsieh, “Concavity of the auxiliary function for classical-quantum channels,” *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5960 – 5965, 2016.



- [16] A. Winter, “Coding theorems of quantum information theory,” *PhD Thesis, Universität Bielefeld*, 1999.
- [17] M. Dalai, “Lower bounds on the probability of error for classical and classical-quantum channels,” *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8027–8056, dec 2013.
- [18] M. Dalai and A. Winter, “Constant Compositions in the Sphere Packing Bound for Classical-Quantum Channels,” *IEEE Transactions on Information Theory*, vol. 63, no. 9, Sept 2017.
- [19] D. Petz, “Quasi-entropies for finite quantum systems,” *Reports on Mathematical Physics*, vol. 23, no. 1, pp. 57–65, feb 1986.
- [20] S. Golden, “Lower bounds for the Helmholtz function,” *Physical Review*, vol. 137, no. 4B, pp. B1127–B1128, feb 1965.
- [21] C. J. Thompson, “Inequality with applications in statistical mechanics,” *Journal of Mathematical Physics*, vol. 6, no. 11, p. 1812, 1965.
- [22] Y. Altuğ and A. B. Wagner, “A refinement of the random coding bound,” in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. Institute of Electrical and Electronics Engineers (IEEE), oct 2012.
- [23] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, “Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations,” *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2647–2666, may 2014.
- [24] J. Scarlett, “Reliable communication under mismatched decoding,” *PhD Thesis (University of Cambridge)*, 2014.
- [25] J. Honda, “Exact asymptotics for the random coding error probability,” [arXiv:1506.03355 \[cs.IT\]](#).
- [26] R. R. Bahadur and R. R. Rao, “On deviations of the sample mean,” *The Annals of Mathematical Statistics*, vol. 31, no. 4, pp. 1015–1027, dec 1960.
- [27] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer, 1998.
- [28] Y. Altuğ and A. B. Wagner, “Moderate deviations in channel coding,” *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4417–4426, aug 2014.
- [29] H.-C. Cheng and M.-H. Hsieh, “Moderate Deviation Analysis for Classical-Quantum Channels and Quantum Hypothesis Testing,” *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 1385–1403, feb 2018.
- [30] C. T. Chubb, V. Y. F. Tan, and M. Tomamichel, “Moderate Deviation Analysis for Classical Communication over Quantum Channels,” *Communications in Mathematical Physics*, vol. 355, no. 3, pp. 1283–1315, nov 2017.
- [31] H.-C. Cheng, E. P. Hanson, N. Datta, and M.-H. Hsieh, “Non-Asymptotic Classical Data Compression with Quantum Side Information,” [arXiv:1803.07505 \[quant-ph\]](#).
- [32] H.-C. Cheng, E. P. Hanson, N. Datta, and M.-H. Hsieh, “Duality between source coding with quantum side information and c-q channel coding,” [arXiv:1809.11143 \[quant-ph\]](#).
- [33] H.-C. Cheng, E. P. Hanson, N. Datta, and M.-H. Hsieh, “Non-Asymptotic Joint Source-Channel Coding with Quantum Side Information,” (in preparation).
- [34] J. K. Omura, “A lower bounding method for channel and source coding probabilities,” *Information and Control*, vol. 27, no. 2, pp. 148–177, feb 1975.
- [35] Y. Altuğ and A. B. Wagner, “Refinement of the sphere-packing bound: Asymmetric channels,” *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1592–1614, mar 2014.
- [36] N. Elkayam and M. Feder, “Sphere packing bound for constant composition,” 2016, (in preparation).
- [37] U. Augustin, “Error estimates for low rate codes,” *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 14, no. 1, pp. 61–88, 1969.
- [38] —, “Noisy channels,” 1978, habilitation thesis, Universität Erlangen (<http://bit.ly/2ID8h7m>).
- [39] B. Nakiboğlu, “The Renyi Capacity and Center,” *IEEE Transactions on Information Theory*, vol. 65, no. 2, pp. 841–860, feb 2019.
- [40] B. Nakiboğlu, “The Sphere Packing Bound via Augustin’s Method,” *IEEE Transactions on Information Theory*, vol. 65, no. 2, pp. 816–840, feb 2019.
- [41] B. Nakiboğlu, “The The Augustin Capacity and Center,” [arXiv:1803.07937 \[cs.IT\]](#).
- [42] B. Nakiboğlu, “The Sphere Packing Bound For Memoryless Channels,” [arXiv:1804.06372 \[cs.IT\]](#).

- [43] H.-C. Cheng, L. Gao, and M.-H. Hsieh, “Properties of Noncommutative Rényi and Augustin Information,” [arXiv:1811.04218 \[quant-ph\]](#).
- [44] H. Umegaki, “Conditional expectation in an operator algebra. IV. entropy and information,” *Kodai Mathematical Seminar Reports*, vol. 14, no. 2, pp. 59–85, 1962.
- [45] F. Hiai and D. Petz, “The proper formula for relative entropy and its asymptotics in quantum probability,” *Communications in Mathematical Physics*, vol. 143, no. 1, pp. 99–114, dec 1991.
- [46] M. Tomamichel and M. Hayashi, “A hierarchy of information quantities for finite block length analysis of quantum tasks,” *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7693–7710, nov 2013.
- [47] K. Li, “Second-order asymptotics for quantum hypothesis testing,” *The Annals of Statistics*, vol. 42, no. 1, pp. 171–189, feb 2014.
- [48] M. Tomamichel and V. Y. F. Tan, “Second-order asymptotics for the classical capacity of image-additive quantum channels,” *Communications in Mathematical Physics*, vol. 338, no. 1, pp. 103–137, may 2015.
- [49] T. Ogawa and H. Nagaoka, “Strong converse and Stein’s lemma in quantum hypothesis testing,” *IEEE Transaction on Information Theory*, vol. 46, no. 7, pp. 2428–2433, 2000.
- [50] M. Mosonyi and T. Ogawa, “Strong Converse Exponent for Classical-Quantum Channel Coding,” *Communications in Mathematical Physics*, vol. 355, no. 1, pp. 373–426, Jun 2017.
- [51] T. Ando, “Concavity of certain maps on positive definite matrices and applications to Hadamard products,” *Linear Algebra and its Applications*, vol. 26, pp. 203–241, 1979.
- [52] E. H. Lieb, “Convex trace functions and the Wigner-Yanase-Dyson conjecture,” *Advances in Mathematics*, vol. 11, no. 3, pp. 267–288, dec 1973.
- [53] B. Schumacher and M. D. Westmoreland, “Optimal signal ensembles,” *Physical Review A*, vol. 63, no. 2, Jan 2001.
- [54] B. Schumacher and M. D. Westmoreland, “Sending classical information via noisy quantum channels,” *Physical Review A*, vol. 56, no. 1, pp. 131–138, jul 1997.
- [55] A. Holevo, “The capacity of the quantum channel with general signal states,” *IEEE Transaction on Information Theory*, vol. 44, no. 1, pp. 269–273, 1998.
- [56] M. Hayashi and M. Tomamichel, “Correlation detection and an operational interpretation of the rényi mutual information,” *Journal of Mathematical Physics*, vol. 57, no. 10, p. 102201, oct 2016.
- [57] M. M. Wilde, A. Winter, and D. Yang, “Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy,” *Communications in Mathematical Physics*, vol. 331, no. 2, pp. 593–622, jul 2014.
- [58] N. Sharma and N. A. Warsi, “Fundamental bound on the reliability of quantum information transmission,” *Physical Review Letters*, vol. 110, no. 8, feb 2013.
- [59] M. Hayashi, *Quantum Information: An Introduction*. Springer, 2006.
- [60] H. Nagaoka and M. Hayashi, “An information-spectrum approach to classical and quantum hypothesis testing for simple hypotheses,” *IEEE Transactions on Information Theory*, vol. 53, no. 2, pp. 534–549, feb 2007.
- [61] K. M. R. Audenaert, M. Nussbaum, A. Szkoła, and F. Verstraete, “Asymptotic error rates in quantum hypothesis testing,” *Communications in Mathematical Physics*, vol. 279, no. 1, pp. 251–283, feb 2008.
- [62] M. Nussbaum and A. Szkoła, “The Chernoff lower bound for symmetric quantum hypothesis testing,” *Annals of Statistics*, vol. 37, no. 2, pp. 1040–1057, apr 2009.
- [63] H. Nagaoka, “The converse part of the theorem for quantum Hoeffding bound,” [arXiv:quant-ph/0611289](#).
- [64] R. E. Blahut, *Principles and practice of information theory*. Addison-Wesley, 1987.
- [65] V. Y. F. Tan, “Asymptotic estimates in information theory with non-vanishing error probabilities,” *Foundations and Trends® in Communications and Information Theory*, vol. 10, no. 4, pp. 1–184, 2014.

- [66] Y. Altuğ and A. B. Wagner, “The third-order term in the normal approximation for singular channels,” in *2014 IEEE International Symposium on Information Theory*. Institute of Electrical and Electronics Engineers (IEEE), jun 2014.
- [67] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Channel coding rate in the finite blocklength regime,” *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2307–2359, may 2010.
- [68] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. Masanes, A. Acín, and F. Verstraete, “Discriminating states: The quantum Chernoff bound,” *Physical Review Letters*, vol. 98, p. 160501, apr 2007.
- [69] M. A. Nielsen and Issac L. Chuang, “Quantum Computation and Quantum Information” Cambridge University Press, 2009.
- [70] M. Tomamichel, *Quantum Information Processing with Finite Resources*. Springer International Publishing, 2016.
- [71] R. Bhatia, *Matrix Analysis*, Springer New York, 1997.
- [72] C. Berge, *Topological Spaces*. Oliver & Boyd, 1963.
- [73] B. Pshenichnyi, *Necessary Conditions for an Extremum*. CRC Press, 1971.
- [74] R. T. Rockafellar, “Minimax theorems and conjugate saddle-functions,” *Mathematica Scandinavica*, vol. 14, p. 151, jun 1964.
- [75] —, *Convex Analysis*. Walter de Gruyter GmbH, jan 1970.
- [76] S. Weis, “Quantum convex support,” *Linear Algebra and its Applications*, vol. 435, no. 12, pp. 3168–3188, dec 2011.
- [77] F. Hiai and D. Petz, *Introduction to Matrix Analysis and Applications*. Springer International Publishing, 2014.
- [78] H.-C. Cheng and M.-H. Hsieh, “New characterizations of matrix  $\Phi$ -entropies, Poincaré and Sobolev inequalities and an upper bound to Holevo quantity,” [arXiv:1506.06801 \[quant-ph\]](https://arxiv.org/abs/1506.06801).
- [79] H.-C. Cheng, M.-H. Hsieh, and Tomamichel, “Exponential decay of matrix  $\Phi$ -entropies on Markov semigroups with applications to dynamical evolutions of quantum ensembles,” *Journal of Mathematical Physics*, vol. 58, no. 9, p. 092202, sep 2017.
- [80] H.-C. Cheng and M.-H. Hsieh, “Characterizations of matrix and operator-valued  $\Phi$ -entropies, and operator EfronStein inequalities,” *Proceedings of the Royal Society of London A*, p. 20150563, 2016.
- [81] R. M. Dudley, *Real Analysis and Probability*. Cambridge University Press (CUP), 2002.
- [82] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*. Springer Nature, 2001.
- [83] A. S. Holevo, “Bounds for the quantity of information transmitted by a quantum communication channel,” *Problems of Information Transmission*, vol 9, no. 3, pp. 177–183, 1973.
- [84] A. N. Kolmogorov and S. V. Fomin, *Introductory Real Analysis*. Dover Publications, New York, 1975.
- [85] Hao-Chung Cheng, Li Gao, and Min-Hsiu Hsieh, “Properties of Scaled Noncommutative Rényi and Augustin Information”, [arXiv:1811.0421 \[quant-ph\]](https://arxiv.org/abs/1811.0421).
- [86] Hao-Chung Cheng, *Error Exponent Analysis in Quantum Information Theory*, PhD Thesis (University of Technology Sydney), 2018.