


# Simple and Tighter Derivation of Achievability for Classical Communication Over Quantum Channels

Hao-Chung Cheng (鄭皓中) <sup>1,2,3,4,5,\*</sup>


<sup>1</sup>Department of Electrical Engineering and Graduate Institute of Communication Engineering, National Taiwan University, Taiwan, Republic of China

<sup>2</sup>Department of Mathematics, National Taiwan University, Taiwan, Republic of China

<sup>3</sup>Center for Quantum Science and Engineering, National Taiwan University, Taiwan, Republic of China

<sup>4</sup>Physics Division, National Center for Theoretical Sciences, Taiwan, Republic of China

<sup>5</sup>Hon Hai (Foxconn) Quantum Computing Center, Taiwan, Republic of China

 (Received 28 April 2023; revised 20 September 2023; accepted 16 October 2023; published 22 November 2023)

Achievability in information theory refers to demonstrating a coding strategy that accomplishes a prescribed performance benchmark for the underlying task. In quantum information theory, the crafted Hayashi-Nagaoka operator inequality is an essential technique in proving a wealth of one-shot achievability bounds since it effectively resembles a union bound in various problems. In this work, we show that the so-called pretty-good measurement naturally plays a role as the union bound as well. A judicious application of it considerably simplifies the derivation of one-shot achievability for classical-quantum channel coding via an elegant three-line proof. The proposed analysis enjoys the following favorable features. (i) The established one-shot bound admits a closed-form expression as in the celebrated Holevo-Helstrom theorem. Namely, the average error probability of sending  $M$  messages through a classical-quantum channel is upper bounded by the minimum error of distinguishing the joint channel input-output state against  $(M - 1)$  decoupled product states. (ii) Our bound directly yields asymptotic achievability results in the large deviation, small deviation, and moderate deviation regimes in a unified manner. (iii) The coefficients incurred in applying the Hayashi-Nagaoka operator inequality or the quantum union bound are no longer needed. Hence, the derived one-shot bound sharpens existing results relying on the Hayashi-Nagaoka operator inequality. In particular, we obtain the tightest achievable  $\varepsilon$ -one-shot capacity for classical communication over quantum channels heretofore, improving the third-order coding rate in the asymptotic scenario. (iv) Our result holds for infinite-dimensional Hilbert space. (v) The proposed method applies to deriving one-shot achievability for classical data compression with quantum side information, entanglement-assisted classical communication over quantum channels, and various quantum network information-processing protocols.

DOI: [10.1103/PRXQuantum.4.040330](https://doi.org/10.1103/PRXQuantum.4.040330)

## I. INTRODUCTION

Communicating classical information over a noisy quantum channel is a foundational task in quantum information science. To protect the transmitted messages against potential noise, an indispensable coding strategy is employed. At the transmitter, Alice initiates the procedure by encoding each message  $m \in \{1, 2, \dots, M\}$  into an  $n$ -qubit quantum state. Suppose in the communication process that each qubit

suffers from independent and identically distributed (IID) quantum noise, which is characterized by an IID quantum channel. Then, at the receiver Bob performs a quantum measurement on the corrupted quantum system to extract the decoded message  $\hat{m}$ .

Via a coding strategy based on the so-called *quantum typicality*, the well-known Holevo-Schumacher-Westmoreland (HSW) theorem [1–7] states that the probability of erroneous decoding,  $\varepsilon := \Pr\{\hat{m} \neq m\}$ , vanishes asymptotically in the limit of  $n \rightarrow \infty$ , whenever the number of bits to be sent per qubit ( $\lim_{n \rightarrow \infty} 1/n \log M$ ) is below the *channel capacity*.

The HSW theorem extends the seminal work of Shannon [8] to the quantum scenario, and, hence, it is one of the fundamental core stones in quantum information theory. However, the HSW coding strategy relies on certain technical assumptions that could be physically demanding.

\*haochung.ch@gmail.com

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

First, the asymptotically large qubit number  $n$  requires the quantum devices to implement arbitrarily large encoding and decoding. Second, the actual quantum noise may be correlated among several qubit systems; hence, the underlying quantum noises are not independent. Third, even if the noises are independent, they may not be stationary; namely, the noise acting on the first qubit is not identical to that on the last qubit.

To circumvent the aforementioned technical requirements, *one-shot* quantum information theory emerges as a new research stream to consider the scenario that no structural hypotheses are imposed on the underlying quantum state or channel. The ultimate goal is to characterize the *optimal trade-off* between the error probability  $\varepsilon$  and the message size  $M$  of transmission, for which the channel is used only once. Such a study allows us to better understand the fundamental capability of one-shot communication. Therefore it may serve as a general guideline for designing the next-generation quantum information-processing systems. However, without the IID repetitions of channel use, conventional methods based on quantum typicality are no longer applicable. Hence, more refined and sophisticated coding techniques are requisite for the one-shot analysis [9].

Why is it challenging to design and analyze good coding strategies for a one-shot quantum information-processing task? Essentially, a proper coding scheme aims to enforce the error probability,  $\Pr\{\bigcup_{\hat{m} \neq m} \mathcal{E}_{\hat{m}|m}\}$ , for sending each message  $m$  small. Here,  $\mathcal{E}_{\hat{m}|m}$  denotes the event of decoding to  $\hat{m}$  when message  $m$  was sent. Yet, the analysis and computational evaluation of such a union error event for a nontrivial quantum measurement could be quite difficult (even in the classical scenario). A useful trick in this effort is the following *union bound*:

$$\Pr\left\{\bigcup_{\hat{m} \neq m} \mathcal{E}_{\hat{m}|m}\right\} \leq \sum_{\hat{m} \neq m} \Pr\{\mathcal{E}_{\hat{m}|m}\}. \quad (1)$$

In view of the right-hand side of Eq. (1), the decoding rule remains to minimize  $(M - 1)$  pairwise error probabilities of deciding  $m$  against each  $\hat{m} \neq m$ . This then serves as the general principle of coding design.

The above coding strategy has achieved prevailing success in classical information theory, channel coding, and modern communication systems [10–23]. However, the quantum union bound of the form (1) is highly nontrivial due to the noncommutative nature of quantum mechanics [24–29]. The first attempt to design a good one-shot coding scheme for general classical-quantum ( $c$ - $q$ ) channels (without the IID condition) was proposed by Hayashi and Nagaoka, in which a powerful operator inequality [30, Lemma 2] was proved: for any positive semidefinite

operators  $0 \leq A \leq \mathbb{1}$ ,  $B \geq 0$ ,

$$\mathbb{1} - \frac{A}{A+B} \leq (1+c)(I-A) + (2+c+c^{-1})B$$

for all  $c > 0$ , (2)

where we denote a *noncommutative quotient* by

$$\frac{A}{B} := B^{-1/2}AB^{-1/2} \quad (3)$$

(here, the inverse is defined only on the support of the operator in the denominator). At the first glimpse of Eq. (2), it is not obvious how it resembles the union bound as Eq. (1) and how it is applied in analyzing the error probability in channel coding; nonetheless, an ingenious application of it by Hayashi and Nagaoka [30] yields a Feinstein-type bound for achieving the  $c$ - $q$  channel capacity [16, Theorem 1], [31]. Later, Oskouei, Mancini, and Wilde proposed a quantum union bound with similar coefficients as in Eq. (2), and hence achieved the same error bound as Refs. [29,30,32]. Subsequently, Hayashi and Nagaoka’s analysis based on Eq. (2) lays a technical cornerstone in a wealth of one-shot and asymptotic achievability results in quantum information theory, wherein a quantum measurement for extracting classical information from a quantum system is needed. For example, letting the coefficient  $c$  in Eq. (2) be a fixed constant along with a *quantum Chernoff bound* [33–36] delivers a large deviation bound for  $c$ - $q$  channel coding [35]. Letting  $c = 1/\sqrt{n}$  for an  $n$ -fold IID repetition of a  $c$ - $q$  channel, Eq. (2) achieves the second-order coding rate [22,37–40] in the small deviation regime. Later, both results were extended to the moderate deviation regime [41,42] accordingly. In addition, Anshu, Jain, and Warsi proposed a *position-based coding* for achieving entanglement-assisted classical communication over quantum channels [43], which also relies on the Hayashi-Nagaoka operator inequality in Eq. (2).

Apart from the success and significance of Hayashi and Nagaoka’s approach, there are still conceptual and practical subtleties. First, the technically sophisticated proof of the operator inequality (2) may blind the insight of the analysis and hide the reason why such a coding strategy works. Does there exist a good quantum coding strategy that naturally reflects the union bound as in Eq. (1) so that the analysis is more interpretable? Second, is it possible to tighten the one-shot achievability bound for quantum information-theoretic tasks by eliminating the incurred coefficients in terms of  $c$  [44,45]? Removing those coefficients may seem superficial. However, we remark that every bit in an analytical bound counts in the one-shot setting; one cannot ignore any coefficient. On top of that, the unnecessary coefficients may often *trivialize* the  $(\varepsilon, M)$  trade-off. For instance, existing one-shot bounds on the error probability  $\varepsilon$  could trivially be greater than

1 for  $\log M$  close to the channel capacity. This analysis then provides no useful characterizations of certain system configurations for practical communication. Lastly, the Hayashi-Nagaoka decoder [30,46] involves solving a positive semidefinite program to obtain the mathematical description of quantum measurement, for which the computational complexity is exponential in the number of qubits. Moreover, a quantum algorithm for implementing the Hayashi-Nagaoka decoder is still missing.

In this paper, we give affirmative answers to the above concerns and questions by showing that the so-called *pretty good measurement* (PGM) [47,48] naturally plays a role as the union bound. Together with the random coding technique, it yields a one-shot achievability bound via a much simpler and self-explainable analysis, which is merely based on previously known facts. The coefficients mentioned above in terms of  $c$  are not required anymore, and, hence, the established one-shot bound is sharpened. Furthermore, the proof itself provides a more transparent connection between  $c$ - $q$  channel coding and binary quantum hypothesis testing.

To present our result, we first introduce a *noncommutative minimal* between two positive semidefinite operators  $A$  and  $B$  as [49]

$$A \wedge B := \frac{A + B - |A - B|}{2}.$$

This quantity is prominent in quantum state discrimination since the celebrated *Holevo-Helstrom theorem* [50–52] endowed it with an operational meaning [53]:

$\text{Tr}[A \wedge B]$  determines the minimum ‘error’ of discrimination between operators  $A$  and  $B$ .

The coding strategy for a  $c$ - $q$  channel  $x \mapsto \rho_B^x$  (which maps each classical symbol  $x$  to a density operator  $\rho_B^x$ ) proceeds as follows. The encoding is via a random codebook  $\{x(1), x(2), \dots, x(M)\}$ , in which each codeword  $x(m)$  is drawn pairwise independently according to an arbitrary probability distribution  $p_X$ . The decoding is via the PGM with respect to the corresponding channel output states [47,48]:

$$\left\{ \frac{\rho_B^{x(m)}}{\sum_{\bar{m}=1}^M \rho_B^{x(\bar{m})}} \right\}_{m=1}^M.$$

We show that the associated average error probability is upper bounded by (Theorem 1):

$$\text{Tr}[\rho_{XB} \wedge (M-1)\rho_X \otimes \rho_B] \quad (4)$$

with  $\rho_{XB} = \sum_{x \in X} p_X(x) |x\rangle\langle x| \otimes \rho_B^x$  the resulting joint bipartite state between the channel input and output. Via the Holevo-Helstrom theorem, the established bound in Eq. (4) provides us the following interpretation for sending  $M$  messages over a  $c$ - $q$  channel (see Sec. III for the detailed explanation):

The average error probability is upper bounded by the error of discriminating  $\rho_{XB}$  and  $(M-1)\rho_X \otimes \rho_B$ .

Below, let us elaborate on the intuition of the proposed coding strategy and why PGM works well. The key observation is that using the PGM to discriminate  $M$  states at the channel output is exactly equivalent to (an average of) binary discrimination between each channel output state, say, e.g.,  $\rho_B^{x(m)}$  against the remaining  $(M-1)$  states  $\rho_B^{x(\bar{m})}$  for all  $\bar{m} \neq m$  using a two-outcome PGM. In this regard, PGM works as a *one-versus-rest* classification strategy; see Fig. 1(a). Most importantly, this manifests the fact that PGM effectively resembles the *quantum union bound* as shown in the right-hand side of Eq. (1). By taking the conditional expectation  $\mathbb{E}_{x(\bar{m})|x(m)}$  over the random codebook, the remaining states are hence *averaged* to  $(M-1)$  identical marginal states  $\rho_B$ ; see Figure 1(b). After taking expectation  $\mathbb{E}_{x(m)}$ , the error bound is equivalent to discriminating the joint state  $\rho_{XB}$  between channel input and output against  $(M-1)$  product states  $\rho_X \otimes \rho_B$ , as shown in Fig. 1(c). This gives the elegant and clean bound in Eq. (4).

The proposed simple derivation enjoys the following favorable features.

- (I) The one-shot achievability bound in Eq. (4) admits a closed-form expression as the Holevo-Helstrom theorem. Computing such a bound is more time efficient than the previous results in terms of entropic quantities involving optimizations (see Remark 4 in Sec. III).
- (II) The proposed coding scheme based on the pretty-good measurement is directly implementable via the existing quantum algorithm by Gilyén *et al.* [54].
- (III) The self-explainable proof signifies a more lucid connection between  $c$ - $q$  channel coding and hypothesis testing. Moreover, our coding strategy and analysis show that PGM effectively works as a union bound by itself. Hence, neither the operator inequality (2) nor a quantum union bound is needed.
- (IV) The proposed bound in Eq. (4) is free of parameter  $c$ , as in Eq. (2). This then shows that the established one-shot achievable error bound is tighter than previously known results based on the Hayashi-Nagaoka operator inequality, Eq. (2); see Sec. III A and Table II therein for a comparison with existing results. Moreover, it unifies asymptotic derivations in the large, small, and moderate derivation regimes. We refer the reader to Fig. 2 for a schematic flow chart.
- (V) The proposed analysis applies to infinite-dimensional quantum systems [55] as well, e.g., communication over infinite-dimensional channels with energy constraints or cost constraints [56,57].
- (VI) The proposed methods via the pretty-good measurement naturally extend to various quantum

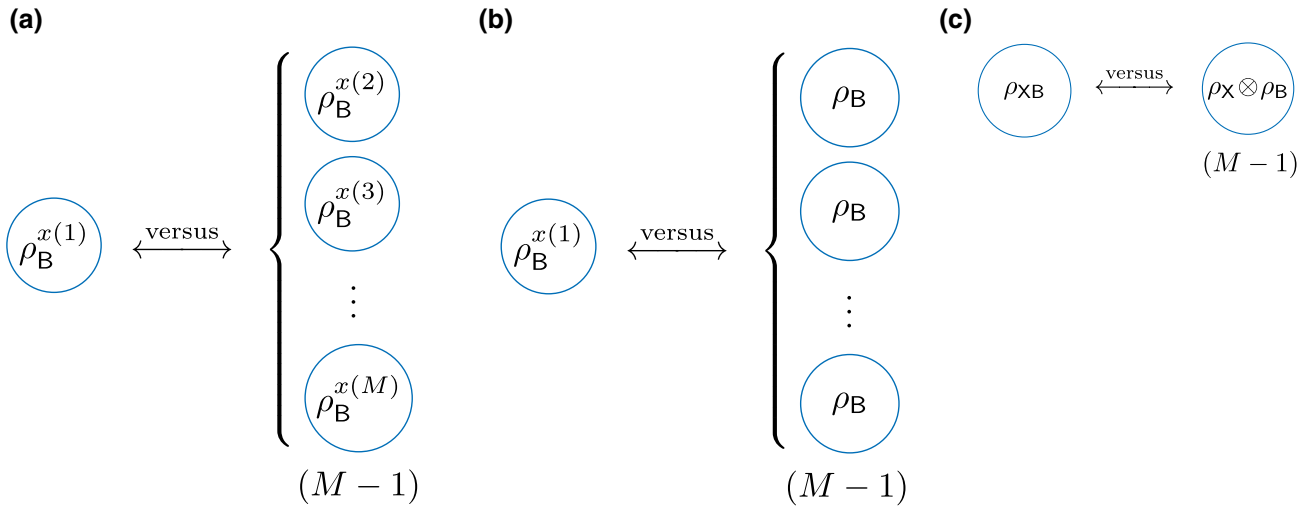


FIG. 1. Schematic illustration of the proposed achievability analysis for classical-quantum channel ( $x \mapsto \rho_B^x$ ) coding. (a) Given a realization of a codebook  $\mathcal{C} = \{x(1), \dots, x(M)\}$ , the error probability of sending message  $m = 1$  using the pretty-good measurement (PGM) is upper bounded by the error of distinguishing  $\rho_B^{x(1)}$  against the remaining states, i.e.,  $\sum_{\tilde{m}=2}^M \rho_B^{x(\tilde{m})}$ . We take the sum of the remaining channel output states because the PGM effectively works as a *one-versus-rest* classification strategy. (b) Taking the conditional expectation over the random codebook  $\mathcal{C}$  conditioned on codeword  $x(1)$ , the error probability of sending message  $m = 1$  is upper bounded by the error of distinguishing  $\rho_B^{x(1)}$  against  $(M - 1)\rho_B$ . Namely, by randomly drawing a codeword  $x(1) \sim p_X$ , we are distinguishing the associated channel output state against  $(M - 1)$  average channel output states. (c) Taking the expectation over the transmitted codeword  $x(1) \sim p_X$ , panel (b) is equivalent to distinguishing the joint state  $\rho_{XB}$  between channel input and output against the scaled product of its marginal states  $(M - 1)\rho_X \otimes \rho_B$ ; see Eq. (4). This may be viewed as a *one-shot packing lemma* for classical-quantum channel coding.

information-theoretic tasks, leading to more profound and sharpened results. These tasks include

- (i) binary quantum hypothesis testing (Sec. IV A),
- (ii) entanglement-assisted classical communication over point-to-point quantum channels (Sec. IV B),
- (iii) classical data compression with quantum side information (Sec. IV C),
- (iv) entanglement-assisted and unassisted classical communication over quantum multiple-access channels (Sec. IV D),
- (v) entanglement-assisted and unassisted classical communication over quantum broadcast channels (Sec. IV E), and
- (vi) entanglement-assisted and unassisted classical communication over quantum channels with casual state information available at the encoder (Sec. IV F).

We refer the reader to the summary given in Table I below.

Lastly, the established simple analysis applies to the *position-based coding*, a pivotal technique in one-shot quantum information theory (see, e.g., Refs. [43, 43, 58–60]), proposed by Anshu *et al.* [43, Lemma 4], whose decoding strategy again relies on the

Hayashi-Nagaoka operator inequality in Eq. (2). The sharpened position-based coding (Theorem 3 below) constitutes the primary technique of deriving numerous one-shot achievability bounds in Sec. IV. By virtue of its variability, we may term it as a *one-shot quantum packing lemma*, and it might lead to more fruitful applications elsewhere.

This paper is organized as follows. Section II formally introduces the noncommutative minimal and its properties. Section III establishes our main result of the one-shot achievability for  $c$ - $q$  channel coding; we compare it with existing results in Sec. III A. Section IV entails its applications in one-shot quantum information theory. We conclude the paper and discuss possible open problems in Section V. The Appendix proves a useful trace inequality regarding the noncommutative minimal.

## II. THE NONCOMMUTATIVE MINIMAL AND ITS PROPERTIES

We first recall the basic concepts of (binary) quantum state discrimination, which constitutes the central tool for the proposed achievability analysis in Sec. III below. Given arbitrary positive semidefinite operators  $A$  and  $B$ , we define the *minimum error* [61] using two-outcome positive operator-valued measures to distinguish them as

$$\inf_{0 \leq T \leq \mathbb{1}} \text{Tr}[A(\mathbb{1} - T)] + \text{Tr}[BT].$$

TABLE I. Summary of the established one-shot achievability bounds in various quantum information-theoretic tasks. The precise statements and notation can be found in Sec. IV below.

Information-theoretic tasks	One-shot achievability	Bounds on coding error
		Bounds on coding size
Point-to-point quantum channel	$\varepsilon \leq \text{Tr}[\mathcal{N}_{A \rightarrow B}(\theta_{XA}) \wedge (M-1)\theta_X \otimes \mathcal{N}_{A \rightarrow B}(\theta_A)]$	$\varepsilon \leq e^{-\sup_{\alpha \in (1/2, 1)} [(1-\alpha)/\alpha] (I_{2-1/\alpha}^{\downarrow}(\mathbf{X} \mathbf{B})_{\mathcal{N}_{A \rightarrow B}(\theta_{XA})})^{-\log M}}$
Entanglement-assisted point-to-point quantum channel	$\varepsilon \leq \text{Tr}[\mathcal{N}_{A \rightarrow B}(\theta_{RA}) \wedge (M-1)\theta_R \otimes \mathcal{N}_{A \rightarrow B}(\theta_A)]$	$\log M \geq I_h^{\varepsilon-\delta}(\mathbf{X} \mathbf{B})_{\mathcal{N}_{A \rightarrow B}(\theta_{XA})} - \log \frac{1}{\delta}$
Quantum channel with casual state information	$\varepsilon \leq \text{Tr}[\mathcal{N}_{AS \rightarrow B}(\theta_{UAS}) \wedge (M-1)\theta_U \otimes \mathcal{N}_{AS \rightarrow B}(\theta_{AS})]$ for all $\theta_{UAS} : \text{Tr}_A[\theta_{UAS}] = \theta_U \otimes \vartheta_S$	$\varepsilon \leq e^{-\sup_{\alpha \in (1/2, 1)} [(1-\alpha)/\alpha] (I_{2-1/\alpha}^{\downarrow}(\mathbf{R} \mathbf{B})_{\mathcal{N}_{AS \rightarrow B}(\theta_{RA})})^{-\log M}}$
Entanglement-assisted quantum channel with casual state information	$\varepsilon \leq \text{Tr}[\mathcal{N}_{AS \rightarrow B}(\theta_{RAS}) \wedge (M-1)\theta_R \otimes \mathcal{N}_{AS \rightarrow B}(\theta_{AS})]$ for all $\theta_{RAS} : \text{Tr}_A[\theta_{RAS}] = \theta_R \otimes \vartheta_S$	$\log M \geq I_h^{\varepsilon-\delta}(\mathbf{U} \mathbf{B})_{\mathcal{N}_{AS \rightarrow B}(\theta_{UAS})} - \log \frac{1}{\delta}$
Broadcast quantum channel	$\varepsilon_B \leq \text{Tr}[\text{Tr}_C[\mathcal{N}_{A \rightarrow BC}(\theta_{UA})] \wedge (M_B-1)\theta_U \otimes \text{Tr}_C[\mathcal{N}_{A \rightarrow BC}(\theta_A)]]$ $\varepsilon_C \leq \text{Tr}[\text{Tr}_B[\mathcal{N}_{A \rightarrow BC}(\theta_{VA})] \wedge (M_C-1)\theta_V \otimes \text{Tr}_B[\mathcal{N}_{A \rightarrow BC}(\theta_A)]]$ for all $\theta_{UVA} : \text{Tr}_A[\theta_{UVA}] = \theta_U \otimes \theta_V$	$\varepsilon \leq e^{-\sup_{\alpha \in (1/2, 1)} [(1-\alpha)/\alpha] (I_{2-1/\alpha}^{\downarrow}(\mathbf{R} \mathbf{B})_{\mathcal{N}_{AS \rightarrow B}(\theta_{RAS})})^{-\log M}}$
Entanglement-assisted broadcast quantum channel	$\varepsilon_B \leq \text{Tr}[\text{Tr}_C[\mathcal{N}_{A \rightarrow BC}(\theta_{RA})] \wedge (M_B-1)\theta_R \otimes \text{Tr}_C[\mathcal{N}_{A \rightarrow BC}(\theta_A)]]$	$\log M \geq I_h^{\varepsilon-\delta}(\mathbf{R} \mathbf{B})_{\mathcal{N}_{AS \rightarrow B}(\theta_{RAS})} - \log \frac{1}{\delta}$
Multiple-access quantum channel	$\varepsilon_C \leq \text{Tr}[\text{Tr}_B[\mathcal{N}_{A \rightarrow BC}(\theta_{CA})] \wedge (M_C-1)\theta_C \otimes \text{Tr}_B[\mathcal{N}_{A \rightarrow BC}(\theta_A)]]$ for all $\theta_{RBCA} : \text{Tr}_A[\theta_{RBCA}] = \theta_R \otimes \theta_C$	
Entanglement-assisted multiple-access quantum channel	$\varepsilon_{XYC} \leq \text{Tr}[\rho_{XYC} \wedge ((M_A-1)\rho_X \otimes \rho_{YC} + (M_B-1)\rho_Y \otimes \rho_{XC} + (M_A-1)\rho_X \otimes \rho_Y \otimes \rho_C)]$ $\rho_{XYC} := \mathcal{N}_{AB \rightarrow C}(\theta_{XA} \otimes \theta_{YB})$	
	$\varepsilon \leq \text{Tr}[\rho_{RARBC} \wedge ((M_A-1)\rho_{RA} \otimes \rho_{RBC} + (M_B-1)\rho_{RB} \otimes \rho_{RAC} + (M_A-1)(M_B-1)\rho_{RA} \otimes \rho_{RB} \otimes \rho_C)]$ $\rho_{RARBC} := \mathcal{N}_{AB \rightarrow C}(\theta_{RAA} \otimes \theta_{RBB})$	
Classical data compression with quantum side information	$\varepsilon \leq \text{Tr} \left[ \rho_{XB} \wedge \left( \frac{1}{M} \mathbb{1}_X \otimes \rho_B \right) \right]$	$\varepsilon \leq e^{-\sup_{\alpha \in (1/2, 1)} [(1-\alpha)/\alpha] (\log M - H_{2-1/\alpha}^{\downarrow}(\mathbf{X} \mathbf{B})_{\rho})}$
		$\log M \leq H_h^{\varepsilon-\delta}(\mathbf{X} \mathbf{B})_{\rho} + \log \frac{1}{\delta}$

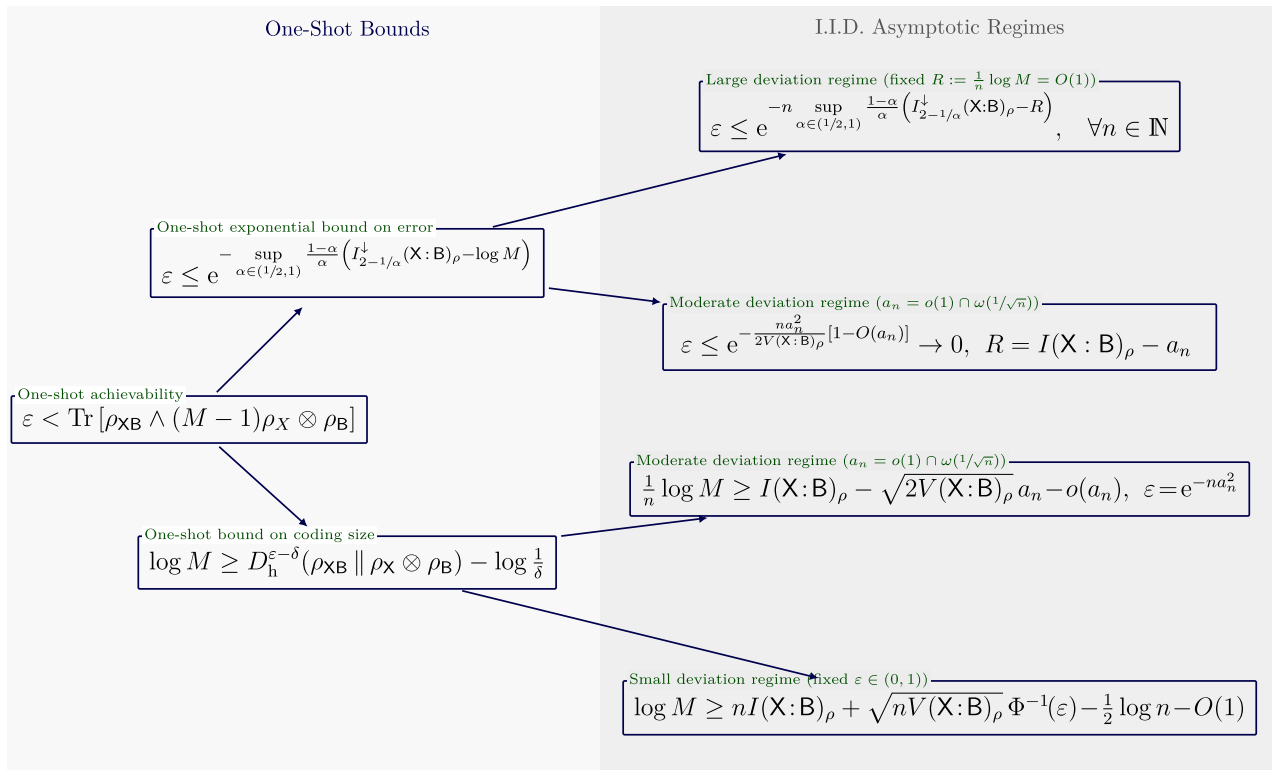


FIG. 2. Flow chart of the implications of the established one-shot achievability bound in the large deviation, small deviation, and moderate deviation regimes. The precise notation is given in Sec. III.

The well-known *Holevo-Helstrom theorem* [50–52,62] shows that the infimum can be attained by a Neyman-Pearson test  $T = \{A - B > 0\}$  that projects onto the positive part of the difference  $A - B$ , and the minimization is given by its dual formulation of a semidefinite program [63], [64, Sec. 1.2.3]:

$$\begin{aligned} & \inf_{0 \leq T \leq 1} \text{Tr}[A(1 - T)] + \text{Tr}[BT] \\ &= \sup_{M=M^\dagger} \{\text{Tr}[M]: M \leq A, M \leq B\} \\ &= \text{Tr}[A \wedge B]. \end{aligned} \tag{5}$$

Here the supremum is attained by the so-called *noncommutative minimal* (i.e., the operator with the greatest trace among the lower bounds in terms of the Loewner partial ordering) [50–52,62] of self-adjoint operators  $A$  and  $B$  [65], i.e.,

$$A \wedge B \in \arg \max_{M=M^\dagger} \{\text{Tr}[M]: M \leq A, M \leq B\}. \tag{6}$$

In other words, the Holevo-Helstrom theorem [50–52] provides an operational meaning to the noncommutative minimal “ $\wedge$ ” for characterizing the minimum *error* of distinguishing positive semidefinite operators  $A$  and  $B$  [66]. We adopt such an interpretation subsequently.

The main goal of this paper is to characterize the average error probability in quantum information-theoretic tasks in terms of the noncommutative minimal “ $\wedge$ .” To that end, we first review the important properties that will be used in the proposed analysis. We note that the following properties can be found in existing literature.

*Fact 1 (Properties of the noncommutative minimal [67]).*—Considering arbitrary self-adjoint operators  $A$  and  $B$ , the following hold the noncommutative minimal defined in Eq. (6) has the following properties.

- (i) (*Unique closed-form expression.*) The noncommutative minimal  $A \wedge B$  is unique and  $A \wedge B = \frac{1}{2}(A + B - |A - B|)$ .
- (ii) (*Monotone increase in the Loewner ordering.*) It holds that  $\text{Tr}[A \wedge B] \leq \text{Tr}[A' \wedge B']$  for  $A \leq A'$  and  $B \leq B'$ .
- (iii) (*Monotone increase under positive trace-preserving maps.*) It holds that  $\text{Tr}[A \wedge B] \leq \text{Tr}[\mathcal{N}(A) \wedge \mathcal{N}(B)]$  for any positive trace-preserving map  $\mathcal{N}$ .
- (iv) (*Concavity.*) The map  $(A, B) \mapsto \text{Tr}[A \wedge B]$  is jointly concave.
- (v) (*Direct sum.*) It holds that  $(A \oplus A') \wedge (B \oplus B') = (A \wedge B) \oplus (A' \wedge B')$  for any self-adjoint  $A'$  and  $B'$ .
- (vi) (*Upper bound.*) It holds that  $\text{Tr}[A \wedge B] \leq \text{Tr}[A^{1-s} B^s]$  for any  $A, B \geq 0$  and  $s \in (0, 1)$ .

TABLE II. Comparisons of the one-shot achievability bounds on the error probability and the coding size and rate established in Propositions 1 and 2 of Sec. III with existing results. We also present the IID asymptotic expansion of the coding rate to highlight the resulting third-order terms, where we use the shorthand  $I \equiv I(\mathbf{X} : \mathbf{B})_\rho$  and  $V \equiv V(\mathbf{X} : \mathbf{B})_\rho$  for brevity.

One-shot exponential bounds on the coding error		
Proposition 1	$\varepsilon \leq e^{-\sup_{\alpha \in (1/2, 1)} [(1-\alpha)\alpha] (I_{2-1/\alpha}^\downarrow(\mathbf{X} : \mathbf{B})_\rho - R)}$	
Hayashi [35]	$\varepsilon \leq 4e^{-\sup_{\alpha \in (1/2, 1)} [(1-\alpha)\alpha] (I_{2-1/\alpha}^\downarrow(\mathbf{X} : \mathbf{B})_\rho - R)}$	
Burnashev and Holevo [89] (pure-state channels)	$\varepsilon \leq 2e^{-\sup_{\alpha \in (1/2, 1)} [(1-\alpha)\alpha] (I_{2-1/\alpha}^\downarrow(\mathbf{X} : \mathbf{B})_\rho - R)}$	
Achievability bounds on the coding size		
	One-shot bounds	IID asymptotic expansion
Proposition 2	$\log M \geq D_h^{\varepsilon-\delta}(\rho_{\mathbf{X}\mathbf{B}} \parallel \rho_{\mathbf{X}} \otimes \rho_{\mathbf{B}}) - \log(1/\delta)$	$\log M \geq nI + \sqrt{nV} \Phi^{-1}(\varepsilon) - \frac{1}{2} \log n - O(1)$
Hayashi and Nagaoka [30] Wang and Renner [46]	$\log M \geq D_h^{\varepsilon-\delta}(\rho_{\mathbf{X}\mathbf{B}} \parallel \rho_{\mathbf{X}} \otimes \rho_{\mathbf{B}}) - \log(4/\delta^2)$	$\log M \geq nI + \sqrt{nV} \Phi^{-1}(\varepsilon) - \log n - O(1)$
Beigi and Gohari [91]	$\log M \geq D_s^{\varepsilon-\delta}(\rho_{\mathbf{X}\mathbf{B}} \parallel \rho_{\mathbf{X}} \otimes \rho_{\mathbf{B}}) - \log[(1-\varepsilon)/\delta]$ $\geq D_h^{\varepsilon-2\delta}(\rho_{\mathbf{X}\mathbf{B}} \parallel \rho_{\mathbf{X}} \otimes \rho_{\mathbf{B}}) - \log[(1-\varepsilon)/\delta^2]$	$\log M \geq nI + \sqrt{nV} \Phi^{-1}(\varepsilon) - \log n - O(1)$
Ogawa [101]	$\log M \geq D_s^{\varepsilon-\delta}(\rho_{\mathbf{X}\mathbf{B}} \parallel \rho_{\mathbf{X}} \otimes \rho_{\mathbf{B}}) - \log(1/\delta)$ $\geq D_h^{\varepsilon-2\delta}(\rho_{\mathbf{X}\mathbf{B}} \parallel \rho_{\mathbf{X}} \otimes \rho_{\mathbf{B}}) - \log(1/\delta^2)$	$\log M \geq nI + \sqrt{nV} \Phi^{-1}(\varepsilon) - \log n - O(1)$

(vii) (*Lower bound [68].*) It holds that  $\text{Tr}[A \wedge B] \geq \text{Tr}\{A[\frac{B}{A+B}]\}$  for any  $A, B \geq 0$ .

*Proof.*—for (i), the uniqueness (also for multiple operators) was proved by Holevo [62, Theorem 2], [52, Sec. 2.2] and later also by Audenaert and Mosonyi [69, Theorem A.3 and Eq. (85)]; the closed-form expression may have already been known by Holevo and Helstrom [50–52] (see also Ref. [69, Lemma A.7]).

Property (ii) follows directly from the definition given in Eq. (6) (see also Ref. [69, Lemma A.8]).

Properties (iii) and (iv) follow from the fact that the trace norm (i.e.,  $\|M\|_1 := \text{Tr}[|M|]$ ) is contractive under positive trace-preserving maps and the triangle inequality (see, e.g., Ref. [70, Theorems 9.2 and 9.3]) of  $\|\cdot\|_1$ . We note that the monotone increase under a positive trace-preserving map and the concavity for multiple operators also hold.

Property (v) with trace is due to the direct-sum structure of the trace norm; the case without trace was proved in Ref. [69, Lemma A.9].

Property (vi) is the celebrated inequality of Audenaert *et al.* [33,34,71] used in proving the *quantum Chernoff bound*; later, it was generalized to infinite-dimensional Hilbert space [36].

Property (vii) in a special case of  $\text{Tr}[A+B]=1$  is an immediate consequence of the Barnum-Knill theorem [64, Theorem 3.10], [72]. That is, the error probability using the pretty good measurement [47,48] is no larger than twice that of using the optimal measurement. The proof for the general case of  $A, B \geq 0$  can be found in the author's previous work [73, Lemma 3]. For completeness, we provide an alternative proof of property (vii) (and a strengthened result of it) in the Appendix. ■

### III. MAIN RESULT: A ONE-SHOT ACHIEVABILITY FOR CLASSICAL-QUANTUM CHANNEL CODING

In this section, we prove our main result of establishing a one-shot achievability bound for classical-quantum channel coding via a direct application of the pretty-good measurement (PGM) [47,48].

*Definition 1 (Classical-quantum channel coding).*—Let  $\mathcal{N}_{\mathbf{X} \rightarrow \mathbf{B}}: x \mapsto \rho_{\mathbf{B}}^x$  be a classical-quantum channel, where each channel output  $\rho_{\mathbf{B}}^x$  is a density operator (i.e., a positive semidefinite operator with unit trace).

- (1) Alice holds classical registers  $\mathbf{M}$  and  $\mathbf{X}$ , and Bob holds a quantum register  $\mathbf{B}$ .
- (2) An encoding  $m \mapsto x(m)$  maps equiprobable messages in  $\mathbf{M}$  to a codeword in  $\mathbf{X}$ .
- (3) The classical-quantum channel  $\mathcal{N}_{\mathbf{X} \rightarrow \mathbf{B}}$  is applied on Alice's register  $\mathbf{X}$  and outputs a state on  $\mathbf{B}$  at Bob.
- (4) A decoding measurement described by a positive operator-valued measure (POVM)  $\{\Pi_{\mathbf{B}}^m\}_{m \in \mathbf{M}}$  is performed on Bob's quantum register  $\mathbf{B}$  to extract the sent message  $m$ .

An  $(M, \varepsilon)$  code for  $\mathcal{N}_{\mathbf{X} \rightarrow \mathbf{B}}$  is a protocol such that  $|\mathbf{M}| = M$  and the average error probability satisfies

$$\frac{1}{M} \sum_{m \in \mathbf{M}} \text{Tr}[\rho_{\mathbf{B}}^{x(m)} (\mathbf{1}_{\mathbf{B}} - \Pi_{\mathbf{B}}^m)] \leq \varepsilon.$$

The encoding is the standard random coding strategy.

- (a) **Encoding.** Consider a random codebook  $\mathcal{C} = \{x(1), x(2), \dots, x(M)\}$ , where each of the codewords  $x(m) \in \mathbf{X}$  is pairwise independently drawn from a probability distribution  $p_X$ . Alice sends codewords according to the codebook  $\mathcal{C}$ .
- (b) **Decoding.** At the receiver, given a realization of the random codebook  $\mathcal{C}$  and the corresponding channel output states  $\{\rho_B^{x(m)}\}_{m \in \mathbf{M}}$ , Bob performs the PGM to decode each message  $m \in \mathbf{M}$ :

$$\Pi_B^m := \frac{\rho_B^{x(m)}}{\sum_{\bar{m} \in \mathbf{M}} \rho_B^{x(\bar{m})}} \quad \text{for all } m \in \mathbf{M}. \quad (7)$$

Our main result is the following.

*Theorem 1 (A one-shot achievability bound for classical-quantum channel coding).*—Consider an arbitrary classical-quantum channel  $\mathcal{N}_{X \rightarrow B} : x \mapsto \rho_B^x$ . Then, there exists an  $(M, \varepsilon)$  code for  $\mathcal{N}_{X \rightarrow B}$  such that, for any probability distribution  $p_X$ ,

$$\varepsilon \leq \text{Tr}[\rho_{XB} \wedge (M - 1)\rho_X \otimes \rho_B]. \quad (8)$$

Here,  $\rho_{XB} := \sum_{x \in \mathbf{X}} p_X(x) |x\rangle\langle x| \otimes \rho_B^x$  and the noncommutative minimal is  $A \wedge B = \frac{1}{2}(A + B - |A - B|)$  (see Fact 1 (i)).

*Proof.*—The claim follows from the lower bound of the noncommutative minimal “ $\wedge$ ” given in Fact 1 (vii) for relating the pretty good measurement to the optimal measurement, and the concavity of “ $\wedge$ ,” i.e., Fact 1 (iv). Precisely, given any realization of codebook  $\mathcal{C} = \{x(m)\}_{m \in \mathbf{M}}$ , we calculate the average probability of erroneous decoding using the PGM given in Eq. (7) as

$$\begin{aligned} & \frac{1}{M} \sum_{m \in \mathbf{M}} \text{Tr} \left[ \rho_B^{x(m)} \frac{\sum_{\bar{m} \neq m} \rho_B^{x(\bar{m})}}{\rho_B^x + \sum_{\bar{m} \neq m} \rho_B^{\bar{m}}} \right] \\ & \leq \frac{1}{M} \sum_{m \in \mathbf{M}} \text{Tr} \left[ \rho_B^{x(m)} \wedge \left( \sum_{\bar{m} \neq m} \rho_B^{x(\bar{m})} \right) \right], \end{aligned} \quad (9)$$

where we have applied Fact 1 (vii) with  $A = \rho_B^{x(m)}$  and  $B = \sum_{\bar{m} \neq m} \rho_B^{x(\bar{m})}$  to relate the error probability under the PGM to the expression in terms of the noncommutative minimal. Next, we take the expectation for each  $x(m) \sim p_X$  to bound the expected average error probability (which is also called

the random-coding error probability), i.e.,

$$\begin{aligned} & \frac{1}{M} \sum_{m \in \mathbf{M}} \mathbb{E}_{x(m), x(\bar{m}) \sim p_X} \text{Tr} \left[ \rho_B^{x(m)} \wedge \left( \sum_{\bar{m} \neq m} \rho_B^{x(\bar{m})} \right) \right] \\ & \stackrel{(a)}{\leq} \frac{1}{M} \sum_{m \in \mathbf{M}} \mathbb{E}_{x(m) \sim p_X} \text{Tr} \left[ \rho_B^{x(m)} \wedge \left( \mathbb{E}_{x(\bar{m})|x(m)} \left[ \sum_{\bar{m} \neq m} \rho_B^{x(\bar{m})} \right] \right) \right] \\ & \stackrel{(b)}{=} \frac{1}{M} \sum_{m \in \mathbf{M}} \mathbb{E}_{x(m) \sim p_X} \text{Tr} [\rho_B^{x(m)} \wedge (M - 1)\rho_B] \\ & = \mathbb{E}_{x \sim p_X} \text{Tr} [\rho_B^x \wedge (M - 1)\rho_B], \end{aligned} \quad (10)$$

where in (a) we used the concavity given in Fact 1 (iv) and in (b) we recalled the pairwise independence of the random codebook.

Invoking the direct sum formula given in Fact 1 (v), we arrive at the claimed inequality at the right-hand side of Eq. (8). Lastly, since the random-coding error probability using any  $p_X$  is larger than the error probability of the optimal code, the proof is completed. ■

Below, we provide a detailed explanation of how PGM works. An important feature of PGM is that the POVM element  $\Pi_B^{x(m)}$  given in Eq. (7) is *proportional* to the sent state  $\rho_B^{x(m)}$ . On the other hand, the complement of the POVM element, i.e.,  $\mathbb{1}_B - \Pi_B^{x(m)}$ , is proportional to the sum of the remaining states  $\sum_{\bar{m} \neq m} \rho_B^{x(\bar{m})}$ . Hence, the average error probability of discriminating  $M$  channel output states, i.e., the left-hand side of Eq. (9), is equivalent to the error of deciding each sent state  $\rho_B^{x(m)}$  using the following two-outcome PGM:

$$\left\{ \frac{\rho_B^{x(m)}}{\rho_B^{x(m)} + \sum_{\bar{m} \neq m} \rho_B^{x(\bar{m})}}, \frac{\sum_{\bar{m} \neq m} \rho_B^{x(\bar{m})}}{\rho_B^{x(m)} + \sum_{\bar{m} \neq m} \rho_B^{x(\bar{m})}} \right\}.$$

Such the discrimination between  $\rho_B^{x(m)}$  with prior probability  $1/M$  against the sum of the remaining states (again each with prior probability  $1/M$ ) reflects the nature of the union bound inherited in the PGM; cf. the right-hand side of Eq. (1). Next, taking the expectation on the rest of the states ensures that we are discriminating  $\rho_B^{x(m)}$  with prior probability  $1/M$  against  $(M - 1)$  identical marginal states  $\rho_B$ , each with prior probability  $1/M$ . Equivalently, this amounts to a binary hypothesis testing between  $\rho_B^{x(m)}$  with prior probability  $1/M$  against the marginal states  $\rho_B$  with prior probability  $(M - 1)/M$ . Lastly, after taking the summation over  $m \in \mathbf{M}$ , the above is equal to the discrimination of the joint state  $\rho_{XB}$  against the scaled decoupled product state  $(M - 1)\rho_X \otimes \rho_B$ . (See Fig. 1 for the illustration.) We hope that this simple proof provides a conceptually clear elucidation on the intimate relation between classical-quantum channel coding and quantum hypothesis testing in a pedagogical way.



*Remark 1.*—In the classical case where  $\{\rho_B^x\}_{x \in X}$  mutually commute, Theorem 1 reduces to a result by Polyanskiy [74, Eq. (2.121)], which is only 1-bit weaker than the *dependence testing bound* by Polyanskiy *et al.* [22, Theorem 17].

*Remark 2.*—As we show shortly in Secs. III A and IV, the one-shot bound established in Theorem 1 already implies (and sharpens) various previously known achievability results in the so-called *achievable rate region*, i.e., rates below the quantum mutual information. Is the bound in Theorem 1 tight outside the achievable rate region? Taking  $c$ - $q$  channel coding as an example, when the message size is too large or the coding rate (i.e.,  $R := 1/n \log M$ ) is way above the mutual information with respect to  $\rho_{XB}$ , the one-shot bound in Theorem 1 might not be very tight. If  $\log(M-1) \geq D_\infty^*(\rho_{XB} \parallel \rho_X \otimes \rho_B)$ , where  $D_\infty^*(A \parallel B) := \inf\{\gamma \in \mathbb{R} : A \leq e^\gamma B\}$  is the *max-relative entropy* [75–77], then Eq. (8) yields a trivial bound:  $\varepsilon \leq 1$ .

In view of this, Theorem 1 can be strengthened to the following more involved form:

$$\varepsilon \leq \left(1 - \frac{1}{M} \text{Tr}[\rho_{XB} \wedge (M-1)\rho_X \otimes \rho_B]\right) \text{Tr}[\rho_{XB} \wedge (M-1)\rho_X \otimes \rho_B]. \quad (11)$$

Bound (11) follows from the tighter inequality (A2) given in Lemma 1 in the Appendix, instead of Fact 1 (vii). Now if  $\log(M-1) \geq D_\infty^*(\rho_{XB} \parallel \rho_X \otimes \rho_B)$  then the random coding error amounts to randomly guessing equiprobable messages, i.e.,  $\varepsilon \leq 1 - 1/M$ . Regardless of the message size  $M$ , Eq. (11) is technically a tighter one-shot bound compared to Eq. (8). This naturally raises the question of whether Eq. (11) can lead to a simple proof of the upper bound on the strong converse exponent of  $c$ - $q$  channel coding; see [78, Section 5.4], [79, Proposition IV.5], and [80, Proposition VI.2.]. We leave this for future work.

The established one-shot achievability in Theorem 1 immediately covers (and sharpens) various known results of deriving the minimal error given a fixed message or coding size  $M$  or deriving the maximal message size given a fixed error  $\varepsilon$ . Let us define the following two important operational quantities for  $c$ - $q$  channel coding:

$$\varepsilon^*(\mathcal{N}, M) := \inf\{\varepsilon \in \mathbb{R} : \exists \text{ an}(M, \varepsilon) \text{ code for } \mathcal{N}\},$$

$$M^*(\mathcal{N}, \varepsilon) := \sup\{M \in \mathbb{N} : \exists \text{ an}(M, \varepsilon) \text{ code for } \mathcal{N}\}.$$

We note that although  $\varepsilon^*(\mathcal{N}, M)$  and  $M^*(\mathcal{N}, \varepsilon)$  are inverse functions to each other in the one-shot setting, they lead to different asymptotic expansions in the large deviation and small deviation regimes, respectively.

*Proposition 1 (Bounding the coding error given a fixed coding rate).*—Consider an arbitrary classical-quantum channel  $\mathcal{N}_{X \rightarrow B} : x \mapsto \rho_B^x$ . Then, for any  $n \in \mathbb{N}$  and  $R > 0$ , there exists an  $(e^{nR}, \varepsilon)$  code for  $\mathcal{N}_{X \rightarrow B}^{\otimes n}$  such that, for any

probability distribution  $p_X$ ,

$$\varepsilon \leq e^{-[n(1-\alpha)/\alpha](I_{2-1/\alpha}^\downarrow(X:B)_\rho - R)} \quad \text{for all } \alpha \in (\tfrac{1}{2}, 1).$$

Here,  $I_\alpha^\downarrow(X:B)_\rho := D_\alpha(\rho_{XB} \parallel \rho_X \otimes \rho_B)$ , the state is evaluated on  $\rho_{XB} := \sum_{x \in X} p_X(x)|x\rangle\langle x| \otimes \rho_B^x$ , and the quantum Petz-Rényi divergence [81] is  $D_\alpha(\rho \parallel \sigma) := [1/(\alpha-1)] \log \text{Tr}[\rho^\alpha \sigma^{1-\alpha}]$ .

The exponent  $\sup_{\alpha \in (1/2, 1)} [(1-\alpha)/\alpha](I_{2-1/\alpha}^\downarrow(X:B)_\rho - R)$  is positive if and only if  $R > I(X:B)_\rho := D(\rho_{XB} \parallel \rho_X \otimes \rho_B)$ .

*Proof.*—For the one-shot case  $n = 1$ , we apply the inequality of Audenaert *et al.*, i.e., Fact 1 (vi), on the one-shot bound given in Theorem 1 with  $A = \rho_{XB}$ ,  $B = (M-1)\rho_X \otimes \rho_B$ , and  $s = (1-\alpha)/\alpha$  to obtain the large-deviation-type bound. When considering product channels in the  $n$ -shot scenario, the exponential decay follows from the fact that  $\rho \mapsto I_{2-1/\alpha}^\downarrow(X:B)_\rho$  is additive for any  $n$ -fold product state. The positivity holds by noting that the map  $\alpha \mapsto I_{2-1/\alpha}^\downarrow(X:B)_\rho$  is nondecreasing on  $[\frac{1}{2}, 1]$  [78, Lemma 3.12]. ■

*Proposition 2 (Bounding the coding rate given a fixed coding error).*—Consider an arbitrary classical-quantum channel  $\mathcal{N}_{X \rightarrow B} : x \mapsto \rho_B^x$ . Then, for any  $\varepsilon \in (0, 1)$ , there exists an  $(M, \varepsilon)$  code for  $\mathcal{N}_{X \rightarrow B}$  such that, for any probability distribution  $p_X$  and any  $\delta \in (0, \varepsilon)$ ,

$$\log M \geq D_h^{\varepsilon-\delta}(\rho_{XB} \parallel \rho_X \otimes \rho_B) - \log \frac{1}{\delta}. \quad (12)$$

Here,  $D_h^\varepsilon(\rho \parallel \sigma) := \sup_{0 \leq T \leq \mathbb{1}} \{-\log \text{Tr}[\sigma T] : \text{Tr}[\rho T] \geq 1 - \varepsilon\}$  is the  $\varepsilon$ -*hypothesis-testing divergence* [38,39,46].

Moreover, for any  $\varepsilon \in (0, 1)$  and sufficiently large  $n \in \mathbb{N}$ , there exists an  $(M, \varepsilon)$  code for  $\mathcal{N}_{X \rightarrow B}^{\otimes n}$  such that, for any probability distribution  $p_X$ ,

$$\log M \geq nI(X:B)_\rho + \sqrt{nV(X:B)_\rho} \Phi^{-1}(\varepsilon) - \frac{1}{2} \log n - O(1),$$

where  $V(X:B)_\rho := V(\rho_{XB} \parallel \rho_X \otimes \rho_B)$ ,  $V(\rho \parallel \sigma) := \text{Tr}[\rho(\log \rho - \log \sigma)^2] - D(\rho \parallel \sigma)^2$ , and  $\Phi^{-1}(\varepsilon) := \sup\{u : \int_{-\infty}^u (1/\sqrt{2\pi})e^{-t^2/2} dt \leq \varepsilon\}$  is the inverse of the cumulative distribution of the standard normal distribution.

*Proof.*—By recalling the definition of the noncommutative minimal given in Eq. (5) and by Theorem 1, for any test  $0 \leq T_{XB} \leq \mathbb{1}_{XB}$  satisfying  $\text{Tr}[\rho_{XB}(\mathbb{1}_{XB} - T_{XB})] \leq \varepsilon - \delta$ , one has

$$\begin{aligned} \varepsilon &\leq \text{Tr}[\rho_{XB}(\mathbb{1}_{XB} - T_{XB})] + (M-1) \text{Tr}[\rho_X \otimes \rho_B T_{XB}] \\ &\leq \varepsilon - \delta + (M-1)e^{-D_h^{\varepsilon-\delta}(\rho_{XB} \parallel \rho_X \otimes \rho_B)}. \end{aligned}$$

The second-order achievability then follows from the expansion of the quantum hypothesis-testing divergence

[29,38,39,82,83] by choosing  $\delta = 1/\sqrt{n}$ :  $D_h^{\varepsilon \pm \delta}(\rho^{\otimes n} \parallel \sigma^{\otimes n}) \geq nD(\rho \parallel \sigma) + \sqrt{nV(\rho \parallel \sigma)}\Phi^{-1}(\varepsilon) - O(1)$ . ■

We remark that both Propositions 1 and 2 extend to the moderate deviation regime by directly following the approaches from Refs. [41,42]. We refer the reader to Figure 2 for the corresponding expressions.

*Remark 3.*—Given that Theorem 1 already provides a one-shot bound on the average error probability, one may wonder why weaken Eq. (8) in Theorem 1 to obtain another one-shot bound in Proposition 1 (note that they both have closed-form expressions). The reason is that the minimum error in terms of the noncommutative minimal on the right-hand side of Eq. (8) is not multiplicative under product states. Nevertheless, it can be further upper bounded by certain multiplicative Rényi-type quantities. That is exactly the spirit of the quantum Chernoff bound [33,34,36,71], and, hence, we term the result of Proposition 1 as a kind of *large deviation type bound* [84] accordingly.

On the other hand, Theorem 1 also gives a one-shot and asymptotic expansions in the small deviation regime (Proposition 2). Hence, to some extent, Theorem 1 may be viewed as a “meta” achievability for classical communication over quantum channels (see also Theorem 3 in Sec. IV B below).

*Remark 4.*—Most existing one-shot achievability bounds to date (e.g., Refs. [30,43,46,59]) are expressed in terms of the quantum hypothesis-testing divergence  $D_h^\varepsilon$ , as in Eq. (12) of Proposition 2, because they directly provide a one-shot characterization (lower bound) on the maximal message or coding size  $M$  given a fixed coding error  $\varepsilon$ , which is also called the  $\varepsilon$ -one-shot channel capacity. To numerically compute  $D_h^\varepsilon$ , one can formulate the quantity in the standard form of a semidefinite program (SDP); namely, it is an optimization over a  $d_B \times d_B$  matrix-valued variable with  $m := d_B^2 + 1$  linear (scalar) constraints, where we use  $d_B$  to denote the dimension of the underlying Hilbert space representing the quantum register  $\mathbf{B}$ . (Here, we only consider the computation on the quantum part of register  $\mathbf{B}$  for simplicity without involving computation on the classical register  $\mathbf{X}$ .) Using the state-of-the-art (classical) SDP solver [85], the running time [86] is  $O^*(m^\omega) = O^*(d_B^{2\omega}) = O^*(d_B^{4.746})$ , where  $\omega \leq 2.373$  is the exponent of matrix multiplication [87].

On the other hand, the one-shot bound provided in Theorem 1 admits a closed-form expression in terms of the trace norm. Using the state-of-the-art algorithm for approximating singular values [88], it requires a running time of  $O^*(d_B^\omega \log^2 d_B) = O^*(d_B^{2.373} \log^2 d_B)$ . This then shows that the computation of the proposed one-shot achievability bound in terms of the noncommutative minimal in Theorem 1 is nearly *quadratically efficient* compared to the computation of the one-shot bounds in terms of the quantum hypothesis-testing divergence.

### A. Comparison to existing results

In the following, we compare the implications of the established one-shot achievability bounds, i.e., Propositions 1 and 2, with existing results. We refer the reader to Table II for a summary.

The exponential decaying rate of the error probability given in Proposition 1 matches that proved by Hayashi [35, Eq. (9)]. However, in the one-shot setting, the large deviation type bound in Proposition 1 is tighter than Eq. (9) of Ref. [35] (without the factor 4). Furthermore, if  $\mathcal{N}_{\mathbf{X} \rightarrow \mathbf{B}}$  is a pure-state channel, one has  $I_{2-1/\alpha}^\downarrow(\mathbf{X} : \mathbf{B})_\rho = I_\alpha(\mathbf{X} : \mathbf{B})_\rho := \inf_{\sigma_B \in \mathcal{S}(\mathcal{H}_B)} D_\alpha(\rho_{\mathbf{X}\mathbf{B}} \parallel \rho_{\mathbf{X}} \otimes \rho_B) = [\alpha/(\alpha - 1)] \log \text{Tr}[(\sum_{x \in \mathcal{X}} p_{\mathbf{X}}(x)(\rho_B^x)^\alpha)^{1/\alpha}]$  [where the minimization is over all density operators on Hilbert space  $\mathcal{H}_B$ , i.e.,  $\mathcal{S}(\mathcal{H}_B)$ ]. Hence, the bound in Proposition 1 is tighter than the bound proved by Burnashev and Holevo [89, Proposition 1] (without the factor 2).

Hayashi and Nagaoka [30, Lemma 3] and Wang and Renner [46, Theorem 1] employed the Hayashi-Nagaoka inequality (2) to obtain a one-shot achievability bound [90] on the message or coding size  $M$ : for any  $0 < \delta < \varepsilon < 1$ , choosing  $c = \delta/(2\varepsilon - \delta)$  in Eq. (2),

$$\log M \geq D_h^{\varepsilon - \delta}(\rho_{\mathbf{X}\mathbf{B}} \parallel \rho_{\mathbf{X}} \otimes \rho_{\mathbf{B}}) - \log \frac{4}{\delta^2}, \quad (13)$$

in terms of the hypothesis-testing divergence  $D_h^\varepsilon$  introduced in Proposition 2. The term  $-\log 4/\delta^2$  results from optimizing coefficient  $c$  when applying the Hayashi-Nagaoka inequality. Compared to Eq. (13), Proposition 2 does not need to choose the appropriate coefficient  $c$ , and, hence, it gives a tighter one-shot achievability bound on  $M$  (especially when  $\delta$  is small):

$$\log M \geq D_h^{\varepsilon - \delta}(\rho_{\mathbf{X}\mathbf{B}} \parallel \rho_{\mathbf{X}} \otimes \rho_{\mathbf{B}}) - \log \frac{1}{\delta}. \quad (14)$$

Specialized to the IID asymptotic scenario of  $n$ -fold product channels with  $\delta = 1/\sqrt{n}$ , Eq. (14) yields an improved third-order coding rate by a factor  $\frac{1}{2} \log n$  compared to the asymptotics based on Eq. (13).

Beigi and Gohari [91] generalized a superb classical achievability approach by Yassaee *et al.* [92] to establish a one-shot achievability bound on  $M$  [91, Corollary 1] as well:

$$\log M \geq D_s^{\varepsilon - \delta}(\rho_{\mathbf{X}\mathbf{B}} \parallel \rho_{\mathbf{X}} \otimes \rho_{\mathbf{B}}) - \log \frac{1 - \varepsilon}{\delta} \quad (15)$$

with  $D_s^\varepsilon(\rho \parallel \sigma) := \sup\{\gamma \in \mathbb{R} : \text{Tr}[\rho \leq e^\gamma \sigma] \leq \varepsilon\}$  the information-spectrum divergence [16,17,30,38,93]. Comparing Eq. (14) to Eq. (15), we recall the relation between the quantum hypothesis-testing divergence  $D_h^\varepsilon$  and the quantum information-spectrum divergence  $D_s^\varepsilon$  [38, Lemma

12]: for all  $0 < \delta' < \varepsilon$ ,

$$D_h^\varepsilon(\rho \parallel \sigma) \geq D_s^\varepsilon(\rho \parallel \sigma) \geq D_h^{\varepsilon-\delta'}(\rho \parallel \sigma) - \log \frac{1}{\delta'}. \quad (16)$$

This indicates that the proposed one-shot bound (14) has a stronger leading term  $D_h^{\varepsilon-\delta}$  instead of  $D_s^{\varepsilon-\delta}$  [94].

When considering the asymptotic expansion of the coding rate in the IID setting, one has to translate  $D_s^\varepsilon$  in Eq. (15) back to  $D_h^\varepsilon$  using Eq. (16) and to employ the second-order achievability [95] of the quantum hypothesis-testing divergence  $D_h^\varepsilon$  [38,39,82,83]:

$$D_h^{\varepsilon \pm \delta}(\rho^{\otimes n} \parallel \sigma^{\otimes n}) \geq nD(\rho \parallel \sigma) + \sqrt{nV(\rho \parallel \sigma)}\Phi^{-1}(\varepsilon) - O(1). \quad (17)$$

Then, Beigi and Gohari's result, Eq. (15), leads to

$$\log M \geq nI(\mathbf{X} : \mathbf{B})_\rho + \sqrt{nV(\mathbf{X} : \mathbf{B})_\rho}\Phi^{-1}(\varepsilon) - \log n - O(1).$$

This achieves the same third-order term as the asymptotic expansion using Eqs. (13) and (17). On the other hand, the established Eq. (14) with Eq. (17) gives a tighter third-order term for the coding rate:

$$\begin{aligned} \log M \geq nI(\mathbf{X} : \mathbf{B})_\rho \\ + \sqrt{nV(\mathbf{X} : \mathbf{B})_\rho}\Phi^{-1}(\varepsilon) - \frac{1}{2} \log n - O(1). \end{aligned}$$

Inspired by the third-order asymptotics of the classical hypothesis-testing divergence proved by Strassen [12, Theorem 3.1] (see also Refs. [22, Lemma 46] and [37, Proposition 2.3]), we conjecture the following third-order achievability of the quantum hypothesis-testing divergence:

$$\begin{aligned} D_h^{\varepsilon \pm \delta}(\rho^{\otimes n} \parallel \sigma^{\otimes n}) \geq nD(\rho \parallel \sigma) + \sqrt{nV(\rho \parallel \sigma)}\Phi^{-1}(\varepsilon) \\ + \frac{1}{2} \log n - O(1). \end{aligned} \quad (18)$$

If Eq. (18) holds then the established Eq. (14) will imply that

$$\log M \geq nI(\mathbf{X} : \mathbf{B})_\rho + \sqrt{nV(\mathbf{X} : \mathbf{B})_\rho}\Phi^{-1}(\varepsilon) - O(1). \quad (19)$$

We remark that Eq. (19) will give the best possible achievable third-order coding rate for  $c$ - $q$  channel coding without further assumptions on the channel [96].

*Remark 5.*—At the writing of this paper, a very recent work by Renes established the optimal error exponent for *symmetric* classical-quantum channels [97]. The result is asymptotic, but it matches the quantum sphere-packing bound [98–100] for the high achievable rate region, and, hence, it is asymptotically optimal and tighter than the error exponent obtained in Proposition 1 in the IID asymptotic setting for symmetric classical-quantum channels. A one-shot bound for that is still missing, which we leave for future work.

## IV. APPLICATIONS IN QUANTUM INFORMATION THEORY

The analysis proposed in Sec. III naturally extends to classical communication over quantum channels, network information theory [102], and beyond; see Table I in Sec. I. We apply our analysis using the pretty-good measurement to binary quantum hypothesis testing in Sec. IV A. We present entanglement-assisted classical communication over quantum channels in Sec. IV B. Section IV C is for classical data compression with quantum side information. Section IV D studies entanglement-assisted and unassisted classical communication over quantum multiple-access channels. Section IV E considers entanglement-assisted and unassisted classical communication over quantum broadcast channels. Section IV F is devoted to entanglement-assisted classical communication over quantum channels with casual state information available at the encoder.

### A. Binary quantum hypothesis testing

Binary quantum hypothesis testing and the optimal quantum measurement is a relatively well-studied topic in quantum information theory due to its simpler mathematical structure and operational significance [33–36,38,39,41,42,50–52,103–114]. The goal of this section is not to re-do the analysis via optimal measurements, but to show how the suboptimal pretty-good measurement along with the properties of the noncommutative minimal given in Sec. II can recover the existing results with only a slightly suboptimal coefficient. Specifically, we show that the pretty-good measurement can also achieve the *quantum Hoeffding bound* [34, Sec. 5.5]. This indicates that the proposed analysis should not be too loose in terms of the one-shot exponential bounds (at least for binary quantum hypothesis testing).

#### 1. Symmetric scenario

We first consider the symmetric scenario, where the two quantum hypotheses are described by density operators  $\rho$  and  $\sigma$  with prior probabilities  $p \in (0, 1)$  and  $1 - p$ , respectively. Note that the one-shot quantum hypothesis testing is also known as the *quantum state discrimination*; the relation between the optimal measurement (i.e., the quantum Neyman-Pearson test) and the pretty-good measurement was proved by Barnum and Knill [72]; see also Ref. [64, Theorem 3.10].

Subsequently, we show that the lower bound on the noncommutative minimal (Fact 1 (vii)) can be interpreted as an adaptation of the Barnum-Knill theorem. On the one hand, the Holevo-Helstrom theorem [50–52] shows that the minimal error for distinguishing  $\rho$  and  $\sigma$  in the symmetric scenario is given by  $\text{Tr}[p\rho \wedge (1-p)\sigma]$ . On the other hand, by using the pretty-good measurement with

respect to the weighted states  $(p\rho, (1-p)\sigma)$  and applying the lower bound of the noncommutative minimal, Fact 1 (vii), the corresponding error probability is given by

$$p \operatorname{Tr} \left[ \rho \frac{(1-p)\sigma}{p\rho + (1-p)\sigma} \right] + (1-p) \operatorname{Tr} \left[ \sigma \frac{p\rho}{p\rho + (1-p)\sigma} \right] \leq 2 \operatorname{Tr} [p\rho \wedge (1-p)\sigma], \quad (20)$$

which is twice the error probability compared to the minimal error via the optimal measurement. This coincides with the claim made by Barnum and Knill on the relation between the error probability using the optimal measurement and that using the pretty-good measurement.

*Remark 6.*—As mentioned in Remark 2 of Sec. III, the upper bound in Eq. (20) can be strengthened to

$$2(1 - \operatorname{Tr} [p\rho \wedge (1-q)\sigma]) \operatorname{Tr} [p\rho \wedge (1-q)\sigma],$$

using Eq. (A2) of Lemma 1 in the Appendix instead of Fact 1 (vii).

On the other hand, one can also use the pretty-good measurement of the form  $\{\frac{\rho}{\rho+\sigma}, \frac{\sigma}{\rho+\sigma}\}$  to obtain an achievable error probability  $(1 - \operatorname{Tr} [\rho \wedge \sigma]) \operatorname{Tr} [\rho \wedge \sigma]$  or simply  $\operatorname{Tr} [\rho \wedge \sigma]$ .

## 2. Asymmetric scenario

We move on to consider the asymmetric scenario, namely, the trade-off between the type-I error and the type-II error without knowing the prior distribution. We use the pretty-good measurement  $\{\rho/(\rho + \mu\sigma), \mu\sigma/(\rho + \mu\sigma)\}$  with a coefficient  $\mu$  that will be specified later and apply Fact 1 (vii) to bound the type-I error  $\alpha$  and the type-II error  $\beta$ :

$$\begin{aligned} \alpha &= \operatorname{Tr} \left[ \rho \frac{\mu\sigma}{\rho + \mu\sigma} \right] \leq \operatorname{Tr} [\rho \wedge \mu\sigma], \\ \beta &= \operatorname{Tr} \left[ \sigma \frac{\rho}{\rho + \mu\sigma} \right] \leq \mu^{-1} \operatorname{Tr} [\rho \wedge \mu\sigma]. \end{aligned} \quad (21)$$

Next, we show how Eq. (21) implies both the small deviation type bound and the large deviation type bound. As in the proof of Proposition 2, we invoke the definition of “ $\wedge$ ” in Eq. (5) with any test  $T$  satisfying  $\operatorname{Tr} [\rho(\mathbb{1} - T)] \leq \varepsilon - \delta$ , i.e.,

$$\begin{aligned} \operatorname{Tr} [\rho \wedge \mu\sigma] &= \inf_{0 \leq T \leq \mathbb{1}} \operatorname{Tr} [\rho(\mathbb{1} - T)] + \operatorname{Tr} [\mu\sigma T] \\ &\leq \varepsilon - \delta + \mu e^{-D_h^{\varepsilon-\delta}(\rho \parallel \sigma)}. \end{aligned} \quad (22)$$

Choosing  $\mu$  such that the right-hand side of Eq. (22) equals  $\varepsilon$  and recalling the upper bound on “ $\wedge$ ” [Fact 1 (vi), taking

$s \rightarrow 0$ ] we obtain the following bound on the type-II error: for all  $0 < \delta < \varepsilon < 1$ ,

$$\beta \leq \mu^{-1} \operatorname{Tr} [\rho \wedge \mu\sigma] \leq \mu^{-1} = e^{-D_h^{\varepsilon-\delta}(\rho \parallel \sigma) - \log \delta}. \quad (23)$$

We remark that Eq. (23) is stronger than the analysis provided by Beigi and Gohari [91, Theorem 6] in view of the relation between the quantum hypothesis-testing divergence and the quantum information-spectrum divergence, Eq. (16). This again magnifies the fact that the pretty-good measurement yields a one-shot achievability bound on the *Stein exponent* [i.e., the maximal exponent of the type-II error provided that the type-I error is at most  $\varepsilon \in (0, 1)$ ], and it can achieve second-order asymptotics in the IID setting as well. Note here that, since the pretty-good measurement is suboptimal, it incurs a cost  $-\log 1/\delta$  on the Stein exponent in the one-shot setting and a third-order term  $-\frac{1}{2} \log n$  in the  $n$ -fold IID scenario. Yet, it is still sufficient to achieve moderate deviation asymptotics [42] (i.e., the inferior third-order term  $-\frac{1}{2} \log n$  does not affect the moderate deviation expansion).

Next, we show that the pretty-good measurement can recover the *quantum Hoeffding bound* [34, Sec. 5.5]. Applying the upper bound on “ $\wedge$ ” [Fact 1 (vi) with  $\alpha = 1 - s \in (0, 1)$ ] in Eq. (21), we obtain

$$\begin{aligned} \alpha &\leq \operatorname{Tr} [\rho \wedge \mu\sigma] \leq \mu^{1-\alpha} \operatorname{Tr} [\rho^\alpha \sigma^{1-\alpha}], \\ \beta &\leq \mu^{-1} \operatorname{Tr} [\rho \wedge \mu\sigma] \leq \mu^{-\alpha} \operatorname{Tr} [\rho^\alpha \sigma^{1-\alpha}]. \end{aligned}$$

Choosing  $\mu = e^{[(\alpha-1)/\alpha]D_\alpha(\rho \parallel \sigma) + r/\alpha}$  with the quantum Petz-Rényi divergence  $D_\alpha$  introduced in Proposition 1, we arrive at the one-shot quantum Hoeffding bound: for all  $r > 0$  and  $\alpha \in (0, 1)$ ,

$$\alpha \leq e^{-[(1-\alpha)/\alpha](D_\alpha(\rho \parallel \sigma) - r)}, \quad \beta \leq e^{-r}.$$

To the best of our knowledge, this is the first time the quantum Hoeffding bound has been achieved using the pretty-good measurement.

## B. Entanglement-assisted classical communication over quantum channels

In this section, we elaborate on how the achievability of entanglement-assisted (EA) classical communication [56, 115–119] follows in the same fashion from the proposed simple derivation in Sec. III.

*Definition 2 (Entanglement-assisted classical communication over quantum channels).*—Let  $\mathcal{N}_{\mathbf{A} \rightarrow \mathbf{B}}$  be a quantum channel.

- (1) Alice holds a classical register  $\mathbf{M}$  and quantum registers  $\mathbf{A}$  and  $\mathbf{A}'$ , and Bob holds quantum registers  $\mathbf{B}$  and  $\mathbf{B}'$ .

- (2) A resource of an arbitrary state  $\theta_{R'A'}$  is shared between Bob and Alice beforehand.
- (3) For any (equiprobable) message  $m \in \mathbf{M}$  Alice wanted to send, she performs an encoding quantum operation  $\mathcal{E}_{A' \rightarrow A}^m$  on  $\theta_{R'A'}$ .
- (4) The quantum channel  $\mathcal{N}_{A \rightarrow B}$  is applied on Alice's quantum register  $\mathbf{A}$  and outputs a state on Bob's quantum register  $\mathbf{B}$ .
- (5) Bob performs a decoding measurement  $\{\Pi_{R'B}^m\}_{m \in \mathbf{M}}$  on registers  $\mathbf{R}'$  and  $\mathbf{B}$  to extract the sent message  $m$ .

An  $(M, \varepsilon)$ -EA code for  $\mathcal{N}_{A \rightarrow B}$  is a protocol such that  $|\mathbf{M}| = M$  and the average error probability satisfies

$$\frac{1}{M} \sum_{m \in \mathbf{M}} \text{Tr}[(\mathbb{1} - \Pi_{R'B}^m) \mathcal{N}_{A \rightarrow B} \circ \mathcal{E}_{A' \rightarrow A}^m(\theta_{R'A'})] \leq \varepsilon.$$

We adopt the encoder of the *position-based coding* [43], but with the pretty-good measurement as the decoder.

- (a) **Preparations.** Alice and Bob preshare an  $M$ -fold product state  $\theta_{R'A'} := \theta_{RA}^{\otimes M} = \theta_{R_1 A_1} \otimes \cdots \otimes \theta_{R_M A_M}$ .
- (b) **Encoding.** For sending each  $m \in \mathbf{M}$ , Alice simply sends her system  $\mathbf{A}_m$ , i.e.,  $\mathcal{E}_{A' \rightarrow A}^m = \text{Tr}_{\mathbf{A} \setminus \{m\}}$ , for tracing out systems  $\mathbf{A}_{\bar{m}}$  for all  $\bar{m} \neq m$ .
- (c) **Decoding.** At the receiver, the channel output states for all  $m \in \mathbf{M}$  are

$$\rho_{R^M B}^m := \theta_R^{\otimes(M-1)} \otimes \mathcal{N}_{A \rightarrow B}(\theta_{R_m A_m}) \otimes \theta_R^{\otimes(M-m)}. \quad (24)$$

Then, Bob performs the pretty-good measurement with respect to the channel output states:

$$\Pi_{R^M B}^m := \frac{\rho_{R^M B}^m}{\sum_{\bar{m} \in \mathbf{M}} \rho_{R^M B}^{\bar{m}}} \quad \text{for all } m \in \mathbf{M}.$$

Note that the decoding part constitutes the main difference from previous results, such as the original position-based coding [43,60] based on the Hayashi-Nagaoka operator inequality [30, Lemma 2], the sequential decoding strategy [27] with an auxiliary probe system, and a quantum union bound [29, Theorem 2.1] [120].

Below, we analyze the conditional error probability for sending each message  $m \in \mathbf{M}$ . Let  $\text{Tr}_{R^M \setminus \{m\}}$  be the partial trace for tracing out systems  $\mathbf{R}_{\bar{m}}$  for all  $\bar{m} \neq m$ , except  $\mathbf{R}_m$ . By Eq. (24), we have the following identities: for all  $m \in \mathbf{M}$  and all  $\bar{m} \neq m$ ,

$$\text{Tr}_{R^M \setminus \{m\}}[\rho_{R^M B}^m] = \mathcal{N}_{A \rightarrow B}(\theta_{R_m A_m}) = \mathcal{N}_{A \rightarrow B}(\theta_{RA}),$$

$$\text{Tr}_{R^M \setminus \{m\}}[\rho_{R^M B}^{\bar{m}}] = \theta_{R_m} \otimes \mathcal{N}_{A \rightarrow B}(\theta_{A_m}) = \theta_R \otimes \mathcal{N}_{A \rightarrow B}(\theta_A).$$

Then, the error probability conditioned on sending each message  $m \in \mathbf{M}$  is

$$\begin{aligned} & \text{Tr} \left[ \rho_{R^M B}^m \frac{\sum_{\bar{m} \neq m} \rho_{R^M B}^{\bar{m}}}{\rho_{R^M B}^m + \sum_{\bar{m} \neq m} \rho_{R^M B}^{\bar{m}}} \right] \\ & \stackrel{(a)}{\leq} \text{Tr} \left[ \rho_{R^M B}^m \wedge \left( \sum_{\bar{m} \neq m} \rho_{R^M B}^{\bar{m}} \right) \right] \\ & = \text{Tr} \left[ \text{Tr}_{R^M \setminus \{m\}} \left[ \rho_{R^M B}^m \wedge \left( \sum_{\bar{m} \neq m} \rho_{R^M B}^{\bar{m}} \right) \right] \right] \\ & \stackrel{(b)}{\leq} \text{Tr} \left[ \left( \text{Tr}_{R^M \setminus \{m\}}[\rho_{R^M B}^m] \right) \wedge \left( \sum_{\bar{m} \neq m} \text{Tr}_{R^M \setminus \{m\}}[\rho_{R^M B}^{\bar{m}}] \right) \right] \\ & = \text{Tr}[\mathcal{N}_{A \rightarrow B}(\theta_{RA}) \wedge (M-1)\theta_R \otimes \mathcal{N}_{A \rightarrow B}(\theta_A)], \quad (25) \end{aligned}$$

where, as in the proof of Theorem 1, (a) follows from the lower bound of the noncommutative minimal, Fact 1 (vii), and (b) is due to the monotonicity of the noncommutative minimal under positive trace-preserving maps, Fact 1 (iii). Hence, we establish the following one-shot achievability for entanglement-assisted classical communication over quantum channels.

*Theorem 2 (A one-shot achievability bound for EA classical communication over quantum channels).—*Consider an arbitrary quantum channel  $\mathcal{N}_{A \rightarrow B}$ . Then, there exists an  $(M, \varepsilon)$ -EA code for  $\mathcal{N}_{A \rightarrow B}$  such that, for any density operator  $\theta_{RA}$ ,

$$\varepsilon \leq \text{Tr}[\mathcal{N}_{A \rightarrow B}(\theta_{RA}) \wedge (M-1)\theta_R \otimes \mathcal{N}_{A \rightarrow B}(\theta_A)].$$

*Remark 7.*—The above derivations reemphasize the central idea of the position-based coding proposed by Anshu *et al.* [43]. Namely, the preshared entanglement  $\theta_{R^M A^M} = \theta_{RA}^{\otimes M}$  along with the encoding  $m \mapsto \rho_{R^M B}^m = \mathcal{N}_{A \rightarrow B}(\theta_{R^M A_m})$  ensure the *mutual independence* between each subsystem  $\mathbf{R}_m \mathbf{A}_m$ ,  $m \in \mathbf{M}$ , and, accordingly,  $\text{Tr}_{R^M \setminus \{m\}}[\rho_{R^M B}^{\bar{m}}] = \theta_R \otimes \mathcal{N}_{A \rightarrow B}(\theta_A)$  for all  $\bar{m} \neq m$ . Here, the partial trace  $\text{Tr}_{R^M \setminus \{m\}}$  may be considered as an *expectation conditioned on  $m$*  (see Remark 8 below for a detailed discussion). Such independence between register  $\mathbf{R}_m$  associated with each channel output state thus plays the same role as the independent random codebook used in classical-quantum channel coding (Theorem 1). On the other hand, we would like to point out that, normally, a communication system operates on a message set whose size is exponentially large, i.e.,  $M \geq e^{nI(\mathbf{R}:\mathbf{B})_{\mathcal{N}(\theta)} - O(n)}$ . Preparing exponentially many copies of the preshared state  $\theta_{RA}^{\otimes M}$  might be practically challenging. (Note that even in the classical case, resources required to generate mutual independence among an exponentially large set could not be considered as nonexpensive [121, Corollary 3.34]). Nevertheless, Anshu *et al.* [43, Sec. IV] adapted an *entanglement recycling* technique by Strelchuk *et al.* [122] to reduce the required amount of entanglement resource.

From our analysis given above, only *pairwise independence* among each subsystem  $\mathbf{R}_m \mathbf{A}_m$ ,  $m \in \mathbf{M}$ , is needed. That is, we only require that  $\text{Tr}_{\mathbf{R}^{\mathbf{M} \setminus \{m, \bar{m}\}} \mathbf{A}^{\mathbf{M} \setminus \{m, \bar{m}\}}} [\theta_{\mathbf{R}^{\mathbf{M}} \mathbf{A}^{\mathbf{M}}}] = \theta_{\mathbf{R}_m \mathbf{A}_m} \otimes \theta_{\mathbf{R}_{\bar{m}} \mathbf{A}_{\bar{m}}}$  for each  $m \neq \bar{m}$ . This point of view may provide another angle to reduce the required entanglement resource. Though, to the best of our knowledge, its explicit construction is not clear in noncommutative probability space. We leave this for future work.

*Remark 8.*—The analysis of Theorem 2 actually shares the same flavor as that of Theorem 1. More precisely, the partial trace  $\text{Tr}_{\mathbf{R}^{\mathbf{M} \setminus \{m\}}}$  in step (b) of Eq. (25) plays the same role as the averaging over the random codebook in step (a) of Eq. (10). In other words, the partial trace  $\text{Tr}_{\mathbf{R}^{\mathbf{M} \setminus \{m\}}}$  can be interpreted as a *conditional expectation* [123–126] (which is a completely positive and trace-preserving map) from the operator algebra of bounded operators on  $\mathbf{R}^{\mathbf{M}} \mathbf{B}$ , i.e.,  $\mathcal{B}(\mathbf{R}^{\mathbf{M}} \mathbf{B})$ , to its subalgebra [127]  $\mathbb{1}_{\mathbf{R}^{m-1}} \otimes \mathcal{B}(\mathbf{R}_m \mathbf{B}) \otimes \mathbb{1}_{\mathbf{R}^{M-m}}$ .

Directly applying the pretty-good measurement as above allows us to obtain a tighter and cleaner one-shot achievability bound in a more general form. This then revisits the *position-based coding* proposed by Anshu *et al.* [43, Lemma 4]. We summarize it as the following *one-shot quantum packing lemma* that is not only prominent to Theorems 1 and 2, and all the forthcoming results in this section, but we believe that it is applicable elsewhere in quantum information theory as well.

*Theorem 3 (A one-shot quantum packing lemma).*—Let  $\rho_{\mathbf{R} \mathbf{B}}$  and  $\tau_{\mathbf{R}}$  be arbitrary density operators, and let  $M$  be an integer. For every  $m \in \mathbf{M} := \{1, \dots, M\}$ , define

$$\omega_{\mathbf{R}_1 \mathbf{R}_2 \dots \mathbf{R}_M \mathbf{B}}^m := \rho_{\mathbf{R}_m \mathbf{B}} \otimes \tau_{\mathbf{R}_1} \otimes \tau_{\mathbf{R}_2} \otimes \dots \otimes \tau_{\mathbf{R}_{m-1}} \otimes \tau_{\mathbf{R}_{m+1}} \otimes \dots \otimes \tau_{\mathbf{R}_M},$$

where  $\rho_{\mathbf{R}_m \mathbf{B}} = \rho_{\mathbf{R} \mathbf{B}}$  and  $\tau_{\mathbf{R}_m} = \tau_{\mathbf{R}}$  for every  $m \in \mathbf{M}$ . Then, there exists a measurement

$$\Pi_{\mathbf{R}_1 \mathbf{R}_2 \dots \mathbf{R}_M \mathbf{B}}^m := \frac{\omega_{\mathbf{R}_1 \mathbf{R}_2 \dots \mathbf{R}_M \mathbf{B}}^m}{\sum_{\bar{m} \in \mathbf{M}} \omega_{\mathbf{R}_1 \mathbf{R}_2 \dots \mathbf{R}_M \mathbf{B}}^{\bar{m}}} \quad \text{for all } m \in \mathbf{M}$$

satisfying, for every  $m \in \mathbf{M}$ ,

$$\text{Tr} [\omega_{\mathbf{R} \mathbf{B}}^m (\mathbb{1} - \Pi_{\mathbf{R} \mathbf{B}}^m)] \leq \text{Tr} [\rho_{\mathbf{R} \mathbf{B}} \wedge (M - 1) \tau_{\mathbf{R}} \otimes \rho_{\mathbf{B}}].$$

To see how the one-shot quantum packing lemma is applied to the previous achievability bounds, we make the following substitutions:  $\rho_{\mathbf{R}_m \mathbf{B}} \rightarrow \mathcal{N}_{\mathbf{A} \rightarrow \mathbf{B}}(\theta_{\mathbf{R}_m \mathbf{A}_m})$  and  $\tau_{\mathbf{R}_{\bar{m}}} \rightarrow \theta_{\mathbf{R}_{\bar{m}}}$  for all  $m \in \mathbf{M}$  and  $\bar{m} \neq m$ . Then, Theorem 3 covers Theorem 2 for entanglement-assisted classical communication over quantum channels.

On the other hand, in the scenario where  $\mathbf{R}_m \mathbf{B} \rightarrow \mathbf{X}_m \mathbf{B}$ ,  $\rho_{\mathbf{R}_m \mathbf{B}} \rightarrow \rho_{\mathbf{X}_m \mathbf{B}}$ , and  $\tau_{\mathbf{R}_{\bar{m}}} \rightarrow \rho_{\mathbf{X}_{\bar{m}}}$  for all  $m \in \mathbf{M}$  and  $\bar{m} \neq m$ , the setting in Theorem 3 corresponds to the *randomness-assisted communication over  $c$ - $q$  channels*, where the

$\mathbf{X}_m$  is the shared randomness at Bob and the joint state  $\rho_{\mathbf{X}_m \mathbf{B}}$  results from Alice sending her  $m$ th classical system through the  $c$ - $q$  channel (see also the papers by Wilde [59] and Anshu *et al.* [128]). Then, Theorem 3 yields the achievability bound on the average error probability over the ensemble of codes, i.e., the right-hand side of Eq. (8) in Theorem 1. Via derandomization, one can always claim the existence of a good code in the ensemble to achieve such an error bound without randomness assistance. This concludes the statement of Theorem 1 for  $c$ - $q$  channel coding [129].

Following the same reasoning as in Proposition 1, Theorem 2 (or Theorem 3) leads to a large deviation type bound, which is tighter than [60, Theorem 6] without a prefactor 4; following the same reasoning as in Proposition 2, Theorem 2 provides a tighter lower bound on the  $\varepsilon$ -one-shot entanglement-assisted capacity for  $\mathcal{N}_{\mathbf{A} \rightarrow \mathbf{B}}$  (i.e., the maximal logarithmic size of messages with average error probability below  $\varepsilon$ ) than Refs. [29, Theorem 5.1], [43, Theorem 1], and [60, Theorem 8] (with the same improvements as the comparison made in Sec. III A).

*Proposition 3 (Bounding the coding error given a fixed coding rate).*—Consider an arbitrary quantum channel  $\mathcal{N}_{\mathbf{A} \rightarrow \mathbf{B}}$ . Then, for any  $R > 0$ , there exists an  $(e^R, \varepsilon)$ -EA code for  $\mathcal{N}_{\mathbf{A} \rightarrow \mathbf{B}}$  such that, for any  $\theta_{\mathbf{R} \mathbf{A}}$ ,

$$\varepsilon \leq e^{-[(1-\alpha)/\alpha](I_{2-1/\alpha}^{\downarrow}(\mathbf{R} : \mathbf{B})_{\mathcal{N}(\theta)} - R)} \quad \text{for all } \alpha \in (\frac{1}{2}, 1).$$

Here, we follow the notation given in Proposition 1.

*Proposition 4 (Bounding the coding rate given a fixed coding error).*—Consider an arbitrary quantum channel  $\mathcal{N}_{\mathbf{A} \rightarrow \mathbf{B}}$ . Then, for any  $\varepsilon \in (0, 1)$ , there exists an  $(M, \varepsilon)$ -EA code for  $\mathcal{N}_{\mathbf{A} \rightarrow \mathbf{B}}$  such that, for any  $\theta_{\mathbf{R} \mathbf{A}}$  and any  $\delta \in (0, \varepsilon)$ ,

$$\log M \geq D_h^{\varepsilon-\delta}(\mathcal{N}_{\mathbf{A} \rightarrow \mathbf{B}}(\theta_{\mathbf{R} \mathbf{A}}) \parallel \theta_{\mathbf{R}} \otimes \mathcal{N}_{\mathbf{A} \rightarrow \mathbf{B}}(\theta_{\mathbf{A}})) - \log \frac{1}{\delta}.$$

Here, we follow the notation given in Proposition 2.

### C. Classical data compression with quantum side information

In this section, we show that how the proposed method in Sec. III can be applied to classical data compression with quantum side information [80, 113, 130–132]. In the following, we refer to such an protocol as CQSW.

*Definition 3 (Classical data compression with quantum side information).*—Let  $\rho_{\mathbf{X} \mathbf{B}} = \sum_{x \in \mathbf{X}} p_{\mathbf{X}}(x) |x\rangle\langle x| \otimes \rho_{\mathbf{B}}^x$  be a classical-quantum state.

- (1) Alice holds classical registers  $\mathbf{X}$  and  $\mathbf{M}$ , and Bob holds a quantum register  $\mathbf{B}$ .
- (2) Alice performs an encoding  $\mathcal{E} : \mathbf{X} \rightarrow \mathbf{M}$  that compresses the source in  $\mathbf{X}$  to an index in  $\mathbf{M}$ .
- (3) Bob performs a decoding measurement described by a family of POVMs indexed by  $m \in \mathbf{M}$ , i.e.,  $\{\Pi_{\mathbf{B}}^{x,m}\}_{x \in \mathbf{X}}$  on register  $\mathbf{B}$ , to recover the source  $x \in \mathbf{X}$ .

An  $(M, \varepsilon)$ -CQSW code for  $\rho_{\mathbf{XB}}$  is a protocol such that  $|\mathbf{M}| = M$  and the error probability satisfies

$$\sum_{x \in \mathbf{X}} p_{\mathbf{X}}(x) \text{Tr}[\rho_{\mathbf{B}}^x (\mathbb{1}_{\mathbf{B}} - \Pi_{\mathbf{B}}^{x, \mathcal{E}(x)})] \leq \varepsilon.$$

Without loss of generality, we assume that the prior distribution of the source,  $p_{\mathbf{X}}$ , has full support for brevity. We also adopt the standard random coding strategy given in Sec. III.

- (a) **Encoding.** The encoder maps each  $x \in \mathbf{X}$  pairwise independently to uniform index  $m \in \mathbf{M}$ .
- (b) **Decoding.** We use the following pretty-good measurement (again given the realization of the above encoding):

$$\Pi_{\mathbf{B}}^{x,m} := \frac{p_{\mathbf{X}}(x) \rho_{\mathbf{B}}^x}{\sum_{\bar{x}: \mathcal{E}(\bar{x})=m} p_{\mathbf{X}}(\bar{x}) \rho_{\mathbf{B}}^{\bar{x}}} \quad \text{for all } x \in \mathbf{X}, m \in \mathbf{M}. \quad (26)$$

*Theorem 4 (A one-shot achievability bound for classical data compression with quantum side information).—* Consider an arbitrary classical-quantum state  $\rho_{\mathbf{XB}} = \sum_{x \in \mathbf{X}} p_{\mathbf{X}}(x) |x\rangle\langle x| \otimes \rho_{\mathbf{B}}^x$ . Then, there exists an  $(M, \varepsilon)$ -CQSW code for  $\rho_{\mathbf{XB}}$  such that

$$\varepsilon \leq \text{Tr} \left[ \rho_{\mathbf{XB}} \wedge \frac{1}{M} \mathbb{1}_{\mathbf{X}} \otimes \rho_{\mathbf{B}} \right].$$

*Proof.*—We use the pretty-good measurement given in Eq. (26) to calculate the expected error probability (over the random encoding):

$$\begin{aligned} & \mathbb{E}_{x \sim p_{\mathbf{X}}} \mathbb{E}_{m \sim \frac{1}{M}} \text{Tr} \left[ \rho_{\mathbf{B}}^x \frac{\sum_{\bar{x} \neq x, \mathcal{E}(\bar{x})=m} p_{\mathbf{X}}(\bar{x}) \rho_{\mathbf{B}}^{\bar{x}}}{\sum_{\bar{x}: \mathcal{E}(\bar{x})=m} p_{\mathbf{X}}(\bar{x}) \rho_{\mathbf{B}}^{\bar{x}}} \right] \\ & \stackrel{(a)}{\leq} \mathbb{E}_{x \sim p_{\mathbf{X}}} \mathbb{E}_{m \sim \frac{1}{M}} \text{Tr} \left[ \rho_{\mathbf{B}}^x \wedge \left( \sum_{\bar{x} \neq x, \mathcal{E}(\bar{x})=m} \frac{p_{\mathbf{X}}(\bar{x})}{p_{\mathbf{X}}(x)} \rho_{\mathbf{B}}^{\bar{x}} \right) \right] \\ & \stackrel{(b)}{\leq} \mathbb{E}_{x \sim p_{\mathbf{X}}} \text{Tr} \left[ \rho_{\mathbf{B}}^x \wedge \mathbb{E}_{m \sim \frac{1}{M}} \left[ \sum_{\bar{x} \neq x} \mathbf{1}_{\{\mathcal{E}(\bar{x})=m\}} \frac{p_{\mathbf{X}}(\bar{x})}{p_{\mathbf{X}}(x)} \rho_{\mathbf{B}}^{\bar{x}} \right] \right] \\ & \stackrel{(c)}{=} \mathbb{E}_{x \sim p_{\mathbf{X}}} \text{Tr} \left[ \rho_{\mathbf{B}}^x \wedge \left( \frac{1}{M} \sum_{\bar{x} \neq x} \frac{p_{\mathbf{X}}(\bar{x})}{p_{\mathbf{X}}(x)} \rho_{\mathbf{B}}^{\bar{x}} \right) \right] \\ & \stackrel{(d)}{\leq} \mathbb{E}_{x \sim p_{\mathbf{X}}} \text{Tr} \left[ \rho_{\mathbf{B}}^x \wedge \left( \frac{1}{M p_{\mathbf{X}}(x)} \rho_{\mathbf{B}} \right) \right] \\ & \stackrel{(e)}{\leq} \text{Tr} \left[ \rho_{\mathbf{XB}} \wedge \left( \frac{1}{M} \mathbb{1}_{\mathbf{X}} \otimes \rho_{\mathbf{B}} \right) \right]. \end{aligned}$$

Here (a) uses the lower bound of the noncommutative minimal given in Fact 1 (vii); (b) follows from the concavity given in Fact 1 (iv); (c) follows from the pairwise independent and uniform random encoding; (d) follows

from the monotone increase in the Loewner ordering and  $\sum_{\bar{x} \neq x} p_{\mathbf{X}}(\bar{x}) \rho_{\mathbf{B}}^{\bar{x}} \leq \sum_{\bar{x}} p_{\mathbf{X}}(\bar{x}) \rho_{\mathbf{B}}^{\bar{x}} = \rho_{\mathbf{B}}$ ; and, lastly, (e) follows from the direct sum formula given in Fact 1 (v). ■

Using the same reasoning as in Propositions 1 and 2 of Sec. III, we have the following one-shot bounds for CQSW.

*Proposition 5 (Bounding the coding error given a fixed coding rate).—* Consider an arbitrary classical-quantum state  $\rho_{\mathbf{XB}} = \sum_{x \in \mathbf{X}} p_{\mathbf{X}}(x) |x\rangle\langle x| \otimes \rho_{\mathbf{B}}^x$ . Then, for any  $R > 0$ , there exists an  $(e^R, \varepsilon)$ -CQSW code for  $\rho_{\mathbf{XB}}$  such that

$$\varepsilon \leq e^{-[(1-\alpha)/\alpha](R-H_{2^{-1/\alpha}}^{\downarrow}(\mathbf{X}|\mathbf{B})_{\rho})} \quad \text{for all } \alpha \in \left(\frac{1}{2}, 1\right),$$

where  $H_{\alpha}^{\downarrow}(\mathbf{X}|\mathbf{B})_{\rho} := -D_{\alpha}(\rho_{\mathbf{XB}} \| \mathbb{1}_{\mathbf{X}} \otimes \rho_{\mathbf{B}})$ .

*Proposition 6 (Bounding the coding rate given a fixed coding error).—* Consider an arbitrary classical-quantum state  $\rho_{\mathbf{XB}} = \sum_{x \in \mathbf{X}} p_{\mathbf{X}}(x) |x\rangle\langle x| \otimes \rho_{\mathbf{B}}^x$ . Then, for any  $\varepsilon \in (0, 1)$ , there exists an  $(M, \varepsilon)$ -CQSW code for  $\rho_{\mathbf{XB}}$  such that, for any  $\delta \in (0, \varepsilon)$ ,

$$\log M \leq H_h^{\varepsilon-\delta}(\mathbf{X}|\mathbf{B})_{\rho} + \log \frac{1}{\delta},$$

where  $H_h^{\varepsilon-\delta}(\mathbf{X}|\mathbf{B})_{\rho} := -D_h^{\varepsilon-\delta}(\rho_{\mathbf{XB}} \| \mathbb{1}_{\mathbf{X}} \otimes \rho_{\mathbf{B}})$ .

## D. Multiple-access channel coding

In this section, we show one-shot achievability bounds for classical-quantum multiple-access channel (MAC) coding and entanglement-assisted classical communication over quantum MACs [60,133–135]. Note that the former naturally extends to (unassisted) classical communication over quantum MACs. We present the scenario for only two senders with one receiver; the result applies to multiple senders in the same fashion.

*Definition 4 (Classical-quantum multiple-access channel coding).—* Let  $\mathcal{N}_{\mathbf{XY} \rightarrow \mathbf{C}} : (x, y) \mapsto \rho_{\mathbf{C}}^{x,y}$  be a classical-quantum multiple-access channel.

- (1) Alice holds classical registers  $\mathbf{M}_A$  and  $\mathbf{X}$ , Bob holds  $\mathbf{M}_B$  and  $\mathbf{Y}$ , and Charlie holds quantum register  $\mathbf{C}$ .
- (2) Alice performs an encoding  $m_A \mapsto x(m_A) \in \mathbf{X}$  for any equiprobable message  $m_A \in \mathbf{M}_A$  she wanted to send. Bob performs an encoding  $m_B \mapsto y(m_B) \in \mathbf{Y}$  for any equiprobable message  $m_B \in \mathbf{M}_B$  he wanted to send.
- (3) The channel  $\mathcal{N}_{\mathbf{XY} \rightarrow \mathbf{C}}$  is applied on Alice's and Bob's registers  $\mathbf{X}$  and  $\mathbf{Y}$  and outputs a state on  $\mathbf{C}$  at Charlie.
- (4) A decoding measurement  $\{\Pi_{\mathbf{C}}^{m_A, m_B}\}_{(m_A, m_B) \in \mathbf{M}_A \times \mathbf{M}_B}$  is performed on register  $\mathbf{C}$  to extract the sent message  $(m_A, m_B)$ .

An  $(M_A, M_B, \varepsilon)$  code for  $\mathcal{N}_{\mathbf{XY} \rightarrow \mathbf{C}}$  is a protocol such that  $|\mathbf{M}_A| = M_A$ ,  $|\mathbf{M}_B| = M_B$ , and the average error probability

satisfies

$$\frac{1}{M_A M_B} \sum_{(m_A, m_B) \in M_A \times M_B} \text{Tr}[(1 - \Pi_C^{m_A, m_B}) \rho_C^{x(m_A), y(m_B)}] \leq \varepsilon.$$

We follow the strategy presented in Sec. III.

- (a) **Encoding.** Each pair of messages  $(m_A, m_B) \in M_A \times M_B$  is mapped to a codeword  $(x(m_A), y(m_B)) \in X \times Y$  pairwise independently according to some probability distribution  $p_X \otimes p_Y$ .
- (b) **Decoding.** We use the pretty-good measurement with respect to the corresponding channel output states (given the realization of the random codebook): for  $(m_A, m_B) \in M_A \times M_B$ ,

$$\Pi_C^{m_A, m_B} = \frac{\rho_C^{x(m_A), y(m_B)}}{\sum_{(\bar{m}_A, \bar{m}_B)} \rho_C^{x(\bar{m}_A), y(\bar{m}_B)}}.$$

Then, we obtain the following result (without duplicating the proof).

*Theorem 5 (A one-shot achievability bound for classical-quantum MAC coding).*—Consider an arbitrary classical-quantum multiple-access channel  $\mathcal{N}_{XY \rightarrow C}: (x, y) \mapsto \rho_C^{x, y}$ . Then, there exists an  $(M_A, M_B, \varepsilon)$  code for  $\mathcal{N}_{XY \rightarrow C}$  such that, for any probability distributions  $p_X$  and  $p_Y$ ,

$$\varepsilon \leq \text{Tr}[\rho_{XYC} \wedge ((M_A - 1)\rho_X \otimes \rho_{YC} + (M_B - 1)\rho_Y \otimes \rho_{XC} + (M_A - 1)(M_B - 1)\rho_X \otimes \rho_Y \otimes \rho_C)],$$

where  $\rho_{XYC} := \sum_{(x, y) \in X \times Y} p_X(x)|x\rangle\langle x| \otimes p_Y(y)|y\rangle\langle y| \otimes \rho_C^{x, y}$ .

Next, we consider the entanglement-assisted setting.

*Definition 5 (Entanglement-assisted classical communication over quantum multiple-access channels).*—Let  $\mathcal{N}_{AB \rightarrow C}$  be a quantum multiple-access channel.

- (1) Alice holds classical register  $M_A$  and quantum registers  $A$  and  $A'$ . Bob holds classical register  $M_B$  and quantum registers  $B$  and  $B'$ . Charlie holds quantum registers  $C$ ,  $R'_A$ , and  $R'_B$ .
- (2) Charlie and Alice share an arbitrary state  $\theta_{R'_A A'}$ . Charlie and Bob share an arbitrary state  $\theta_{R'_B B'}$ .
- (3) Alice performs an encoding  $\mathcal{E}_{A' \rightarrow A}^{m_A}$  on  $\theta_{R'_A A'}$  for any equiprobable message  $m_A \in M_A$  she wanted to send; Bob performs an encoding  $\mathcal{E}_{B' \rightarrow B}^{m_B}$  on  $\theta_{R'_B B'}$  for any equiprobable message  $m_B \in M_B$  he wanted to send.
- (4) The channel  $\mathcal{N}_{AB \rightarrow C}$  is applied on Alice's and Bob's registers  $A$  and  $B$ , and outputs a state on quantum register  $C$  at Charlie.
- (5) Charlie performs a decoding measurement  $\{\Pi_{R'_A R'_B C}^{m_A, m_B}\}_{(m_A, m_B) \in M_A \times M_B}$  on registers  $R'_A$ ,  $R'_B$ , and  $C$  to extract the sent message  $(m_A, m_B)$ .

An  $(M_A, M_B, \varepsilon)$ -EA code for  $\mathcal{N}_{AB \rightarrow C}$  is a protocol such that  $|M_A| = M_A$ ,  $|M_B| = M_B$ , and the average error probability satisfies

$$\frac{1}{M_A M_B} \sum_{(m_A, m_B) \in M_A \times M_B} \text{Tr}[(1 - \Pi_{R'_A R'_B C}^{m_A, m_B}) \mathcal{N}_{AB \rightarrow C}(\mathcal{E}_{A' \rightarrow A}^{m_A}(\theta_{R'_A A'}) \otimes \mathcal{E}_{B' \rightarrow B}^{m_B}(\theta_{R'_B B'}))] \leq \varepsilon.$$

As in Sec. IV B, we use the encoder of the position-based coding (see also Ref. [60]) and the pretty-good measurement for decoding.

- (a) **Preparation.** The  $M_A$ -fold product states  $\theta_{R'_A A'} := \theta_{R'_A A}^{\otimes M_A}$  are shared between Charlie and Alice, and the  $M_B$ -fold product states  $\theta_{R'_B B'} := \theta_{R'_B B}^{\otimes M_B}$  are shared between Charlie and Bob.
- (b) **Encoding.** Alice adopts encoding  $\mathcal{E}_{A' \rightarrow A}^{m_A} = \text{Tr}_{A \setminus \{m_A\}}$  for each  $m_A \in M_A$ , and Bob adopts encoding  $\mathcal{E}_{B' \rightarrow B}^{m_B} = \text{Tr}_{B \setminus \{m_B\}}$  for each  $m_B \in M_B$ .
- (c) **Decoding.** Denote the corresponding channel output state for sending message  $(m_A, m_B) \in M_A \times M_B$  by

$$\rho_{R'_A R'_B C}^{m_A, m_B} := \theta_{R'_A}^{\otimes (M_A - 1)} \otimes \theta_{R'_B}^{\otimes (M_B - 1)} \otimes \mathcal{N}_{AB \rightarrow C}(\theta_{R'_A A} \otimes \theta_{R'_B B}) \otimes \theta_{R'_A}^{\otimes (M_A - m_A)} \otimes \theta_{R'_B}^{\otimes (M_B - m_B)}.$$

Then, we use the associated pretty-good measurement: for all  $(m_A, m_B) \in M_A \times M_B$ ,

$$\Pi_{R'_A R'_B C}^{m_A, m_B} = \frac{\rho_{R'_A R'_B C}^{m_A, m_B}}{\sum_{(\bar{m}_A, \bar{m}_B)} \rho_{R'_A R'_B C}^{\bar{m}_A, \bar{m}_B}}.$$

Following the analysis presented in Sec. IV B, we immediately obtain the following result (without duplicating the proof).

*Theorem 6 (A one-shot achievability bound for EA classical communication over quantum MAC).*—Consider an arbitrary quantum multiple-access channel  $\mathcal{N}_{AB \rightarrow C}$ . Then, there exists an  $(M_A, M_B, \varepsilon)$ -EA code for  $\mathcal{N}_{AB \rightarrow C}$  such that, for any  $\theta_{R'_A A}$  and  $\theta_{R'_B B}$ ,

$$\varepsilon \leq \text{Tr}[\rho_{R'_A R'_B C} \wedge ((M_A - 1)\rho_{R'_A} \otimes \rho_{R'_B C} + (M_B - 1)\rho_{R'_B} \otimes \rho_{R'_A C} + (M_A - 1)(M_B - 1)\rho_{R'_A} \otimes \rho_{R'_B} \otimes \rho_C)],$$

where  $\rho_{R'_A R'_B C} := \mathcal{N}_{AB \rightarrow C}(\theta_{R'_A A} \otimes \theta_{R'_B B})$ .

## E. Broadcast channel coding

In this section, we study entanglement-assisted and unassisted classical communication over quantum broadcast channels [43, 117, 128, 136–141].



*Definition 6 (Classical-quantum broadcast channel coding).*—Let  $\mathcal{N}_{X \rightarrow BC} : x \mapsto \rho_{BC}^x$  be a classical-quantum broadcast channel.

- (1) Alice holds classical registers  $M_B$ ,  $M_C$ , and  $X$ . Bob holds quantum register  $B$  and Charlie holds quantum register  $C$ .
- (2) Alice performs an encoding  $(m_B, m_C) \mapsto x(m_B, m_C) \in X$  for any equiprobable message  $(m_B, m_C) \in M_B \times M_C$  she wanted to send to Bob and Charlie, respectively.
- (3) The channel  $\mathcal{N}_{X \rightarrow BC}$  is applied on Alice's register  $X$ , and outputs a marginal state on  $B$  at Bob and a marginal state on  $C$  at Charlie.
- (4) Bob performs a decoding measurement  $\{\Pi_B^{m_B}\}_{m_B \in M_B}$  on register  $B$  to extract the sent message  $m_B$ , and Charlie performs a decoding measurement  $\{\Pi_C^{m_C}\}_{m_C \in M_C}$  on register  $C$  to extract the sent message  $m_C$ .

An  $(M_B, M_C, \varepsilon_B, \varepsilon_C)$  code for  $\mathcal{N}_{X \rightarrow BC} : x \mapsto \rho_{BC}^x$  is a protocol such that  $|M_B| = M_B$ ,  $|M_C| = M_C$ , and the average error probabilities satisfy

$$\frac{1}{M_B M_C} \sum_{(m_B, m_C) \in M_B \times M_C} \text{Tr}[(\mathbb{1} - \Pi_B^{m_B}) \rho_B^{x(m_B, m_C)}] \leq \varepsilon_B,$$

$$\frac{1}{M_B M_C} \sum_{(m_B, m_C) \in M_B \times M_C} \text{Tr}[(\mathbb{1} - \Pi_C^{m_C}) \rho_C^{x(m_B, m_C)}] \leq \varepsilon_C.$$

We follow the analysis proposed in Sec. III by considering communication from Alice to Bob and from Alice to Charlie, separately.

- (a) **Encoding.** We introduce two auxiliary classical registers  $U$  and  $V$  for precoding. Message  $m_B \in M_B$  for Bob is encoded to a precodeword  $u(m_B)$  pairwise independently according to some probability distribution  $p_U$ ; message  $m_C \in M_C$  for Charlie is encoded to a precodeword  $v(m_C)$  pairwise independently according to some probability distribution  $p_V$ . Then, Alice picks a (deterministic) encoding  $(u(m_B), v(m_C)) \mapsto x(u(m_B), v(m_C)) \in X$ .
- (b) **Decoding.** Denoting (with a slight abuse of notation) the marginal channel output states at Bob by  $\rho_B^{m_B} := (1/M_C) \sum_{m_C \in M_C} \rho_B^{x(u(m_B), v(m_C))}$ ,  $m_B \in M_B$ , Bob performs the corresponding pretty-good measurement:

$$\Pi_B^{m_B} := \frac{\rho_B^{m_B}}{\sum_{\bar{m}_B \in M_B} \rho_B^{\bar{m}_B}}, \quad m_B \in M_B.$$

Similarly, denoting (with a slight abuse of notation again) the marginal channel output states at Charlie by  $\rho_C^{m_C} := (1/M_B) \sum_{m_B \in M_B} \rho_C^{x(u(m_B), v(m_C))}$ ,  $m_C \in M_C$ ,

$M_C$ , Charlie performs the corresponding pretty-good measurement:

$$\Pi_C^{m_C} := \frac{\rho_C^{m_C}}{\sum_{\bar{m}_C \in M_C} \rho_C^{\bar{m}_C}}, \quad m_C \in M_C.$$

Then, we obtain the following result (without duplicating the proof).

*Theorem 7 (A one-shot achievability bound for classical-quantum broadcast channel coding).*—Consider an arbitrary classical-quantum broadcast channel  $\mathcal{N}_{X \rightarrow BC} : x \mapsto \rho_{BC}^x$ . Then, there exists an  $(M_B, M_C, \varepsilon_B, \varepsilon_C)$  code for  $\mathcal{N}_{X \rightarrow BC}$  such that, for any probability distributions  $p_U$  and  $p_V$ , and (deterministic) encoding  $(u, v) \mapsto x(u, v)$ ,

$$\varepsilon_B \leq \text{Tr}[\rho_{UB} \wedge (M_B - 1) \rho_U \otimes \rho_B],$$

$$\varepsilon_C \leq \text{Tr}[\rho_{UC} \wedge (M_C - 1) \rho_V \otimes \rho_C],$$

where  $\rho_{UVBC} := \sum_{(u,v) \in U \times V} p_U(u) |u\rangle \langle u| \otimes p_V(v) |v\rangle \langle v| \otimes \rho_{BC}^{x(u,v)}$ .

Note that Theorem 7 extends to classical communication over quantum broadcast channels straightforwardly (see Table I).

*Remark 9.*—Theorem 7 employs independent precoding  $p_U \otimes p_V$  and hence it provides a simple and clean one-shot achievability bound. We note that such a scenario was considered by Anshu *et al.* [128]. Hence, Theorem 7 improves on the achievability in Ref. [128, Theorem 13].

Generally, Alice can adopt a joint precoding  $p_{UV}$ , which is called *Marton's inner bound* in the classical setting [142, 143] (see also the studies in the quantum setting [43, 117, 136–139]); however, it would require additional covering techniques. We leave this for future work [144].

Next, we present entanglement-assisted classical communication over quantum broadcast channels.

*Definition 7 (Entanglement-assisted classical communication over quantum broadcast channels).*—Let  $\mathcal{N}_{A \rightarrow BC}$  be a quantum broadcast channel.

- (1) Alice holds classical registers  $M_B$  and  $M_C$ , and quantum registers  $A'_B$  and  $A'_C$ . Bob holds quantum registers  $B$  and  $R'_B$ . Charlie holds quantum registers  $C$  and  $R'_C$ .
- (2) Bob and Alice share an arbitrary state  $\theta_{R'_B A'_B}$ . Charlie and Alice share an arbitrary state  $\theta_{R'_C A'_C}$ .
- (3) Alice performs an encoding  $\mathcal{E}_{A'_B A'_C \rightarrow A}^{m_B, m_C}$  on  $\theta_{R'_B A'_B} \otimes \theta_{R'_C A'_C}$  for any equiprobable message  $(m_B, m_C) \in M_B \times M_C$  she wanted to send to Bob and Charlie, respectively.
- (4) The channel  $\mathcal{N}_{A \rightarrow BC}$  is applied on Alice's register  $A$ , and outputs a marginal state on  $B$  at Bob and a marginal state on  $C$  at Charlie.
- (5) Bob performs a decoding measurement  $\{\Pi_{R'_B}^{m_B}\}_{m_B \in M_B}$  on register  $B$  to extract the sent message  $m_B$ ,

and Charlie performs a decoding measurement  $\{\Pi_{R'_C C}^{m_C}\}_{m_C \in M_C}$  on register  $C$  to extract the sent message  $m_C$ .

An  $(M_B, M_C, \varepsilon_B, \varepsilon_C)$ -EA code for  $\mathcal{N}_{A \rightarrow BC}$  is a protocol such that  $|M_B| = M_B$ ,  $|M_C| = M_C$  and the average error probabilities satisfy

$$\begin{aligned} & \frac{1}{M_B M_C} \sum_{(m_B, m_C) \in M_B \times M_C} \text{Tr} \left[ (1 - \Pi_{R'_B B}^{m_B}) \mathcal{N}_{A \rightarrow BC} \right. \\ & \quad \left. \circ \mathcal{E}_{A'_B A'_C \rightarrow A}^{m_B, m_C} (\theta_{R'_B A'_B} \otimes \theta_{R'_C A'_C}) \right] \leq \varepsilon_B, \\ & \frac{1}{M_B M_C} \sum_{(m_B, m_C) \in M_B \times M_C} \text{Tr} \left[ (1 - \Pi_{R'_C C}^{m_C}) \mathcal{N}_{A \rightarrow BC} \right. \\ & \quad \left. \circ \mathcal{E}_{A'_B A'_C \rightarrow A}^{m_B, m_C} (\theta_{R'_B A'_B} \otimes \theta_{R'_C A'_C}) \right] \leq \varepsilon_C. \end{aligned}$$

We follow the similar analysis as above by considering communication from Alice to Bob and from Alice to Charlie, separately. Again, we also employ the encoder of the position-based coding as in Refs. [43, Theorem 6] and [128, Theorem 6], and apply the pretty-good measurement for decoding.

- (a) **Preparation.** Consider an arbitrary state  $\theta_{R_B R_C A}$  satisfying  $\theta_{R_B R_C} = \theta_{R_B} \otimes \theta_{R_C}$ . Let  $\theta_{R_B A_B}$  be a purified state of  $\theta_{R_B}$ , and let  $\theta_{R_C A_C}$  be a purified state of  $\theta_{R_C}$ . Then, Alice and Bob share the  $M_B$ -fold product states  $\theta_{R'_B A'_B} := \theta_{R_B A_B}^{\otimes M_B}$ , and Alice and Charlie share the  $M_C$ -fold product states  $\theta_{R'_C A'_C} := \theta_{R_C A_C}^{\otimes M_C}$ .
- (b) **Encoding.** For each  $(m_B, m_C) \in M_B \times M_C$ , Alice sends the  $(m_B, m_C)$ th registers  $A_B$  and  $A_C$  and then performs an isometry transformation  $\mathcal{V}_{A_B A_C \rightarrow EA}$  such that  $\mathcal{V}_{A_B A_C \rightarrow EA} (\theta_{R_B A_B} \otimes \theta_{R_C A_C})$  equals a purified state  $\theta_{E R_B R_C A}$  of  $\theta_{R_B R_C A}$  with an additional purifying register  $E$ . The overall encoding map is then  $\mathcal{E}_{A'_B A'_C \rightarrow A}^{m_B, m_C} = \text{Tr}_E \circ \mathcal{V}_{A_B A_C \rightarrow EA} \circ \text{Tr}_{A_B \setminus \{m_B\} A_C \setminus \{m_C\}}$ .
- (c) **Decoding.** Denoting (with a slight abuse of notation) the marginal channel output states at Bob for sending  $m_B \in M_B$  by

$$\begin{aligned} \rho_{R'_B B}^{m_B} & := \theta_{R_B}^{\otimes (m_B - 1)} \otimes \text{Tr}_{C E R_C} \circ \mathcal{N}_{A \rightarrow BC} \\ & \quad \circ \mathcal{V}_{A_B A_C \rightarrow EA} (\theta_{R_B A_B} \otimes \theta_{R_C A_C}) \otimes \theta_{R_B}^{\otimes (M_B - m_B)}, \end{aligned}$$

Bob performs the corresponding pretty-good measurement:

$$\Pi_{R'_B B}^{m_B} := \frac{\rho_{R'_B B}^{m_B}}{\sum_{\bar{m}_B \in M_B} \rho_{R'_B B}^{\bar{m}_B}}, \quad m_B \in M_B.$$

Similarly, denoting (with a slight abuse of notation again) the marginal channel output states at Charlie for sending  $m_C \in M_C$  by

$$\begin{aligned} \rho_{R'_C C}^{m_C} & := \theta_{R_C}^{\otimes (m_C - 1)} \otimes \text{Tr}_{B E R_B} \circ \mathcal{N}_{A \rightarrow BC} \circ \mathcal{V}_{A_B A_C \rightarrow EA} \\ & \quad (\theta_{R_B A_B} \otimes \theta_{R_C A_C}) \otimes \theta_{R_C}^{\otimes (M_C - m_C)}, \end{aligned}$$

Charlie performs the corresponding pretty-good measurement:

$$\Pi_{R'_C C}^{m_C} := \frac{\rho_{R'_C C}^{m_C}}{\sum_{\bar{m}_C \in M_C} \rho_{R'_C C}^{\bar{m}_C}}, \quad m_C \in M_C.$$

Then, we obtain the following result (without duplicating the proof).

*Theorem 8 (A one-shot achievability bound for EA classical communication over quantum broadcast channels).—*Consider an arbitrary quantum broadcast channel  $\mathcal{N}_{A \rightarrow BC}$ . Then, there exists an  $(M_B, M_C, \varepsilon_B, \varepsilon_C)$ -EA code for  $\mathcal{N}_{A \rightarrow BC}$  such that, for any  $\theta_{R_B R_C A}$  satisfying  $\theta_{R_B R_C} = \theta_{R_B} \otimes \theta_{R_C}$ ,

$$\begin{aligned} \varepsilon_B & \leq \text{Tr} [\rho_{R_B B} \wedge (M_B - 1) \rho_{R_B} \otimes \rho_B], \\ \varepsilon_C & \leq \text{Tr} [\rho_{R_C C} \wedge (M_C - 1) \rho_{R_C} \otimes \rho_C], \end{aligned}$$

where  $\rho_{R_B R_C B C} := \mathcal{N}_{A \rightarrow BC} (\theta_{R_B R_C A})$ .

## F. Communication with casual state information at the encoder

In this section, we consider entanglement-assisted and unassisted classical communication over quantum channels with causal channel state information available at the encoder [43, 117, 128], which is the quantum generalization of the classical Gel'fand-Pinsker channel [145]; see also Ref. [102, Sec. 7].

*Definition 8 (Classical-quantum channel coding with causal state information).—*Let  $\mathcal{N}_{X S \rightarrow B} : (x, s) \mapsto \rho_B^{x, s}$  be a classical-quantum channel parameterized by  $s \in \mathbf{S}$  and assume that a channel state  $\rho_S$  is available at the encoder.

- (1) The channel holds a classical register  $\mathbf{S}$ . Alice holds classical registers  $\mathbf{M}$ ,  $\mathbf{X}$ , and  $\mathbf{S}'$  (an identical copy of  $\mathbf{S}$ ). Bob holds a quantum register  $\mathbf{B}$ .
- (2) Given a realization of the channel state  $s \in \mathbf{S}'$ , an encoding  $(m, s) \mapsto x(m, s)$  maps an equiprobable message  $m \in \mathbf{M}$  to a codeword in  $\mathbf{X}$ .
- (3) The classical-quantum channel  $\mathcal{N}_{X S \rightarrow B}$  is applied on Alice's register  $\mathbf{X}$  given the realization of the channel state  $s \in \mathbf{S}$ , and outputs a state on  $\mathbf{B}$  at Bob. (The realizations  $s \in \mathbf{S}'$  at Alice and  $s \in \mathbf{S}$  at the channel are identical.)

- (4) Bob performs a decoding measurement described by a measurement  $\{\Pi_{\mathbf{B}}^m\}_{m \in \mathbf{M}}$  on  $\mathbf{B}$  to extract the sent message  $m \in \mathbf{M}$ . (Note that Bob is aware of the mathematical description of the probability distribution  $p_{\mathbf{S}}$  but not of the realization of a specific  $s \in \mathbf{S}$ .)

An  $(M, \varepsilon)$  code for  $\mathcal{N}_{\mathbf{X}\mathbf{S} \rightarrow \mathbf{B}}$  with state information  $p_{\mathbf{S}}$  is a protocol such that  $|\mathbf{M}| = M$  and the average error probability satisfies

$$\frac{1}{M} \sum_{(m,s) \in \mathbf{M} \times \mathbf{S}} p_{\mathbf{S}}(s) \text{Tr}[\rho_{\mathbf{B}}^{x(m,s),s} (\mathbb{1}_{\mathbf{B}} - \Pi_{\mathbf{B}}^m)] \leq \varepsilon.$$

We adopt the standard random coding strategy as follows.

- (a) **Encoding.** We introduce an auxiliary classical register  $\mathbf{U}$  for precoding. The message  $m \in \mathbf{M}$  is mapped to a precodeword  $u \in \mathbf{U}$  pairwise independently according to  $p_{\mathbf{U}}$ . With the realization of the channel state  $s \in \mathbf{S}$ , Alice picks a (deterministic) encoding  $(u(m), s) \mapsto x(u(m), s) \in \mathbf{X}$ .
- (b) **Decoding.** At the receiver, denoting (with a slight abuse of notation) the channel output state by  $\rho_{\mathbf{B}}^m := \sum_{s \in \mathbf{S}} p_{\mathbf{S}}(s) \rho_{\mathbf{B}}^{x(u(m),s),s}$  for each  $m \in \mathbf{M}$ , Bob performs the corresponding pretty-good measurement:

$$\Pi_{\mathbf{B}}^m := \frac{\rho_{\mathbf{B}}^m}{\sum_{\bar{m} \in \mathbf{M}} \rho_{\mathbf{B}}^{\bar{m}}} \quad \text{for all } m \in \mathbf{M}.$$

Then, following the analysis in Sec. III, we obtain a one-shot achievability bound (without duplicating the proof).

*Theorem 9 (A one-shot achievability bound for classical-quantum channel coding with casual state information).*—Consider an arbitrary classical-quantum channel  $\mathcal{N}_{\mathbf{X}\mathbf{S} \rightarrow \mathbf{B}} : (x, s) \mapsto \rho_{\mathbf{B}}^{x,s}$  with state information  $p_{\mathbf{S}}$ . Then, there exists an  $(M, \varepsilon)$  code for  $\mathcal{N}_{\mathbf{X}\mathbf{S} \rightarrow \mathbf{B}}$  with state information  $p_{\mathbf{S}}$  such that, for any probability distribution  $p_{\mathbf{U}}$  and (deterministic) map  $(u, s) \mapsto x(u, s)$ ,

$$\varepsilon \leq \text{Tr}[\rho_{\mathbf{U}\mathbf{B}} \wedge (M - 1)\rho_{\mathbf{U}} \otimes \rho_{\mathbf{B}}].$$

Here,  $\rho_{\mathbf{U}\mathbf{B}} := \sum_{(u,s) \in \mathbf{U} \times \mathbf{S}} p_{\mathbf{U}}(u) |u\rangle \langle u| \otimes p_{\mathbf{S}}(s) \rho_{\mathbf{B}}^{x(u,s),s}$ .

The result extends to classical communication over quantum channels  $\mathcal{N}_{\mathbf{A}\mathbf{S} \rightarrow \mathbf{B}}$  with quantum state information  $\vartheta_{\mathbf{S}}$  (see also the definition below in the entanglement-assisted setting). We refer the reader to Table I for the corresponding results.

*Remark 10.*—In the precoding phase of Theorem 9, the chosen probability distribution  $p_{\mathbf{U}}$  is independent of the channel state  $p_{\mathbf{S}}$ . This coding strategy is called *casual state information* at the encoder [102, Sec. 7.5] and it was also studied in the quantum setting [128, Theorem 12]. (Hence, Theorem 9 improves on Ref. [128, Theorem 12].)

For the scenario of *noncasual state information* at the encoder, the precoding probability distribution on  $\mathbf{U}$  may be correlated with the channel state  $p_{\mathbf{S}}$  [117, Sec. 4], [43, Sec. V]. This would require additional techniques. We leave this for future work [144].

Next, we move on to the entanglement-assisted setting.

*Definition 9 (Entanglement-assisted classical communication over quantum channels with causal state information).*—Let  $\mathcal{N}_{\mathbf{A}\mathbf{S} \rightarrow \mathbf{B}}$  be a quantum channel with a channel state  $\vartheta_{\mathbf{S}}$ .

- (1) The channel holds a quantum register  $\mathbf{S}$ . Alice holds a classical register  $\mathbf{M}$  and quantum registers  $\mathbf{A}'$  and  $\mathbf{S}'$ . Bob holds quantum registers  $\mathbf{R}'$  and  $\mathbf{B}$ .
- (2) A resource of arbitrary state  $\theta_{\mathbf{R}'\mathbf{A}'}$  is shared between Bob and Alice. Moreover, let  $\vartheta_{\mathbf{S}'\mathbf{S}}$  be a purified state of  $\vartheta_{\mathbf{S}}$  shared between Alice and the channel.
- (3) Alice performs an encoding  $\mathcal{E}_{\mathbf{A}'\mathbf{S}' \rightarrow \mathbf{A}}^m$  on registers  $\mathbf{A}'$  and  $\mathbf{S}'$  of state  $\theta_{\mathbf{R}'\mathbf{A}'} \otimes \vartheta_{\mathbf{S}'\mathbf{S}}$  for any equiprobable message  $m \in \mathbf{M}$ .
- (4) The quantum channel  $\mathcal{N}_{\mathbf{A}\mathbf{S} \rightarrow \mathbf{B}}$  is applied on Alice's register  $\mathbf{A}$  and the register  $\mathbf{S}$  of the channel state, and outputs a state on  $\mathbf{B}$  at Bob.
- (5) Bob performs a decoding measurement  $\{\Pi_{\mathbf{R}'\mathbf{B}}^m\}_{m \in \mathbf{M}}$  on  $\mathbf{R}'\mathbf{B}$  to extract the sent message  $m$ . (Note that Bob is aware of the mathematical description of the channel state  $\vartheta_{\mathbf{S}}$ , but Bob cannot access the channel's register  $\mathbf{S}$  nor operate on such a channel state.)

An  $(M, \varepsilon)$ -EA code for  $\mathcal{N}_{\mathbf{A}\mathbf{S} \rightarrow \mathbf{B}}$  with state information  $\vartheta_{\mathbf{S}}$  is a protocol such that  $|\mathbf{M}| = M$  and the average error probability satisfies

$$\frac{1}{M} \sum_{m \in \mathbf{M}} \text{Tr}[(\mathbb{1} - \Pi_{\mathbf{R}'\mathbf{B}}^m) \mathcal{N}_{\mathbf{A}\mathbf{S} \rightarrow \mathbf{B}} \circ \mathcal{E}_{\mathbf{A}'\mathbf{S}' \rightarrow \mathbf{A}}^m(\theta_{\mathbf{R}'\mathbf{A}'} \otimes \vartheta_{\mathbf{S}'\mathbf{S}})] \leq \varepsilon.$$

We also use the encoder of the position-based coding as in Refs. [43, Theorem 5], [128, Theorem 4], and the pretty-good measurement for decoding.

- (a) **Preparation.** Consider an arbitrary state  $\theta_{\mathbf{R}\mathbf{A}\mathbf{S}}$  satisfying  $\theta_{\mathbf{R}\mathbf{S}} = \theta_{\mathbf{R}} \otimes \vartheta_{\mathbf{S}}$ . Let  $\theta_{\mathbf{R}\mathbf{U}}$  be a purified state of  $\theta_{\mathbf{R}}$  with an additional quantum register  $\mathbf{U}$  at Alice. Then, Alice and Bob share the  $M$ -fold product states  $\theta_{\mathbf{R}'\mathbf{A}'} := \theta_{\mathbf{R}\mathbf{U}}^{\otimes M}$ .
- (b) **Encoding.** For each  $m \in \mathbf{M}$ , Alice sends the  $m$ th register of  $\mathbf{U}$  and applies an isometry transformation  $\mathcal{V}_{\mathbf{S}'\mathbf{U} \rightarrow \mathbf{E}\mathbf{A}}$  such that  $\mathcal{V}_{\mathbf{S}'\mathbf{U} \rightarrow \mathbf{E}\mathbf{A}}(\theta_{\mathbf{R}\mathbf{U}} \otimes \vartheta_{\mathbf{S}'\mathbf{S}})$  equals a purified state  $\theta_{\mathbf{E}\mathbf{R}\mathbf{A}\mathbf{S}}$  of  $\theta_{\mathbf{R}\mathbf{A}\mathbf{S}}$  with an additional purifying register  $\mathbf{E}$ . Then, the overall encoding map is  $\mathcal{E}_{\mathbf{U}^M \rightarrow \mathbf{A}}^m = \text{Tr}_{\mathbf{E}} \circ \mathcal{V}_{\mathbf{S}'\mathbf{U} \rightarrow \mathbf{E}\mathbf{A}} \circ \text{Tr}_{\mathbf{U}^{M \setminus \{m\}}}$ .
- (c) **Decoding.** Denoting (with a slight abuse of notation again) the marginal channel output states at Bob for

sending  $m \in \mathbf{M}$  by

$$\rho_{\mathbf{R}^m \mathbf{B}}^m := \theta_{\mathbf{R}}^{\otimes(m-1)} \otimes \mathcal{N}_{\mathbf{A} \rightarrow \mathbf{B}} \circ \text{Tr}_{\mathbf{E}} \\ \circ \mathcal{V}_{\mathbf{S}' \mathbf{U} \rightarrow \mathbf{E} \mathbf{A}}(\theta_{\mathbf{R} \mathbf{U}} \otimes \vartheta_{\mathbf{S}' \mathbf{S}}) \otimes \theta_{\mathbf{R}}^{\otimes(M-m)},$$

Bob performs the corresponding pretty-good measurement:

$$\Pi_{\mathbf{R}^m \mathbf{B}}^m := \frac{\rho_{\mathbf{R}^m \mathbf{B}}^m}{\sum_{\bar{m} \in \mathbf{M}} \rho_{\mathbf{R}^{\bar{m}} \mathbf{B}}^{\bar{m}}}, \quad m \in \mathbf{M}.$$

Then, applying the analysis in Sec. IV B, we obtain the following result (without duplicating the proof).

*Theorem 10 (A one-shot achievability bound for EA classical communication over quantum channels with casual state information).*—Consider an arbitrary quantum channel  $\mathcal{N}_{\mathbf{A} \rightarrow \mathbf{B}}$  with state information  $\vartheta_{\mathbf{S}}$ . Then, there exists an  $(M, \varepsilon)$ -EA code for  $\mathcal{N}_{\mathbf{A} \rightarrow \mathbf{B}}$  with state information  $\vartheta_{\mathbf{S}}$  such that, for any  $\theta_{\mathbf{R} \mathbf{A} \mathbf{S}}$  satisfying  $\theta_{\mathbf{R} \mathbf{S}} = \theta_{\mathbf{R}} \otimes \vartheta_{\mathbf{S}}$ ,

$$\varepsilon \leq \text{Tr}[\rho_{\mathbf{R} \mathbf{B}} \wedge (M - 1)\rho_{\mathbf{R}} \otimes \rho_{\mathbf{B}}].$$

Here,  $\rho_{\mathbf{R} \mathbf{B}} := \mathcal{N}_{\mathbf{A} \rightarrow \mathbf{B}}(\theta_{\mathbf{R} \mathbf{A} \mathbf{S}})$ .

## V. CONCLUSIONS

We propose a conceptually simple analysis of one-shot achievability for processing classical information in quantum systems. The key point of this work is to demonstrate that the pretty-good measurement directly translates the conditional error probability of a multiple-state discrimination to the error of discriminating a state against the rest. This can be viewed as the *one-versus-rest* strategy, and, hence, the pretty-good measurement effectively resembles the quantum union bound in quantum coding design and analysis. We obtain an elegant closed-form expression of the average error probability for classical communication over quantum channels with standard random coding and basic properties of the noncommutative minimal. The proposed method is tight in the sense that it gives tighter one-shot achievability bounds for channels without further constraints such as symmetry (see Sec. III A), and it unifies the asymptotic derivations in the large, small, and moderate deviation regimes (Fig. 2). Moreover, the analysis naturally applies to various quantum information-theoretic tasks (see Sec. IV and Table I). This manifests that the proposed method may be considered a fundamental and unified approach to deriving achievable error bounds in quantum information theory. In this regard, we may term it as a *one-shot quantum packing lemma* (Theorem 3). Essentially, the proposed analysis can be applied to and can sharpen almost all existing results that rely on the Hayashi-Nagaoka operator inequality [30, Lemma 2]; see, e.g., Refs. [30,43,46,59,60,80,128,132,146]. The improvement is crucial because every bit counts in a one-shot bound

because weaker one-shot bounds could be trivial in certain practical scenarios. Hence, we expect more applications of the proposed analysis to emerge. As for the computational aspect, we point out a recently developed quantum algorithm for implementing the pretty-good measurement [54]. Last but not least, the proposed achievability analysis also applies to the converse analysis for the covering-type problems [73,144,147,148].

We list some open problems along these research directions.

- (a) Standard second-order analysis of the achievable coding rate consists of two steps: (i) reducing the underlying task to binary quantum hypothesis testing and (ii) an asymptotic expansion of the quantum hypothesis-testing divergence [29,38,39,82,83]. The proposed approach simplifies step (i), and, hence, now the bottleneck lies in step (ii). Specifically, we conjecture a third-order achievable expansion of the quantum hypothesis-testing divergence in Eq. (18). If Eq. (18) holds then Proposition 2 will lead to the best possible third-order coding rate for general classical-quantum channels without further assumptions, i.e.,
 
$$\log M \geq nI(\mathbf{X} : \mathbf{B})_{\rho} + \sqrt{nV(\mathbf{X} : \mathbf{B})_{\rho}} \Phi^{-1}(\varepsilon) - O(1).$$
- (b) To the best of our knowledge, conjectures made by Mosonyi and Audenaert [69, Conjecture 4.2], and Qi *et al.* [60, Conjecture 18] are still open. If they were true then the established one-shot achievability bound for classical-quantum multiple-access channels (Theorem 5) will directly imply an upper bound on the error probability by a sum of exponential decays [149].
- (c) Can the established strengthened one-shot bound in Eq. (11) provide a simple proof for the upper bound on the strong converse exponent of classical-quantum channel coding (see Refs. [78, Sec. 5.4], [79, Proposition IV.5], and [80, Proposition VI.2])?
- (d) In the classical setting (where all the channel output states  $\{\rho_{\mathbf{B}}^x\}_{x \in \mathbf{X}}$  mutually commute), the derived bound in Theorem 1 is still weaker than the *random-coding union bound* proved by Polyanskiy *et al.* [22, Theorem 16], i.e., the latter implies Theorem 1 in the commuting case. Nevertheless, we remark that Eq. (8) can already yield Gallager’s random-coding bound in the commuting case. Hence, there are technical noncommutativity difficulties that remain to be solved.
- (e) It is not clear whether the derived bound in Theorem 1 can lead to Gallager’s random-coding exponent [14,15] for general classical-quantum channels. It is interesting to see if the recent techniques proposed by Dupuis [150] (see also Ref. [147]) and Renes [97] can be combined with the proposed analysis.

## ACKNOWLEDGMENTS

H.-C. C. is supported by the Young Scholar Fellowship (Einstein Program) of the Ministry of Science and Technology, Taiwan (R.O.C.) under Grants No. NSTC 111-2636-E-002-026, No. NSTC 112-2636-E-002-009, No. NSTC 112-2119-M-007-006, No. NSTC 112-2119-M-001-006, No. NSTC 112-2124-M-002-003, by the Yushan Young Scholar Program of the Ministry of Education, Taiwan (R.O.C.) under Grants No. NTU-111V1904-3, No. NTU-112V1904-4, and by the research project ‘‘Pioneering Research in Forefront Quantum Computing, Learning and Engineering’’ of National Taiwan University under Grant No. NTC-CC-112L893405. H.-C. C. acknowledges support from the ‘‘Center for Advanced Computing and Imaging in Biomedicine (NTU-112L900702)’’ through the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

## APPENDIX: A TRACE INEQUALITY

This section is devoted to proving the trace inequality in Fact 1 (vii):

$$\mathrm{Tr}[A \wedge B] \geq \mathrm{Tr}\left[A \frac{B}{A+B}\right]. \quad (\text{A1})$$

Note that a special case of Eq. (A1) for  $\mathrm{Tr}[A] = \mathrm{Tr}[B] = 1$  is a consequence of Barnum and Knill’s theorem [72] (see also Ref. [64, Theorem 3.10]). The result has been extended to the case of general positive semidefinite  $A$  and  $B$  in the author’s previous work [73, Lemma 3]. For completeness, we provide a strengthened proof in the following lemma that implies the desired Eq. (A1) by extending the ideas of Sason and Verdú [151] and Renes [152].

*Lemma 1 (A trace inequality for a noncommutative parallel sum).*—Let  $A$  and  $B$  be arbitrary positive semidefinite operators satisfying  $\mathrm{Tr}[A+B] > 0$ . Then, it holds that

$$\mathrm{Tr}\left[A \frac{B}{A+B}\right] \leq \frac{\mathrm{Tr}[A \vee B] \cdot \mathrm{Tr}[A \wedge B]}{\mathrm{Tr}[A+B]} \leq \mathrm{Tr}[A \wedge B]. \quad (\text{A2})$$

$$\begin{aligned} \mathrm{Tr}\left[A \frac{A}{A+B}\right] + \mathrm{Tr}\left[B \frac{B}{A+B}\right] &= e^{D_2^*(A \oplus B \| (A+B)^{\oplus 2})} \geq e^{D_2^*(\Lambda(A \oplus B) \| \Lambda((A+B)^{\oplus 2}))} \\ &= \exp\left\{D_2^*\left(\begin{bmatrix} \mathrm{Tr}[A \vee B] & 0 \\ 0 & \mathrm{Tr}[A \wedge B] \end{bmatrix} \parallel \begin{bmatrix} \mathrm{Tr}[A+B] & 0 \\ 0 & \mathrm{Tr}[A+B] \end{bmatrix}\right)\right\} = \frac{(\mathrm{Tr}[A \vee B])^2}{\mathrm{Tr}[A+B]} + \frac{(\mathrm{Tr}[A \wedge B])^2}{\mathrm{Tr}[A+B]}. \end{aligned} \quad (\text{A3})$$

Noting that the left-hand side of Eq. (A3) can be written as

$$\mathrm{Tr}\left[A \frac{A}{A+B}\right] + \mathrm{Tr}\left[B \frac{B}{A+B}\right] = \mathrm{Tr}[A+B] - 2 \mathrm{Tr}\left[A \frac{B}{A+B}\right],$$

Here,  $A \vee B := (A+B+|A-B|)/2$  and  $A \wedge B := (A+B-|A-B|)/2$ .

*Remark 11.*—In the scalar case of positive  $a$  and  $b$ , the inequality  $ab/(a+b) \leq a \wedge b$  is obvious, and the term  $ab/(a+b) = (a^{-1}+b^{-1})^{-1}$  is called the *parallel sum* of  $a, b$ . Hence, Lemma 1 may be viewed as a noncommutative generalization of it. We note that an operator parallel sum  $(A^{-1}+B^{-1})^{-1}$  has been studied before (e.g., Refs. [153, Sec. 3], [154, Sec. 5]), and it is related to the Kubo-Ando operator (harmonic) mean [155, 156]. Furthermore, it can be shown that  $\mathrm{Tr}[(A^{-1}+B^{-1})^{-1}] \leq \mathrm{Tr}[A \frac{B}{A+B}]$ , and hence Lemma 1 also provides an upper bound (in trace) to the operator parallel sum.

*Proof of Lemma 1.*—We define the *collision divergence* [76] for  $A, B \geq 0$  as

$$D_2^*(A \| B) := \log \mathrm{Tr}[(B^{-1/4}AB^{-1/4})^2].$$

Let  $\{\Pi_A, \Pi_B\}$  be the optimal measurement for distinguishing positive semidefinite operators  $A$  and  $B$ , i.e., by recalling the Holevo-Helstrom theorem [50–52],

$$\begin{aligned} \sup_{0 \leq T \leq \mathbb{1}} \mathrm{Tr}[AT] + \mathrm{Tr}[B(\mathbb{1}-T)] &= \mathrm{Tr}[A\Pi_A] + \mathrm{Tr}[B\Pi_B] \\ &= \mathrm{Tr}[A \vee B]. \end{aligned}$$

Denote by ‘‘ $\oplus$ ’’ the direct sum operation. We introduce a measure-and-prepare operation  $\Lambda$ , which is a completely positive and trace-preserving (CPTP) map:

$$\Lambda : (\cdot) \mapsto \mathrm{Tr}[(\cdot)\Pi_A \oplus \Pi_B] \oplus \mathrm{Tr}[(\cdot)(\mathbb{1} - \Pi_A \oplus \Pi_B)].$$

We calculate

$$\begin{aligned} \Lambda(A \oplus B) &= \mathrm{Tr}[A \vee B] \oplus \mathrm{Tr}[A \wedge B], \\ \Lambda((A+B)^{\oplus 2}) &= \mathrm{Tr}[A+B] \oplus \mathrm{Tr}[A+B]. \end{aligned}$$

Since the map  $e^{D_2^*(\cdot \| \cdot)}$  is monotonically decreasing under CPTP maps [157–159], we obtain

then the above inequality translates to

$$2 \operatorname{Tr} \left[ A \frac{B}{A+B} \right] \leq \operatorname{Tr}[A+B] - \frac{(\operatorname{Tr}[A \vee B])^2}{\operatorname{Tr}[A+B]} - \frac{(\operatorname{Tr}[A \wedge B])^2}{\operatorname{Tr}[A+B]} = \frac{(\operatorname{Tr}[A+B])^2 - (\operatorname{Tr}[A \vee B])^2 - (\operatorname{Tr}[A \wedge B])^2}{\operatorname{Tr}[A+B]}$$

$$\stackrel{(a)}{=} \frac{2 \operatorname{Tr}[A \vee B] \cdot \operatorname{Tr}[A \wedge B]}{\operatorname{Tr}[A+B]} \stackrel{(b)}{\leq} 2 \operatorname{Tr}[A \wedge B].$$

Here, (a) follows from the identity  $A+B = A \vee B + A \wedge B$ ; and the inequality  $\operatorname{Tr}[A \vee B] \leq \operatorname{Tr}[A+B]$  used in (b) is because  $A+B$  is a feasible solution to the infimum representation of the noncommutative maximum [51,52,62]:

$$A \vee B = \frac{A+B+|A-B|}{2}$$

$$= \arg \min_{M=M^\dagger} \{\operatorname{Tr}[M] : M \geq A, M \geq B\}.$$

This completes the proof.  $\blacksquare$

- [1] R. Jozsa and B. Schumacher, A new proof of the quantum noiseless coding theorem, *J. Mod. Opt.* **41**, 2343 (1994).
- [2] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters, Classical information capacity of a quantum channel, *Phys. Rev. A* **54**, 1869 (1996).
- [3] B. Schumacher and M. D. Westmoreland, Sending classical information via noisy quantum channels, *Phys. Rev. A* **56**, 131 (1997).
- [4] A. Holevo, The capacity of the quantum channel with general signal states, *IEEE Trans. Inf. Theory* **44**, 269 (1998).
- [5] A. Winter, Coding theorem and strong converse for quantum channels, *IEEE Trans. Inf. Theory* **45**, 2481 (1999).
- [6] A. Winter, Ph.D. thesis, Universität Bielefeld, [ArXiv: quant-ph/9907077](https://arxiv.org/abs/quant-ph/9907077) (1999).
- [7] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, Cambridge, England, 2016).
- [8] C. E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.* **27**, 379 (1948).
- [9] To the best of our knowledge, the first achievability analysis for one-shot coding in quantum information theory was made by Burnashev and Holevo [89] in studying pure-state classical-quantum ( $c$ - $q$ ) channels (based on a Gram matrix technique), and a one-shot strong converse for  $c$ - $q$  channel coding was proved by Ogawa and Nagaoka [160] by extending the approach of Arimoto [161] to the noncommutative scenario. Technically speaking, the quantum dense coding [162] and teleportation [163] are both one-shot protocols—however, the original references only concerned noiseless channels. Furthermore, the early work of quantum state discrimination [50,103] also considered the one-shot setting. This paper focuses on *packing-type problems* in one-shot quantum information theory with general (noisy) quantum channels without specific constraints. We skip the *covering-type problems*

such as quantum covering [43,58,147,148,164–167], privacy amplification [73,76,148,150,165,168–171], decoupling [117,172,173], and simulation [174–177] since their achievability techniques are different from that of packing problems.

- [10] A. Feinstein, Error bounds in noisy channels without memory, *IEEE Trans. Inf. Theory* **1**, 13 (1955).
- [11] R. M. Fano, *Transmission of Information, A Statistical Theory of Communications* (The MIT Press, Cambridge, Massachusetts, United States, 1961).
- [12] V. Strassen, in *Transactions of the Third Prague Conference on Information Theory*, p. 689 (Academic Press Inc., Prague, Czech Republic, 1962).
- [13] R. G. Gallager, *Low-Density Parity-Check Codes* (The MIT Press, Cambridge, Massachusetts, United States, 1963).
- [14] R. Gallager, A simple derivation of the coding theorem and some applications, *IEEE Trans. Inf. Theory* **11**, 3 (1965).
- [15] R. Gallager, *Information Theory and Reliable Communication* (Wiley, Hoboken, New Jersey, United States, 1968).
- [16] S. Verdú and T. S. Han, A general formula for channel capacity, *IEEE Trans. Inf. Theory* **40**, 1147 (1994).
- [17] T. S. Han, *Information-Spectrum Methods in Information Theory* (Springer Berlin Heidelberg, Midtown Manhattan, New York City, United States, 2003).
- [18] I. Sason and S. Shamai, *Performance Analysis of Linear Codes under Maximum-Likelihood Decoding: A Tutorial* (Now Foundations and Trends, Norwell, Massachusetts, United States, 2006), Vol. 3, p. 1.
- [19] R. G. Gallager, *Principles of Digital Communication* (Cambridge University Press, Cambridge, England, 2008).
- [20] T. Richardson and R. Urbanke, *Modern Coding Theory* (Cambridge University Press, Cambridge, England, 2008).
- [21] M. Hayashi, Information spectrum approach to second-order coding rate in channel coding, *IEEE Trans. Inf. Theory* **55**, 4947 (2009).
- [22] Y. Polyanskiy, H. V. Poor, and S. Verdú, Channel coding rate in the finite blocklength regime, *IEEE Trans. Inf. Theory* **56**, 2307 (2010).
- [23] A. Lapidoth, *A Foundation in Digital Communication* (Cambridge University Press, Cambridge, England, 2017), 2nd ed.
- [24] S. Aaronson, in *Twenty-First Annual IEEE Conference on Computational Complexity* (IEEE, Prague, Czech Republic, 2006), p. 261.

- [25] V. Giovannetti, S. Lloyd, and L. Maccone, Achieving the Holevo bound via sequential measurements, *Phys. Rev. A* **85**, 012302 (2012).
- [26] P. Sen, in *2012 IEEE International Symposium on Information Theory Proceedings* (IEEE, Cambridge, MA, United States, 2012), p. 736.
- [27] M. M. Wilde, Sequential decoding of a general classical-quantum channel, *Proc. R. Soc. A: Math. Phys. Eng. Sci.* **469**, 20130259 (2013).
- [28] J. Gao, Quantum union bounds for sequential projective measurements, *Phys. Rev. A* **92**, 052331 (2015).
- [29] S. Khabbazi Oskouei, S. Mancini, and M. M. Wilde, Union bound for quantum information processing, *Proc. R. Soc. A: Math. Phys. Eng. Sci.* **475**, 20180612 (2019).
- [30] M. Hayashi and H. Nagaoka, General formulas for capacity of classical-quantum channels, *IEEE Trans. Inf. Theory* **49**, 1753 (2003).
- [31] A. Feinstein, A new basic theorem of information theory, *Trans. IRE Prof. Group Inf. Theory* **4**, 2 (1954).
- [32] We refer the reader to Ref. [29, Theorem 2.1] for a precise statement of the quantum union bound. An application of it with the sequential decoding strategy [27] leads to the same achievability bound as in Refs. [30,46] (for infinite-dimensional Hilbert space as well), [29, Theorem 5.1].
- [33] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete, Discriminating states: The quantum Chernoff bound, *Phys. Rev. Lett.* **98**, 160501 (2007).
- [34] K. M. R. Audenaert, M. Nussbaum, A. Szkoła, and F. Verstraete, Asymptotic error rates in quantum hypothesis testing, *Commun. Math. Phys.* **279**, 251 (2008).
- [35] M. Hayashi, Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding, *Phys. Rev. A* **76**, 062301 (2007).
- [36] V. Jakšić, Y. Ogata, C.-A. Pillet, and R. Seiringer, Quantum hypothesis testing and non-equilibrium statistical mechanics, *Rev. Math. Phys.* **24**, 1230002 (2012).
- [37] V. Y. F. Tan, Asymptotic estimates in information theory with non-vanishing error probabilities, *Found. Trends Commun. Inf. Theory* **10**, 1 (2014).
- [38] M. Tomamichel and M. Hayashi, A hierarchy of information quantities for finite block length analysis of quantum tasks, *IEEE Trans. Inf. Theory* **59**, 7693 (2013).
- [39] K. Li, Second-order asymptotics for quantum hypothesis testing, *Ann. Stat.* **42**, 171 (2014).
- [40] M. Tomamichel and V. Y. F. Tan, Second-order asymptotics for the classical capacity of image-additive quantum channels, *Commun. Math. Phys.* **338**, 103 (2015).
- [41] H.-C. Cheng and M.-H. Hsieh, Moderate deviation analysis for classical-quantum channels and quantum hypothesis testing, *IEEE Trans. Inf. Theory* **64**, 1385 (2018).
- [42] C. T. Chubb, V. Y. F. Tan, and M. Tomamichel, Moderate deviation analysis for classical communication over quantum channels, *Commun. Math. Phys.* **355**, 1283 (2017).
- [43] A. Anshu, R. Jain, and N. A. Warsi, Building blocks for communication over noisy quantum networks, *IEEE Trans. Inf. Theory* **65**, 1287 (2019).
- [44] It is generally believed that the coefficients in Eq. (2) or in the quantum union bound cannot be removed [29]. However, achieving a tighter one-shot bound in quantum information theory without such coefficients is possible. That is the keynote of the present paper.
- [45] A one-shot achievability bound without the coefficients in terms of  $c$  was actually proved by Beigi and Gohari [91], but the comparison to Hayashi and Nagaoka's bound [30, Lemma 3], [46, Theorem 1] was missing. We compare Beigi and Gohari's result [91, Corollary 1] with ours (Theorem 1) in Sec. III A.
- [46] L. Wang and R. Renner, One-shot classical-quantum capacity and hypothesis testing, *Phys. Rev. Lett.* **108**, eid 200501 (2012).
- [47] V. P. Belavkin, Optimal multiple quantum statistical hypothesis testing, *Stochastics* **1**, 315 (1975).
- [48] P. Hausladen and W. K. Wootters, A "pretty good" measurement for distinguishing quantum states, *J. Mod. Opt.* **41**, 2385 (1994).
- [49] The noncommutative minimal is indeed unique. We refer the reader to Sec. II for a more precise statement.
- [50] C. W. Helstrom, Detection theory and quantum mechanics, *Inf. Control* **10**, 254 (1967).
- [51] A. Holevo, The analogue of statistical decision theory in the noncommutative probability theory, *Proc. Moscow Math. Soc.* **26**, 133 (1972).
- [52] A. S. Holevo, in *Proc. Steklov Inst. Math.*, Vol. 124 (American Mathematical Soc., New York City, United States, 1978), p. 1.
- [53] Note here that such an operational interpretation in quantum statistical decision theory applies to general self-adjoint positive semidefinite operators  $A$  and  $B$  whose traces are not necessarily normalized. One can think that operators  $A$  and  $B$  already incorporate the prior probabilities and the penalty for the erroneous decision; see, e.g., Refs. [178, Sec. 2.2.2] and [179, Eq. (1)].
- [54] A. Gilyén, S. Lloyd, I. Marvian, Y. Quek, and M. M. Wilde, Quantum algorithm for Petz recovery channels and pretty good measurements, *Phys. Rev. Lett.* **128**, 220502 (2022).
- [55] A. Serafini, *Quantum Continuous Variables: A Primer of Theoretical Methods* (CRC Press, Boca Raton, Florida, United States, 2017).
- [56] A. S. Holevo, On entanglement-assisted classical capacity, *J. Math. Phys.* **43**, 4326 (2002).
- [57] A. S. Holevo, *Quantum Systems, Channels, Information: A Mathematical Introduction* (De Gruyter, Berlin, Germany, 2012).
- [58] A. Anshu, V. K. Devabathini, and R. Jain, Quantum communication using coherent rejection sampling, *Phys. Rev. Lett.* **119**, 120506 (2017).
- [59] M. M. Wilde, Position-based coding and convex splitting for private communication over quantum channels, *Quantum Inf. Process.* **16**, 264 (2017).
- [60] H. Qi, Q. Wang, and M. M. Wilde, Applications of position-based coding to classical communication over quantum channels, *J. Phys. A: Math. Theor.* **51**, 444002 (2018).
- [61] The quantum state discrimination problems are usually concerned with distinguishing an ensemble of quantum states where each state (i.e., a density operator with unit trace) in the ensemble is endowed with a prior probability. For example, the *minimum error* defined above coincides with the error probability in conventional quantum state

- discrimination when assuming that  $\text{Tr}[A + B] = 1$ . We note that in Holevo's early works [51,62], [52, Sec. II], the scenario of distinguishing positive semidefinite operators (even on infinite-dimensional Hilbert space) was studied (see also Refs. [69] and [64, Sec. 3]).
- [62] A. S. Holevo, Remarks on optimal quantum measurements, *Probl. Inf. Transm.* **10**, 51 (1974).
- [63] K. Yanagi, K. Kuriyama, and S. Furuichi, Generalized Shannon inequalities based on Tsallis relative operator entropy, *Linear Algebra Appl.* **394**, 109 (2005).
- [64] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, England, 2018).
- [65] One can also define the noncommutative minimal among multiple self-adjoint operators. Throughout this paper, we only consider the case of two positive semidefinite operators.
- [66] In quantum state discrimination, one sometimes studies the maximum success probability [50–52] and expresses it via a semidefinite programming formulation as the trace of the so-called *noncommutative maximal* [64,69,103], i.e., the least (in trace ordering) of the upper bound (in Loewner partial ordering). Since the present paper aims to relate the *error* of a quantum information-theoretic task to that of quantum state discrimination, we only focus on the error. Note also that, for distinguishing multiple operators, say  $\{A_i\}_i$ , the error of the discrimination is given by the noncommutative minimal among the set of operators  $\{\sum_{j \neq i} A_j\}_i$  [103, Sec. II].
- [67] Some of the properties can be proved by using the quantum Hockey-Stick divergence [180–182] as well. In this paper, we just stick to the notation of the noncommutative minimal.
- [68] In case that  $A + B$  is not invertible, one just uses the Moore-Penrose pseudoinverse of  $A + B$  in the definition of the noncommutative quotient, Eq. (3).
- [69] K. M. R. Audenaert and M. Mosonyi, Upper bounds on the error probabilities and asymptotic error exponents in quantum multiple state discrimination, *J. Math. Phys.* **55**, 102201 (2014).
- [70] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2009).
- [71] K. M. Audenaert, Comparisons between quantum state distinguishability measures, *Quantum Inf. Comput.* **14**, 31 (2014).
- [72] H. Barnum and E. Knill, Reversing quantum dynamics with near-optimal quantum and classical fidelity, *J. Math. Phys.* **43**, 2097 (2002).
- [73] Y.-C. Shen, L. Gao, and H.-C. Cheng, Strong converse for privacy amplification against quantum side information, *ArXiv:2202.10263* (2022).
- [74] Y. Polyanskiy, Ph.D. thesis, Princeton University, 2010.
- [75] N. Datta, Min- and max-relative entropies and a new entanglement monotone, *IEEE Trans. Inf. Theory* **55**, 2816 (2009).
- [76] R. Renner, Ph.D. thesis (ETH), *ArXiv:quant-ph/0512258* (2005).
- [77] R. König, R. Renner, and C. Schaffner, The operational meaning of min- and max-entropy, *IEEE Trans. Inf. Theory* **55**, 4337 (2009).
- [78] M. Mosonyi and T. Ogawa, Strong converse exponent for classical-quantum channel coding, *Commun. Math. Phys.* **355**, 373 (2017).
- [79] M. Mosonyi and T. Ogawa, Divergence radii and the strong converse exponent of classical-quantum channel coding with constant compositions, *IEEE Trans. Inf. Theory* **67**, 1668 (2021).
- [80] H.-C. Cheng, E. P. Hanson, N. Datta, and M.-H. Hsieh, Non-asymptotic classical data compression with quantum side informations, *IEEE Trans. Inf. Theory* **67**, 902 (2021).
- [81] D. Petz, Quasi-entropies for finite quantum systems, *Rep. Math. Phys.* **23**, 57 (1986).
- [82] N. Datta, Y. Pautrat, and C. Rouzé, Second-order asymptotics for quantum hypothesis testing in settings beyond i.i.d.-quantum lattice systems and more, *J. Math. Phys.* **57**, 062207 (2016).
- [83] Y. Pautrat and S. Wang, Ke Li's lemma for quantum hypothesis testing in general von Neumann algebras, *ArXiv:2010.02177* (2020).
- [84] By large deviations, we meant the scenario where the coding rate  $R = 1/n \log M$  deviates from the fundamental threshold (i.e., the quantum mutual information or the Holevo quantity when optimizing the input distributions) by a *large amount*; namely, it is constant  $O(1)$  away from the fundamental limit. In the small deviation regime, the optimal rate  $R$  for some fixed error  $\varepsilon \in (0, 1)$  converges to the fundamental limit at a speed of  $O(1/\sqrt{n})$ , meaning that  $R$  deviates by a *small amount*. In between, we refer to the *moderate deviation regime*, where  $R$  deviates by an order of  $\omega(1/\sqrt{n}) \cap o(1)$  [41,42]. In this case, the optimal error vanishes at subexponential speed.
- [85] B. Huang, S. Jiang, Z. Song R. Tao, and R. Zhang, in *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, Denver, United States, 2022).
- [86] We use the notation  $O^*$  to hide  $m^{o(1)}$  and  $\log(1/\varepsilon)$  factors, where  $\varepsilon$  is the accuracy parameter.
- [87] J. Demmel, I. Dumitriu, O. Holtz, and R. Kleinberg, Fast matrix multiplication is stable, *Numer. Math.* **106**, 199 (2007).
- [88] J. Banks, J. Garza-Vargas, and A. Kulkarni, Pseudospectral shattering, the sign function, and diagonalization in nearly matrix multiplication time, *ArXiv:1912.08805* (2019).
- [89] M. V. Burnashev and A. S. Holevo, On the reliability function for a quantum communication channel, *Probl. Inf. Transm.* **34**, 97 (1998).
- [90] More precisely, Hayashi and Nagaoka obtained the first one-shot achievability bound (for general  $c$ - $q$  channels) as in Eq. (13) but in terms of the *information-spectrum divergence*  $D_s^\varepsilon$  [16,17,30,38,93] instead of the hypothesis-testing divergence  $D_h^\varepsilon$  defined in Proposition 2. On the other hand, it is known that  $D_h^\varepsilon(\cdot\|\cdot) \geq D_s^\varepsilon(\cdot\|\cdot)$  [38, Lemma 12], and, hence, the one-shot achievability bound in terms of  $D_h^\varepsilon$  is tighter than that in terms of  $D_s^\varepsilon$ . Here, we remark that the approach proposed by Hayashi and Nagaoka [30] allows for choosing any measurement along with applying the Hayashi-Nagaoka inequality (2) in establishing achievability. Namely, the analysis



- in Ref. [30] with Eq. (2) can already lead to Eq. (13). We remark that the terminology and concept of the hypothesis-testing divergence  $D_h^\varepsilon$  might already appear in the contexts of statistical hypothesis testing by Stein and Chernoff [183], Strassen [12], Csiszár–Longo [184], Polyanskiy *et al.* [22], and by Wang and Renner [46] in the quantum setting.
- [91] S. Beigi and A. Gohari, Quantum achievability proof via collision relative entropy, *IEEE Trans. Inf. Theory* **60**, 7980 (2014).
- [92] M. H. Yassaee, M. R. Aref, and A. Gohari, in *2013 IEEE International Symposium on Information Theory* (IEEE, Istanbul, Turkey, 2013), p. 1287.
- [93] H. Nagaoka and M. Hayashi, An information-spectrum approach to classical and quantum hypothesis testing for simple hypotheses, *IEEE Trans. Inf. Theory* **53**, 534 (2007).
- [94] In the IID asymptotic setting, the leading term (i.e., either  $D_h^{\varepsilon-\delta}$  or  $D_s^{\varepsilon-\delta}$ ) dominates both the first-order and the second-order coding rates. On the other hand, Eq. (15) does have a better constant  $-\log(1-\varepsilon)$ , which corresponds to the *fourth-order term* in the small deviation regime. Such a term is negligible for small errors (say, e.g.,  $\varepsilon \leq 10^{-3}$ ). For large errors, one can invoke the strengthened one-shot bound in Eq. (11) (though the formula is more involved).
- [95] The second-order expansion of the quantum hypothesis-testing divergence was concurrently proposed by Tomamichel and Hayashi [38] and Li [39]. Here in Eq. (17), we cite Li’s result [39, Theorem 5] since it has a better third-order term,  $-O(1)$ , than that of Tomamichel and Hayashi [38, Eqs. (28) and (33)], in which the third-order term is  $-(\frac{1}{2} + \min(\lambda(\sigma), \nu(\sigma)) \log n - O(1)$  with  $\lambda(\sigma)$  the *logarithmic condition number* of  $\sigma$  and  $\nu(\sigma)$  the number of distinct eigenvalues of  $\sigma$ . We also remark that Li’s result, Eq. (17), was later generalized to infinite-dimensional separable Hilbert space by Datta *et al.* [82, Proposition 1] and Oskouei *et al.* [29, Lemma A.1]. The same result for general von Neumann algebras was proved by Pautrat and Wang [83, Theorem 1].
- [96] We note that Altuğ and Wagner proved that the third-order coding rate  $O(1)$  is optimal for classical symmetric and singular channels [185, Proposition 1] (see also Ref. [37, Theorem 4.3]). In other words, if the nonsingularity condition is not imposed, a larger third-order coding rate than  $O(1)$  is not possible.
- [97] J. M. Renes, Achievable error exponents of data compression with quantum side information and communication over symmetric classical-quantum channels, [ArXiv:2207.08899](https://arxiv.org/abs/2207.08899) (2022).
- [98] M. Dalai, Lower bounds on the probability of error for classical and classical-quantum channels, *IEEE Trans. Inf. Theory* **59**, 8027 (2013).
- [99] M. Dalai and A. Winter, Constant compositions in the sphere packing bound for classical-quantum channels, *IEEE Trans. Inf. Theory* **63**, 5603 (2017).
- [100] H.-C. Cheng, M.-H. Hsieh, and M. Tomamichel, Quantum sphere-packing bounds with polynomial prefactors, [ArXiv:1704.05703](https://arxiv.org/abs/1704.05703) (2017).
- [101] T. Ogawa, in *The 38th Symposium on Information Theory and its Applications (SITA2015)* (IEEE, Kurashiki, Okayama, Japan, 2015).
- [102] A. E. Gamal and Y.-H. Kim, *Network Information Theory* (Cambridge University Press, Cambridge, England, 2011).
- [103] H. P. Yuen, R. S. Kennedy, and M. Lax, Optimum testing of multiple hypotheses in quantum detection theory, *IEEE Trans. Inf. Theory* **21**, 125 (1975).
- [104] F. Hiai and D. Petz, The proper formula for relative entropy and its asymptotics in quantum probability, *Commun. Math. Phys.* **143**, 99 (1991).
- [105] T. Ogawa and H. Nagaoka, Strong converse and Stein’s lemma in quantum hypothesis testing, *IEEE Trans. Inf. Theory* **46**, 2428 (2000).
- [106] T. Ogawa and M. Hayashi, On error exponents in quantum hypothesis testing, *IEEE Trans. Inf. Theory* **50**, 1368 (2004).
- [107] M. Hayashi, Optimal sequence of quantum measurements in the sense of Stein’s lemma in quantum hypothesis testing, *J. Phys. A: Math. General* **35**, 10759 (2002).
- [108] M. Nussbaum and A. Szkoła, The Chernoff lower bound for symmetric quantum hypothesis testing, *Ann. Stat.* **37**, 1040 (2009).
- [109] M. Mosonyi and T. Ogawa, Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies, *Commun. Math. Phys.* **334**, 1617 (2014).
- [110] M. Hayashi and M. Tomamichel, Correlation detection and an operational interpretation of the Rényi mutual information, *J. Math. Phys.* **57**, 102201 (2016).
- [111] H.-C. Cheng, M.-H. Hsieh, and M. Tomamichel, Quantum sphere-packing bounds with polynomial prefactors, *IEEE Trans. Inf. Theory* **65**, 2872 (2019).
- [112] H.-C. Cheng, L. Gao, and M.-H. Hsieh, in *2019 IEEE International Symposium on Information Theory (ISIT)* (IEEE, Rue Saint-Victor, Paris, 2019).
- [113] H.-C. Cheng, Ph.D. thesis, University of Technology Sydney, 2018.
- [114] H.-C. Cheng, A. Winter, and N. Yu, Discrimination of quantum states under locality constraints in the many-copy setting, [ArXiv:2011.13063](https://arxiv.org/abs/2011.13063).
- [115] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, Entanglement-assisted classical capacity of noisy quantum channels, *Phys. Rev. Lett.* **83**, 3081 (1999).
- [116] C. Bennett, P. Shor, J. Smolin, and A. Thapliyal, Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem, *IEEE Trans. Inf. Theory* **48**, 2637 (2002).
- [117] F. Dupuis, Ph.D. thesis, Université de Montréal, [ArXiv:1004.1641](https://arxiv.org/abs/1004.1641) (2010).
- [118] N. Datta and M.-H. Hsieh, One-shot entanglement-assisted quantum and classical communication, *IEEE Trans. Inf. Theory* **59**, 1929 (2013).
- [119] W. Matthews and S. Wehner, Finite blocklength converse bounds for quantum channels, *IEEE Trans. Inf. Theory* **60**, 7317 (2014).
- [120] The pretty-good measurement was also used in Ref. [186, Sec. 4] by following Beigi-Gohari’s approach [187],

- wherein the obtained one-shot expression and the asymptotic analysis are more involved.
- [121] S. P. Vadhan, *Pseudorandomness* (Now Publishers Inc, Norwell, Massachusetts, United States, 2012).
- [122] S. Strelchuk, M. Horodecki, and J. Oppenheim, Generalized teleportation and entanglement recycling, *Phys. Rev. Lett.* **110**, 010505 (2013).
- [123] H. Umegaki, Conditional expectation in an operator algebra, I, *Tohoku Math. J.* **6**, 177 (1954).
- [124] H. Umegaki, Conditional expectation in an operator algebra, II, *Tohoku Math. J.* **8**, 86 (1956).
- [125] H. Umegaki, Conditional expectation in an operator algebra, III, *Kodai Math. J.* **11**, 51 (1959).
- [126] E. Carlen, in *Contemporary Mathematics*, Vol. 529 (American Mathematical Society (AMS), Providence, Rhode Island, United States, 2010), p. 73.
- [127] Namely, the subalgebra consists of all operators in  $\mathcal{B}(\mathbb{R}^M \mathbb{B})$  of the form  $\mathbb{1}_{\mathbb{R}^{m-1}} \otimes \Upsilon_{\mathbb{R}_m \mathbb{B}} \otimes \mathbb{1}_{\mathbb{R}^{M-m}}$  for all  $\Upsilon_{\mathbb{R}_m \mathbb{B}} \in \mathcal{B}(\mathbb{R}_m \mathbb{B})$ .
- [128] A. Anshu, R. Jain, and N. A. Warsi, On the near-optimality of one-shot classical communication over quantum channels, *J. Math. Phys.* **60**, 012204 (2019).
- [129] We note that, by applying Theorem 3 in randomness-assisted communication over  $c$ - $q$  channels, it is equivalent to calculating the average error probability using a *mutually independent* random codebook, while in Theorem 1, we only require a *pairwise independent* random codebook.
- [130] I. Devetak and A. Winter, Classical data compression with quantum side information, *Phys. Rev. A* **68**, 042301 (2003).
- [131] J. M. Renes and R. Renner, One-shot classical data compression with quantum side information and the distillation of common randomness or secret keys, *IEEE Trans. Inf. Theory* **58**, 1985 (2012).
- [132] H.-C. Cheng, E. P. Hanson, N. Datta, and M.-H. Hsieh, Duality between source coding with quantum side information and classical-quantum channel coding, *IEEE Trans. Inf. Theory (Early Access)* **68**, 7315 (2022).
- [133] A. Winter, The capacity of the quantum multiple-access channel, *IEEE Trans. Inf. Theory* **47**, 3059 (2001).
- [134] M.-H. Hsieh, I. Devetak, and A. Winter, Entanglement-assisted capacity of quantum multiple-access channels, *IEEE Trans. Inf. Theory* **54**, 3078 (2008).
- [135] S. C. Xu and M. M. Wilde, Sequential, successive, and simultaneous decoders for entanglement-assisted classical communication, *Quantum Inf. Process.* **12**, 641 (2012).
- [136] J. Yard, P. Hayden, and I. Devetak, Quantum broadcast channels, *IEEE Trans. Inf. Theory* **57**, 7147 (2011).
- [137] I. Savov and M. M. Wilde, Classical codes for quantum broadcast channels, *IEEE Trans. Inf. Theory* **61**, 7017 (2015).
- [138] F. Dupuis, P. Hayden, and K. Li, A father protocol for quantum broadcast channels, *IEEE Trans. Inf. Theory* **56**, 2946 (2010).
- [139] J. Radhakrishnan, P. Sen, and N. Warsi, One-shot Marton inner bound for classical-quantum broadcast channel, *IEEE Trans. Inf. Theory* **62**, 2836 (2016).
- [140] Q. Wang, S. Das, and M. M. Wilde, Hadamard quantum broadcast channels, *Quantum Inf. Process.* **16**, 1 (2017).
- [141] H.-C. Cheng, N. Datta, and C. Rouze, Strong converse bounds in quantum network information theory, *IEEE Trans. Inf. Theory* **67**, 2269 (2021).
- [142] K. Marton, A coding theorem for the discrete memoryless broadcast channel, *IEEE Trans. Inf. Theory* **25**, 306 (1979).
- [143] A. E. Gamal and E. van der Meulen, A proof of Marton's coding theorem for the discrete memoryless broadcast channel (corresp.), *IEEE Trans. Inf. Theory* **27**, 120 (1981).
- [144] H.-C. Cheng and L. Gao, Tight one-shot analysis for convex splitting with applications in quantum information theory, *ArXiv:2304.12055* (2023).
- [145] S. I. Gel'fand and M. S. Pinsker, Coding for channel with random parameters, *Probl. Control Inf. Theory* **9**, 19 (1980).
- [146] N. Datta and F. Leditzky, Second-order asymptotics for source coding, dense coding, and pure-state entanglement conversions, *IEEE Trans. Inf. Theory* **61**, 582 (2015).
- [147] H.-C. Cheng and L. Gao, Error exponent and strong converse for quantum soft covering, *ArXiv:2202.10995* (2022).
- [148] H.-C. Cheng and L. Gao, Optimal second-order rates for quantum soft covering and privacy amplification, *ArXiv:2202.11590* (2022).
- [149] Otherwise, one could invoke Theorem 4.3 of Ref. [69]. However, the resulting error exponents will be weakened by one half.
- [150] F. Dupuis, Privacy amplification and decoupling without smoothing, *ArXiv:2105.05342* (2021).
- [151] I. Sason and S. Verdú, Arimoto-Rényi conditional entropy and Bayesian  $M$ -ary hypothesis testing, *IEEE Trans. Inf. Theory* **64**, 4 (2018).
- [152] J. M. Renes, Better bounds on optimal measurement and entanglement recovery, with applications to uncertainty and monogamy relations, *Phys. Rev. A* **96**, 042328 (2017).
- [153] F. Hiai, Matrix analysis: Matrix monotone functions, matrix means, and majorization, *Interdiscip. Inf. Sci.* **16**, 139 (2010).
- [154] F. Hiai and D. Petz, *Introduction to Matrix Analysis and Applications* (Springer International Publishing, Midtown Manhattan, New York City, United States, 2014).
- [155] T. Ando, Concavity of certain maps on positive definite matrices and applications to Hadamard products, *Linear Algebra Appl.* **26**, 203 (1979).
- [156] F. Kubo and T. Ando, Means of positive linear operators, *Math. Ann.* **246**, 205 (1980).
- [157] R. L. Frank and E. H. Lieb, Monotonicity of a relative Rényi entropy, *J. Math. Phys.* **54**, 122201 (2013).
- [158] M. Müller-Lennert, F. Dupuis, O. Szechr, S. Fehr, and M. Tomamichel, On quantum Rényi entropies: A new generalization and some properties, *J. Math. Phys.* **54**, 122203 (2013).
- [159] 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, *Commun. Math. Phys.* **331**, 593 (2014).

- [160] T. Ogawa and H. Nagaoka, Strong converse to the quantum channel coding theorem, *IEEE Trans. Inf. Theory* **45**, 2486 (1999).
- [161] S. Arimoto, Computation of random coding exponent functions, *IEEE Trans. Inf. Theory* **22**, 665 (1976).
- [162] C. H. Bennett and S. J. Wiesner, Communication via one- and two-particle operators on Einstein–Podolsky–Rosen states, *Phys. Rev. Lett.* **69**, 2881 (1992).
- [163] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Teleporting an unknown quantum state via dual classical and Einstein–Podolsky–Rosen channels, *Phys. Rev. Lett.* **70**, 1895 (1993).
- [164] R. Ahlswede and A. Winter, Strong converse for identification via quantum channels, *IEEE Trans. Inf. Theory* **48**, 569 (2002).
- [165] M. Hayashi, General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel, *IEEE Trans. Inf. Theory* **52**, 1562 (2006).
- [166] M. Hayashi, Quantum wiretap channel with non-uniform random number and its exponent and equivocation rate of leaked information, *IEEE Trans. Inf. Theory* **61**, 5595 (2015).
- [167] M. Hayashi, *Quantum Information Theory* (Springer Berlin Heidelberg, Berlin, Germany, 2017).
- [168] M. Tomamichel, C. Schaffner, A. Smith, and R. Renner, Leftover hashing against quantum side information, *IEEE Trans. Inf. Theory* **57**, 5524 (2011).
- [169] M. Hayashi, Tight exponential analysis of universally composable privacy amplification and its applications, *IEEE Trans. Inf. Theory* **59**, 7728 (2013).
- [170] M. Tomamichel, *Quantum Information Processing with Finite Resources* (Springer International Publishing, Midtown Manhattan, New York City, United States, 2016).
- [171] K. Li, Y. Yao, and M. Hayashi, Tight exponential analysis for smoothing the max-relative entropy and for quantum privacy amplification, *IEEE Trans. Inf. Theory* **69**, 1680 (2023).
- [172] F. Dupuis, M. Berta, J. Wullschleger, and R. Renner, One-shot decoupling, *Commun. Math. Phys.* **328**, 251 (2014).
- [173] K. Li and Y. Yao, Reliability function of quantum information decoupling, [ArXiv:2111.06343](https://arxiv.org/abs/2111.06343) (2021).
- [174] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels, *IEEE Trans. Inf. Theory* **60**, 2926 (2014).
- [175] Z. Luo and I. Devetak, Channel simulation with quantum side information, [ArXiv:quant-ph/0611008](https://arxiv.org/abs/quant-ph/0611008) (2006).
- [176] M. Berta, Quantum side information: Uncertainty relations, extractors, channel simulations, [ArXiv:1310.4581](https://arxiv.org/abs/1310.4581) (2013).
- [177] K. Li and Y. Yao, Reliable simulation of quantum channels, [ArXiv:2112.04475](https://arxiv.org/abs/2112.04475) (2021).
- [178] A. S. Holevo, *Statistical Structure of Quantum Theory* (Springer Berlin Heidelberg, Berlin, Germany, 2001).
- [179] K. Li, Discriminating quantum states: The multiple Chernoff distance, *Ann. Stat.* **44**, 1661 (2016).
- [180] Y. Polyanskiy and S. Verdú, in *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* (IEEE, Monticello, IL, United States, 2010).
- [181] N. Sharma and N. A. Warsi, Fundamental bound on the reliability of quantum information transmission, *Phys. Rev. Lett.* **110**, 080501 (2013).
- [182] C. Hirche, C. Rouzé, and D. S. França, Quantum differential privacy: An information theory perspective, [ArXiv:2202.10717](https://arxiv.org/abs/2202.10717) (2022).
- [183] H. Chernoff, Large-sample theory: Parametric case, *The Annals of Mathematical Statistics* **27**, 1 (1956).
- [184] I. Csiszár and G. Longo, On the error exponent for source coding and for testing simple statistical hypotheses, *Stud. Sci. Math. Hung.* **6**, 181 (1971).
- [185] Y. Altugğ and A. B. Wagner, in *2014 IEEE International Symposium on Information Theory (ISIT)* (IEEE, Honolulu, HI, USA, 2014).
- [186] N. Datta, M. Tomamichel, and M. M. Wilde, On the second-order asymptotics for entanglement-assisted communication, *Quantum Inf. Process.* **15**, 2569 (2016).
- [187] R. Bhatia and P. Grover, Norm inequalities related to the matrix geometric mean, *Linear Algebra Appl.* **437**, 726 (2012).