Communication Over Quantum Channels With Parameter Estimation

Uzi Pereg, *Member, IEEE*

***Abstract*— Communication over a random-parameter quantum channel when the decoder is required to reconstruct the parameter sequence is considered. We study scenarios that include either strictly-causal, causal, or non-causal channel side information (CSI) available at the encoder, and also when CSI is not available. This model can be viewed as a form of quantum metrology, and as the quantum counterpart of the classical rate-and-state channel with state estimation at the decoder. Regularized formulas for the capacity-distortion regions are derived. In the special case of measurement channels, single-letter characterizations are derived for the strictly-causal and causal settings. Furthermore, in the more general case of entanglement-breaking channels, a single-letter characterization is derived when CSI is not available. As a consequence, we obtain regularized formulas for the capacity of random-parameter quantum channels with CSI, generalizing previous results by Boche *et al.*, 2016, on classical-quantum channels. Bosonic dirty paper coding is introduced as a consequence, where we demonstrate that the optimal coefficient is not necessarily that of minimum mean-square error estimation as in the classical setting.**

***Index Terms*— Quantum communication, Shannon theory, state estimation, rate-and-state channel, bosonic channel, writing on dirty paper, encoding constraints.**

I. Introduction

A FUNDAMENTAL task in classical information theory is to determine the ultimate transmission rate of communication. Various settings of practical significance can be described by a channel $p_{Y|X,S}$ that depends on a random parameter $S \sim q(s)$ when there is channel side information (CSI) available at the transmitter [2]–[4]. For example, a cognitive radio in a wireless system may be aware of the channel state and network configuration [5]. Other applications include memory storage where the writer knows the fault locations [6], digital watermarking [7], and spread-spectrum communication [8], [9], where the CSI represents the host data or a pseudo-random sequence to be modulated.

In the rate-and-state (RnS) model [10], the receiver is not only required to recover the message, but also to estimate the parameter sequence with limited distortion. For example, in digital multicast [10], the message represents digital control information that is multicast on top of an existing analog transmission, which is also estimated by the receiver. Additional applications can be found in [3] and references therein. The capacity-distortion tradeoff region with strictly-causal CSI and with causal CSI was determined by Choudhuri *et al.* [3], and without CSI by Zhang *et al.* [11], [12]. Inner and outer bounds on the tradeoff region with non-causal CSI were derived by Sutivong in [13], with full characterization in the Gaussian case [10]. The RnS channel with feedback was recently considered by Bross and Lapidoth [14].

The field of quantum information is rapidly evolving in both practice and theory [15]–[19]. Quantum information theory is the natural extension of classical information theory. Nevertheless, this generalization reveals astonishing phenomena with no parallel in classical communication [20]. For example, two quantum channels, each with zero quantum capacity, can have a nonzero quantum capacity when used together [21]. This property is known as super-activation.

Communication through quantum channels can be separated into different categories. The Holevo-Schumacher-Westmoreland (HSW) Theorem provides a regularized ("multi-letter") formula for the capacity of a quantum channel [22], [23]. Although calculation of such a formula is intractable in general, it provides computable lower bounds, and there are special cases where the capacity can be computed exactly. The reason for this difficulty is that the Holevo information is not necessarily additive [24], [25]. Shor has demonstrated additivity for the class of entanglement-breaking channels [26], in which case the HSW theorem provides a single-letter computable formula for the capacity. This class includes both classical-quantum channels and measurement (quantum-classical) channels [27, Section 4.6.7]. A similar difficulty occurs with transmission of quantum information [28].

As for quantum channels with random parameters, Boche, Cai, and Nötzel [1] addressed the classical-quantum channel with CSI at the encoder. The capacity was determined given causal CSI, and a regularized formula was provided given non-causal CSI. Warsi and Coon [29] used an information-spectrum approach to derive multi-letter bounds for a similar setting, where the side information

Manuscript received March 21, 2021; revised August 20, 2021; accepted October 20, 2021. Date of publication October 26, 2021; date of current version December 23, 2021. This work was supported in part by the German Federal Ministry of Education and Research (BMBF) under Grant 16KIS0856 and in part by the Israel Council for Higher Education (CHE) Fellowship for Quantum Science and Technology. An earlier version of this paper was presented in part at the 2020 IEEE International Symposium on Information Theory and in part at the 2021 IEEE International Symposium on Information Theory.

The author is with the Institute for Communications Engineering, Technische Universität München, 80333 Munich, Germany (e-mail: uzi.pereg@tum.de).

Communicated by M. M. Wilde, Associate Editor for Quantum Information Theory.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2021.3123221.

Digital Object Identifier 10.1109/TIT.2021.3123221

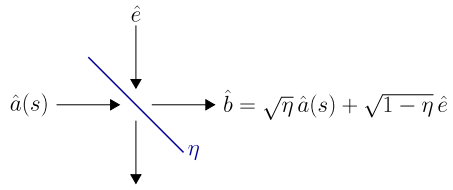

Fig. 1. The beam-splitter relation of the single-mode bosonic channel. The channel input is an electromagnetic field mode with an annihilation operator $\hat{a}$, and the output is another mode with the annihilation operator $\hat{b} = \sqrt{\eta}\,\hat{a} + \sqrt{1-\eta}\,\hat{e}$, where $\hat{e}$ is associated with the environment noise and the parameter η is the transmissivity, where $0 \leq \eta \leq 1$.

has a limited rate. Anshu *et al.* [30] have recently considered the fully quantum wiretap channel with CSI as well. The entanglement-assisted capacity of a quantum channel with non-causal CSI was determined by Dupuis [31], [32], and with causal CSI by the author [33], [34]. One-shot communication with CSI is considered in [35] as well. Luo and Devetak [36] considered channel simulation with source side information (SSI) at the decoder, and also solved the quantum generalization of the Wyner-Ziv problem [37]. Quantum data compression with SSI is also studied in [38]–[40]. State-dependent channels with environment assistance are considered in [41]–[43]. The dual setting of state masking, where the channel state is hidden from the receiver, was recently considered in [44]. Quantum relay channels are treated in [45], [46] using a decode-forward communication scheme with block Markov coding. Parameter estimation of quantum channels was previously studied from the algorithmic point of view in different settings [47]–[49].

Optical communication forms the backbone of the Internet [50]–[52]. The bosonic channel is a simple quantum-mechanical model for optical communication over free space or optical fibers [53], [54], and it can be viewed as the quantum counterpart of the classical channel with additive white Gaussian noise (AWGN). An optical communication system consists of a modulated source of photons, the optical channel, and an optical detector. For a single-mode bosonic channel, the channel input is an electromagnetic field mode with an annihilation operator $\hat{a}$, and the output is another mode with the annihilation operator $\hat{b}$. The input-output relation in the Heisenberg picture [55] is given by

$$\hat{b} = \sqrt{\eta}\,\hat{a} + \sqrt{1-\eta}\,\hat{e} \tag{1}$$

where $\hat{e}$ is associated with the environment noise and the parameter η is the transmissivity, $0 \leq \eta \leq 1$, which depends on the length of the optical fiber and its absorption length [56] (see Figure 1). For a lossy bosonic channel, the noise mode $\hat{e}$ is in a Gibbs thermal state. Modulation is performed such that the unitary displacement operator $D(\alpha) = \exp(\alpha\hat{a}^\dagger - \alpha^*\hat{a})$ is applied to the vacuum state $|0\rangle\langle 0|$ [53].

In this paper, we consider a random-parameter quantum channel when the decoder is required to reconstruct the parameter sequence in a lossy manner, *i.e.* with limited distortion. Here, we give two applications for this model: digital multicast using *quantum* communication channels, and classical watermarking with a quantum embedding. In the watermarking application, an authentication message is mixed within classical host data ("stegotext"), and this mixture is encoded into a quantum state that is sent to an authenticator. The random parameters in this setting represent the host data, while their estimation at the decoder corresponds to a scenario where the host data itself contains desirable information. Our setting can also be interpreted as a form of quantum metrology [57], where the decoder performs measurements on the received (quantum) systems in order to estimate classical noise parameters, while exploiting the entanglement generated by the encoder.

The scenarios that are studied in the present work include either strictly-causal, causal, or non-causal channel side information (CSI) available at the encoder, as well as the case where CSI is not available. With strictly-causal CSI, Alice knows the *past* parameters at each time instance; given causal CSI, she knows the *past and present* parameters; and with non-causal CSI, the entire sequence of random parameters is available to her a priori. This model can be viewed as the quantum analog of the classical RnS channel. We derive regularized formulas for the capacity-distortion tradeoff regions. In the special case of measurement channels, single-letter characterizations are established for the strictly-causal and causal settings. Furthermore, in the more general case of entanglement-breaking channels, a single-letter characterization is derived when CSI is not available. As a consequence, we obtain regularized formulas for the capacity of random-parameter quantum channels with strictly-causal, causal, or non-causal CSI, generalizing the previous results by Boche *et al.* [1] on classical-quantum channels.

Considering entanglement-breaking channels without CSI, we use a different approach from that of Shor [26]. As opposed to Shor [26], we do not show additivity of the capacity formula, but rather extend the methods of Wang *et al.* [58] to prove the converse part in a more direct manner. To prove achievability with strictly-causal CSI, we extend the classical block Markov coding method from [3] to the quantum setting, and then apply the quantum packing lemma [59] for decoding the message, and the classical covering lemma for the reconstruction of the parameter sequence. The gentle measurement lemma [60], [61] alleviates the proof, as it guarantees that multiple decoding measurements can be performed without collapsing the quantum state and such that the output state after each measurement is almost the same. Thus, we can separate between measurements for recovering the message and for sequence reconstruction. Achievability with causal CSI is proved using similar techniques with the addition of a quantum "Shannon-strategy" encoding operation [62] [34, Section IV.D]. To prove achievability with non-causal CSI, we use an extension of the classical binning technique [6] to the quantum setting.

Furthermore, we introduce the bosonic dirty paper setting as a special case. We consider the single-mode lossy bosonic channel with a coherent-state protocol and a non-ideal displacement operation in the modulation process:

$$|\zeta_1\zeta_2\cdots\zeta_n\rangle = D(\alpha_1 + s_1)|0\rangle \otimes \cdots \otimes D(\alpha_n + s_n)|0\rangle \tag{2}$$

where the parameter s_i represents classical interference in the transmission equipment, which the transmitter becomes aware of, while the receiver is not. It is assumed that the input has an

average power constraint $\frac{1}{n}\sum_{i=1}^{n}|\alpha_i|^2 \leq N_A$. Alternatively, this can be viewed as a watermarking model with a quantum embedding. Given a classical host data sequence $s_1,\ldots,s_n$, Alice encodes an authentication message m into a watermark $(\alpha_i(m,s_1,\ldots,s_n))_{i=1}^n$. Next, Alice performs a quantum embedding of the watermark; she prepares a *watermarked state* $|\zeta_1\zeta_2\cdots\zeta_n\rangle$ as in (2), and transmits it to the authenticator Bob through the optical fiber. The capacity of the random-parameter bosonic channel represents the optimal rate at which the authenticator can recover the messages with high fidelity.

First, we consider homodyne and heterodyne detection. Both of those settings reduce to a classical random-parameter channel with either real or complex-valued Gaussian noise. Thereby, we observe that based on Costa's dirty-paper solution, the effect of the classical interference can be canceled, and the capacity is the same regardless of the intensity of the interference. Then, we consider joint detection, in which case, the problem does not reduce to that of a classical description. We derive a dirty-paper coding lower bound based on the above results, with a general coefficient t (see (77)). Considering the special case of a pure-loss bosonic channel, we show that the optimal coefficient is not necessarily that of minimum mean-square error (MMSE) estimation value as in (78).

The paper is organized as follows. In Section II, we give the definitions and present the models. In Section III, we provide a brief review of related work on channels without random parameters, regularization, additivity, and entanglement-breaking channels; as well as a comparison between Shor's original approach for single-letterization, based on additivity, and the alternative argument that extends the methods in [58]. In Section IV, we state our main results on the random-parameter quantum channel with parameter estimation at the decoder. In Section V, we consider bosonic dirty paper coding as a consequence of the main results. Section VI is dedicated to summary and discussion, where we summarize our main results and conclude with remarks on the comparison between the classical and quantum dirty-paper settings. The proofs are given in the appendix.

II. Definitions

A. Notation, States, and Information Measures

We use the following notation conventions. Calligraphic letters $\mathcal{X},\mathcal{Y},\mathcal{Z},\ldots$ are used for finite sets. Lowercase letters $x,y,z,\ldots$ represent constants and values of classical random variables, and uppercase letters $X,Y,Z,\ldots$ represent classical random variables. The distribution of a random variable X is specified by a probability mass function (pmf) $p_X(x)$ over a finite set $\mathcal{X}$. We use $x^j=(x_1,x_2,\ldots,x_j)$ to denote a sequence of letters from $\mathcal{X}$. A random sequence X^n and its distribution $p_{X^n}(x^n)$ are defined accordingly. For a pair of integers i and j, $1\leq i\leq j$, we write a discrete interval as $[i:j]=\{i,i+1,\ldots,j\}$.

The state of a quantum system A is given by a density operator ρ on the Hilbert space $\mathcal{H}_A$. Unless mentioned otherwise, we assume that the Hilbert spaces have finite dimensions. A density operator is an Hermitian, positive semidefinite operator, with unit trace, *i.e.* $\rho^\dagger=\rho$, $\rho\succeq 0$, and $\mathrm{Tr}(\rho)=1$. The state is said to be pure if $\rho=|\psi\rangle\langle\psi|$, for some vector $|\psi\rangle\in\mathcal{H}_A$, where $\langle\psi|=(|\psi\rangle)^\dagger$. A measurement of a quantum system is any set of operators $\{\Lambda_j\}$ that forms a positive operator-valued measure (POVM), *i.e.* the operators are positive semi-definite and $\sum_j\Lambda_j=\mathbb{1}$, where $\mathbb{1}$ is the identity operator. According to the Born rule, if the system is in state ρ, then the probability of the measurement outcome j is given by $p_A(j)=\mathrm{Tr}(\Lambda_j\rho)$. The qubit Pauli basis is denoted by $\{\mathbb{1},\mathsf{X},\mathsf{Y},\mathsf{Z}\}$.

Define the quantum entropy of the density operator ρ as

$$H(\rho)\triangleq-\mathrm{Tr}[\rho\log(\rho)] \tag{3}$$

which is the same as the Shannon entropy associated with the eigenvalues of ρ. Given a bipartite state σ_{AB} on $\mathcal{H}_A\otimes\mathcal{H}_B$, define the quantum mutual information by

$$I(A;B)_\sigma=H(\sigma_A)+H(\sigma_B)-H(\sigma_{AB}). \tag{4}$$

The conditional quantum entropy and mutual information are defined by $H(A|B)_\sigma=H(\sigma_{AB})-H(\sigma_B)$ and $I(A;B|C)_\sigma=H(A|C)_\sigma+H(B|C)_\sigma-H(A,B|C)_\sigma$, respectively.

A pure bipartite state is called *entangled* if it cannot be expressed as the tensor product of two states in $\mathcal{H}_A$ and $\mathcal{H}_B$. The maximally entangled state between two systems of dimension D is defined by $|\Phi_{AB}\rangle=\frac{1}{\sqrt{D}}\sum_{j=0}^{D-1}|j\rangle_A\otimes|j\rangle_B$, where $\{|j\rangle_A\}_{j=0}^{D-1}$ and $\{|j\rangle_B\}_{j=0}^{D-1}$ are respective orthonormal bases. Note that $I(A;B)_{|\Phi\rangle\langle\Phi|}=2\cdot\log(D)$.

B. Quantum Channels With Random Parameters

A quantum channel maps a quantum state at the sender system to a quantum state at the receiver system. Here, we consider a model of channel uncertainty, where the channel is governed by a random parameter that *changes over time*. Formally, let $\{\mathcal{N}_{A\to B}^{(s)},\ s\in\mathcal{S}\}$ be a collection of linear, completely positive, and trace-preserving (CPTP) maps, indexed by s, each corresponds to a quantum physical evolution. It is assumed that the channel has a product form, *i.e.* if the systems $A^n=(A_1,\ldots,A_n)$ are sent through n channel uses, then the input state ρ_{A^n} undergoes the tensor product mapping

$$\mathcal{N}_{A^n\to B^n}^{(s^n)}\equiv\bigotimes_{i=1}^{n}\mathcal{N}_{A\to B}^{(s_i)}. \tag{5}$$

We consider a quantum channel with a memoryless random parameter sequence, where the parameter sequence $(S_1,S_2,\ldots)$ is i.i.d. $\sim q(s)$. That is, the joint distribution of the parameter sequence is given by $\Pr(S^n=s^n)=q^n(s^n)\equiv\prod_{i=1}^{n}q(s_i)$. Therefore, without CSI, the input-output relation is

$$\begin{aligned}\rho_{B^n}&=\sum_{s^n\in\mathcal{S}^n}q^n(s^n)\mathcal{N}_{A^n\to B^n}^{(s^n)}(\rho_{A^n})\\&=\left(\sum_{s\in\mathcal{S}}q(s)\mathcal{N}_{A\to B}^{(s)}\right)^{\otimes n}(\rho_{A^n}).\end{aligned} \tag{6}$$

The sender and the receiver are often referred to as Alice and Bob.

Equivalently, the random-parameter quantum channel can be defined by a CPTP map $\mathcal{N}_{SA\to B}$ with a bi-partite input, such that the component S is a classical system in a given fixed state, *i.e.*

$$\rho_S = \sum_{s\in\mathcal{S}} q(s)|s\rangle\langle s| \tag{7}$$

where $\{|s\rangle\}_{s\in\mathcal{S}}$ is an orthonormal basis of the Hilbert space $\mathcal{H}_S$. Given CSI at the encoder, *i.e.* when Alice has access to the parameter sequence, or a part of it, then the input system A^n can be correlated with S^n as well.

We will also consider the quantum-classical special case.

Definition 1: A measurement channel (or, q-c channel) $\mathcal{M}_{A\to Y}$ has the following form,

$$\mathcal{M}_{A\to Y}(\rho_A) = \sum_{y\in\mathcal{Y}} \mathrm{Tr}(\Lambda_y\rho_A)|y\rangle\langle y| \tag{8}$$

for some POVM $\{\Lambda_y\}$ and orthonormal vectors $\{|y\rangle\}$. A random-parameter channel is called a measurement channel when the collection of CPTP maps consists of q-c channels. We denote the random-parameter measurement channel by $\mathcal{M}_{SA\to Y}$ to distinguish it from the general channel $\mathcal{N}_{SA\to B}$.

A more general class of channels is that of entanglement-breaking channels. The definition is given below.

Definition 2: A quantum channel $\mathcal{E}_{A\to B}$ is called entanglement breaking if for every input state $\rho_{AA'}$, where A' is an arbitrary reference system, the channel output is separable, *i.e.*

$$(\mathcal{E}_{A\to B}\otimes\mathbb{1})(\rho_{AA'}) = \sum_{x\in\mathcal{X}} p_X(x)\psi_B^x\otimes\psi_{A'}^x \tag{9}$$

for some probability distribution $p_X(x)$ and pure states ψ_B^x and $\psi_{A'}^x$. We say that a random-parameter channel $\mathcal{N}_{SA\to B}$ is entanglement-breaking if each $\mathcal{N}_{A\to B}^{(s)}$ is entanglement breaking, for $s\in\mathcal{S}$.

Every entanglement-breaking channel $\mathcal{E}_{A\to B}$ can be represented as a serial concatenation of a measurement channel followed by a classical-quantum channel [27, Corollary 4.6.1]. That is, if $\mathcal{E}_{A\to B}$ is entanglement breaking, then there exists a pair of channels, $\mathcal{P}_{Y\to B}$ and $\mathcal{M}_{A\to Y}$, such that

$$\mathcal{E}_{A\to B} = \mathcal{P}_{Y\to B}\circ\mathcal{M}_{A\to Y} \tag{10}$$

where Y is classical.

C. Coding

We define a code to transmit classical information. We will address four CSI scenarios. With strictly-causal CSI, Alice knows, the *past* random parameters S^{i-1}; given causal CSI, she knows the *past and present* parameters S^i; with non-causal CSI, the entire sequence S^n is available to her a priori; and without CSI, Alice is ignorant. In all of those cases, Bob is unaware of the random parameters, and he has two tasks to perform. He is required to decode the message and to reconstruct the parameter sequence S^n with a limited distortion. Let $d:\mathcal{S}\times\hat{\mathcal{S}}\to[0,\infty)$ be a bounded distortion function, with $d_{\max}\equiv\max_{s,\hat{s}} d(s,\hat{s})$. Denote the average distortion between a parameter sequence s^n and a reconstruction sequence $\hat{s}^n$ by

$$d^n(s^n,\hat{s}^n) \triangleq \frac{1}{n}\sum_{i=1}^n d(s_i,\hat{s}_i). \tag{11}$$

Definition 3: A $(2^{nR},n)$ code with strictly-causal CSI at the encoder consists of the following: a message set $[1:2^{nR}]$, where 2^{nR} is assumed to be an integer, a sequence of encoding maps (channels) $\mathcal{F}_{MS^{i-1}\to A^i}^{(i)}$ for $i\in[1:n]$, and a decoding POVM $\{\Lambda_{B^n}^{m,\hat{s}^n}\}_{m\in[1:2^{nR}],\hat{s}^n\in\hat{\mathcal{S}}^n}$. The encoding maps must be consistent in the sense that the states $\rho_{A^i}^{m,s^{i-1}} \equiv \mathcal{F}_{M,S^{i-1}\to A^i}^{(i)}(m,s^{i-1})$ satisfy $\mathrm{Tr}_{A_{i+1}^n}(\rho_{A^n}^{m,s^{n-1}}) = \rho_{A^i}^{m,s^{i-1}}$ for $i\in[1:n]$. We denote the code by $(\mathcal{F},\Lambda)$.

The communication scheme is depicted in Figure 2. The sender Alice has the systems A^n and the receiver Bob has the systems B^n. Alice chooses a classical message $m\in[1:2^{nR}]$. At time $i\in[1:n]$, Alice has the sequence of past parameters $s^{i-1}\in\mathcal{S}^{i-1}$, and can thus prepare the state $\rho_{A_i}^{m,s^{i-1}}$ and transmit the system A_i over the channel $\mathcal{N}_{SA\to B}$.

Bob receives the channel output systems B^n and performs the POVM $\{\Lambda_{B^n}^{m,\hat{s}^n}\}_{m\in[1:2^{nR}],\hat{s}^n\in\hat{\mathcal{S}}^n}$. The conditional probability of decoding error, given that the message m was sent, is given by

$$P_{e|m}^{(n)}(\mathcal{F},\Lambda) = \mathrm{Tr}\left[\mathbb{1} - \sum_{\hat{s}^n\in\hat{\mathcal{S}}^n}\Lambda_{B^n}^{m,\hat{s}^n}\sum_{s^n\in\mathcal{S}^n} q^n(s^n)\mathcal{N}_{A^n\to B^n}^{(s^n)}(\rho_{A^n}^{m,s^{n-1}})\right]. \tag{12}$$

The average distortion for the code $(\mathcal{F},\Lambda)$ is

$$\Delta^{(n)}(\mathcal{F},\Lambda) \triangleq \sum_{s^n\in\mathcal{S}^n}\sum_{\hat{s}^n\in\hat{\mathcal{S}}^n} d^n(s^n,\hat{s}^n)\Pr\left(S^n=s^n,\hat{S}^n=\hat{s}^n\right) \tag{13}$$

where

$$\Pr\left(S^n=s^n,\hat{S}^n=\hat{s}^n\right) = q^n(s^n)\cdot \frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}}\sum_{\hat{m}=1}^{2^{nR}}\mathrm{Tr}\left[\Lambda_{B^n}^{\hat{m},\hat{s}^n}\mathcal{N}_{A^n\to B^n}^{(s^n)}(\rho_{A^n}^{m,s^{n-1}})\right]. \tag{14}$$

A $(2^{nR},n,\varepsilon,D)$ rate-distortion code satisfies $P_{e|m}^{(n)}(\mathcal{F},\Lambda)\le\varepsilon$ for all $m\in[1:2^{nR}]$, and $\Delta^{(n)}(\mathcal{F},\Lambda)\le D$. A rate $R>0$ is called achievable with distortion D if for every $\varepsilon>0$ and sufficiently large n, there exists a $(2^{nR},n,\varepsilon,D)$ code. The capacity-distortion region $\mathbb{C}_{\text{s-c}}(\mathcal{N})$ is defined as the set of achievable pairs (R,D) with strictly-causal CSI.

Alternatively, one may fix the average distortion constraint $D>0$ and consider the optimal transmission rate. The capacity-distortion function $C_{\text{s-c}}(\mathcal{N},D)$ is defined as the supremum of achievable rates R for a given distortion D. Note that $C_{\text{s-c}}(\mathcal{N},d_{\max})$ reduces to the standard definition of the capacity of a quantum channel, without a distortion requirement or parameter estimation by the decoder.

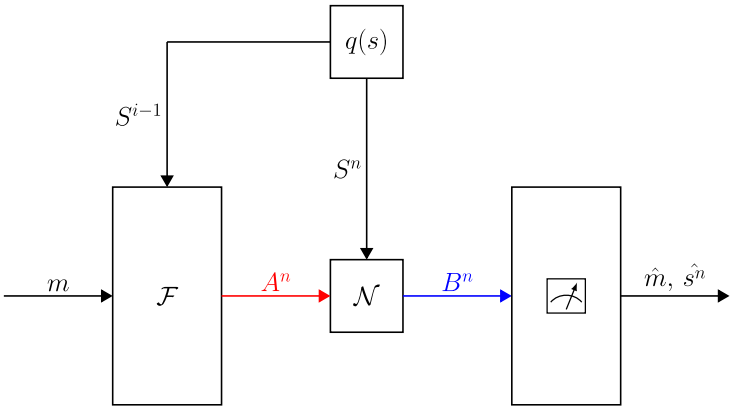

Fig. 2. Coding for a quantum channel $\mathcal{N}_{SA\to B}$ that depends on a random parameter $S\sim q(s)$, with strictly-causal side information at the encoder and parameter estimation at the decoder. The quantum systems of Alice and Bob are marked in red and blue, respectively. Alice chooses a classical message m. At time i, given the parameter sequence s^{i-1}, her encoder $\mathcal{E}$ prepares a state $\rho_{A_i}^{m,s^{i-1}}$, and then transmits the system A_i over the quantum channel $\mathcal{N}_{SA\to B}$. Bob receives the channel output systems B^n and performs a measurement. The outcome is the estimated message $\hat{m}$ and reconstruction sequence $\hat{s}^n$. With causal side information or non-causal side information, S^{i-1} is replaced by S^i or S^n, respectively.

We also address the causal and the non-causal setting. In the causal setting, Alice has the present parameter value S_i as well, and prepares $\rho_{A^n}^{m,s^n}$ such that $\rho_{A^i}^{m,s^n}\equiv\mathcal{E}^{(i)}_{MS^i\to A^i}(m,s^i)$. Whereas, in the non-causal setting, Alice has the entire parameter sequence S^n a priori, and can thus prepare $\rho_{A^n}^{m,s^n}$ of any form. Without CSI, Alice sends a sequence in the state $\rho_{A^n}^m=\mathcal{E}_{M\to A^n}(m)$ that is independent of the parameter sequence. We use the subscripts 's-c', 'caus', or 'n-c' to indicate whether CSI is available at the encoder in a stictly-causal, causal, or non-causal manner, respectively. The notation is summarized in the table in Figure 3.

III. Related Work

In this section, we briefly review known results for a quantum channel that does not depend on a random parameter and has no distortion constraint, *i.e.* $\mathcal{N}^{(s)}_{A\to B}=\mathcal{E}_{A\to B}$ for $s\in\mathcal{S}$, and $D=d_{\max}$. We also bring a general discussion on regularization, additivity, and entanglement-breaking channels. We compare between Shor's original approach for single-letterization, based on additivity, and an alternative argument that follows from the methods by Wang *et al.* [58]. In the sequel, we will use those observations in our capacity-distortion analysis in the absence of CSI.

A. HSW Theorem

Consider a channel $\mathcal{E}_{A\to B}$ without random parameters. Define

$$\chi(\mathcal{E})\triangleq\max_{p_X(x),|\phi_A^x\rangle} I(X;B)_\rho \tag{15}$$

with $\rho_{XB}\equiv\sum_{x\in\mathcal{X}}p_X(x)|x\rangle\langle x|\otimes\mathcal{E}(|\phi_A^x\rangle\langle\phi_A^x|)$ and $|\mathcal{X}|\le|\mathcal{H}_A|^2$. The objective functional $I(X;B)_\rho$ is referred to as the Holevo information with respect to the ensemble $\{p_X(x),\mathcal{E}(|\phi_A^x\rangle\langle\phi_A^x|)\}$ and the channel $\mathcal{E}_{A\to B}$, while the formula $\chi(\mathcal{E})$ itself is sometimes referred to as the Holevo information of the channel [27]. Next, we cite the HSW Theorem, which provides a regularized capacity formula for a quantum channel without parameters or distortion requirement.

Theorem 1 (See [22], [23], [26]):

1) The capacity of a quantum channel $\mathcal{E}_{A\to B}$ without parameters is given by

$$C(\mathcal{E},d_{\max})=\lim_{k\to\infty}\frac{1}{k}\chi\left(\mathcal{E}^{\otimes k}\right). \tag{16}$$

2) If $\mathcal{E}_{A\to B}$ is entanglement-breaking, then

$$C(\mathcal{E},d_{\max})=\chi(\mathcal{E}). \tag{17}$$

In the second part of the lemma, we included Shor's result for the class of entanglement-breaking channels [26] (see Definition 2). We note that this class includes both classical-quantum channels and measurement channels. In particular, the capacity of a measurement channel $\mathcal{M}^{(0)}_{A\to Y}$ without parameters is given by

$$C(\mathcal{M}^{(0)},d_{\max})=\max_{p_X(x),|\phi_A^x\rangle} I(X;Y) \tag{18}$$

with $p_{Y|X}(y|x)=\langle\phi_A^x|\Lambda_y|\phi_A^x\rangle$.

Remark 1: The setting of a random-parameter quantum channel $\mathcal{N}_{SA\to B}$ *without* side information and with $D=d_{\max}$ is equivalent to that of a channel that does not depend on a parameter, with $\mathcal{E}_{A\to B}=\sum_{s\in\mathcal{S}}q(s)\mathcal{N}^{(s)}_{A\to B}$ (see (6)). On the other hand, with side information at the encoder, this equivalence does not hold, as the channel input is correlated with the parameter sequence.

B. Regularization

From a practical perspective, the Holevo information formula in (15) is generally considered to be "easy to compute", given the channel statistics, since there are efficient algorithms to solve this convex optimization problem numerically, as *e.g.* in [63], up to a given precision and provided that the dimensions of the Hilbert spaces, $\mathcal{H}_A$ and $\mathcal{H}_B$, are not too large. Yet, in Shannon theory, it is generally desirable to establish a single-letter computable capacity formula [64].

	none	strictly-causal	causal	non-causal
Region	$\mathbb{C}(\mathcal{N})$	$\mathbb{C}_{\text{s-c}}(\mathcal{N})$	$\mathbb{C}_{\text{caus}}(\mathcal{N})$	$\mathbb{C}_{\text{n-c}}(\mathcal{N})$
Function	$C(\mathcal{N}, D)$	$C_{\text{s-c}}(\mathcal{N}, D)$	$C_{\text{caus}}(\mathcal{N}, D)$	$C_{\text{n-c}}(\mathcal{N}, D)$

Fig. 3. Notation of channel capacity-distortion regions and functions with and without CSI. The notation of the capacity-distortion regions is given in the first row, and of the capacity-distortion functions in the second row. The columns indicate the type of CSI that is available at the encoder.

Beyond computability, the disadvantage of a regularized multi-letter formula, of the form

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{I}(\mathcal{E}^{\otimes n}), \tag{19}$$

is that such characterization is not unique (see [27, Section 13.1.3]). Nonetheless, it should be emphasized that regularized characterizations are yet significant, since in many cases, the capacity can still be computed. Furthermore, there are interesting properties that can be derived even without a closed-form expression for the capacity [21], [44].

From a technical mathematical perspective, the difficulty in proving a single-letter converse part for a quantum channel is the hybrid nature of the Holevo mutual information $I(X;B)_\rho$, which involves a classical auxiliary variable X and a quantum system B (see Theorem 1). Specifically, consider a channel $\mathcal{E}_{A\to B}$ without a random parameter. From the familiar exercise of Fano's inequality and the chain rule, one obtains the bound

$$\begin{aligned} R - \varepsilon_n &\le \frac{1}{n}\sum_{i=1}^{n} I(M; B_i|B^{i-1})_\rho \\ &\le \frac{1}{n}\sum_{i=1}^{n} I(M, B^{i-1}; B_i)_\rho \end{aligned} \tag{20}$$

where ε_n tends to zero as $n \to \infty$ [22], [23]. In the attempt to establish an upper bound on the achievable rates in terms of the Holveo information $I(X;B)_\rho$, one is free to choose the auxiliary X in the converse proof, in principle. Yet, X needs to satisfy a certain Markov property, and more importantly in our discussion, X must be classical. Thereby, we cannot identify the auxiliary sequence X_i with (M, B^{i-1}). We note that this stands in contrast to the entanglement-assisted capacity formula [65] [44, Remark 5], where the auxiliary can be quantum. A deeper perspective is given in Subsection III-C below.

Remark 2: One may look at the regularization problem from a different angle. In the book by Nielsen and Chuang [66, Chapter 12], the single-letter Holevo information is associated with the *product-state capacity*. Specifically, the authors consider a simplified setting where the encoder is constrained such that the channel input must be a product state. This means that not only entanglement is prohibited, but classical correlation is not allowed either. In this remark, we propose a more general encoding constraint, which makes more sense for a practical system. For example, the model is suitable when the transmitter has access to multiple small or moderate-size quantum computers without interaction between them, where each computer has b qubits. In addition, in some qubit architectures, the physical limitations do not allow all qubits to "talk to each other". That is, one cannot apply a quantum gate to any pair of qubits, but only to qubits that are at certain proximity to each other. In order to account for those limitations, we impose the following encoding constraint. Assume that the encoder's quantum systems A^n are partitioned into sub-blocks of size b, such that the input state has the form

$$\rho_{A^n} = \rho_{A_1^b} \otimes \rho_{A_{b+1}^{2b}} \otimes \cdots \otimes \rho_{A_{(\ell-1)b+1}^{n}} \tag{21}$$

with $\ell \equiv \frac{n}{b}$. As usual, the capacity $C_b(\mathcal{E})$ under encoding constraint $b > 0$ is defined as the supremum of the achievable rates with codes that satisfy the constraint above. Following the lines of [66, Chapter 12], it can be shown that the capacity of a quantum channel $\mathcal{E}_{A\to B}$, without parameters, under encoding constraint $b > 0$, is given by

$$C_b(\mathcal{E}) = \frac{1}{b}\chi(\mathcal{E}^{\otimes b}) \tag{22}$$

where $\chi(\mathcal{E}) \equiv \max_{p_X, |\phi_A^x\rangle} I(X;B)_\rho$ is the Holevo information of the channel $\mathcal{E}_{A\to B}$. Observe that this capacity formula is computable, since $b > 0$ is assumed to be a small constant. One may think of the formula on the RHS of (22) as finite regularization. By taking the limit $b \to \infty$, we recover the HSW theorem without encoding constraints.

C. Additivity

Additivity is a central problem in the field of quantum Shannon theory [22]. An information measure $\mathsf{I}(\mathcal{E})$ is called additive if the information of a product of two channels is equal to the sum of the respective informations. That is, for every pair of channels $\mathcal{E}$ and $\mathcal{G}$,

$$\mathsf{I}(\mathcal{E} \otimes \mathcal{G}) = \mathsf{I}(\mathcal{E}) + \mathsf{I}(\mathcal{G}). \tag{23}$$

It is well-known that this property holds for the Shannon capacity formula of a classical channel $\mathcal{E}_{X\to Y}$. The merit of this property is that regularized capacity formulas reduce to a single-letter computable formula when the corresponding information measure is additive.

For more than a decade, it was believed by many researchers that the Holevo information $\chi(\mathcal{E})$, as defined in (15), is also additive and that entanglement between input states does not increase the classical capacity of a quantum channel [66, p. 554]. If the Holevo information of a channel is additive, then the regularization in the HSW characterization can be removed and the capacity can be expressed as $C(\mathcal{E}, d_{\max}) = \chi(\mathcal{E})$ (see Theorem 1). In fact, Fukuda and Wolf [67] established that n-fold additivity of the Holevo information is equivalent to its pairwise additivity. That is, when considering $\mathsf{I}(\mathcal{E}) = \chi(\mathcal{E})$, we have that $\mathsf{I}(\mathcal{E}^{\otimes n}) = n \cdot \mathsf{I}(\mathcal{E})$ holds for every quantum channel $\mathcal{E}_{A\to B}$ *if and only if* (23) holds for every pair of quantum channels $\mathcal{E}_{A_1\to B_1}$

and $\mathcal{G}_{A_2\to B_2}$. Nevertheless, the additivity conjecture has been refuted as Hastings [24] demonstrated strict super-additivity of quantum channels in 2009. That is, it was shown that there exist two channels $\mathcal{E}_{A_1\to B_1}$ and $\mathcal{G}_{A_2\to B_2}$ such that the Holevo informations satisfy $\chi(\mathcal{E}\otimes\mathcal{G}) > \chi(\mathcal{E})+\chi(\mathcal{G})$.

Remark 3: The super-additivity of the Holevo information implies that the Holevo information does not provide a full characterization of the capacity, *i.e.* $C(\mathcal{E}, d_{\max}) \neq \chi(\mathcal{E})$ in general. However, it does *not* imply that the operational capacity can be super-additive. As pointed out in [25, Section 8.4], it is an open problem whether there exist two channels $\mathcal{E}_{A_1\to B_1}$ and $\mathcal{G}_{A_2\to B_2}$ such that the operational capacity satisfies $C(\mathcal{E}\otimes\mathcal{G}, d_{\max}) > C(\mathcal{E}, d_{\max})+C(\mathcal{G}, d_{\max})$.

D. Entanglement-Breaking Channels

Given the HSW characterization, it is straightforward to obtain a single-letter formula for measurement channels and classical-quantum channels. Shor [26] considered the more general class of entanglement-breaking channels, which includes both measurement and classical-quantum channels. To obtain a single-letter characterization, Shor [26] has shown that the Holevo information of an entanglement-breaking channel is additive. On the other hand, we do not show additivity, but rather extend the methods of Wang *et al.* [58], and prove the converse part in a more direct manner. We note that Shor's approach in [26] has more insight than ours, as it characterizes the fundamental properties of an entanglement-breaking channel. Yet, we believe that the alternative argument is easier to extend to more complex models, including channel uncertainty.

First, we demonstrate this argument for a channel $\mathcal{E}_{A\to B}$ without parameters. Consider the bound in (20). As mentioned in Subsection II-B, if $\mathcal{E}_{A\to B}$ is an entanglement-breaking channel, then it can be presented as a concatenation of a measurement channel, followed by a state-preparation channel, i.e.

$$\mathcal{E}_{A\to B} = \mathcal{P}_{Y\to B}\circ\mathcal{M}_{A\to Y} \tag{24}$$

where Y is classical. Therefore, by the quantum data processing theorem due to Schumacher and Nielsen [68] [27, Theorem 11.9.4], $I(M, B^{i-1}; B_i)_\rho \leq I(M, Y^{i-1}; B_i)_\rho$. Since the sequence Y^{n-1} is classical, we can identify the auxiliary sequence as $X_i = (M, Y^{i-1})$, hence

$$R - \varepsilon_n \leq \frac{1}{n}\sum_{i=1}^{n} I(X_i; B_i)_\rho \tag{25}$$

which is bounded by the single-letter Holevo information of the channel.

In the sequel, we will use this argument to establish a single-letter characterization of the capacity-distortion region in the absence of CSI (see Subsection IV-D and Part 2 of Appendix F).

IV. Main Results

We state our results on the random-parameter quantum channel $\mathcal{N}_{SA\to B}$ with and without CSI at the encoder. The analysis is based on the information-theoretic tools that are presented in Appendix A.

A. Strictly-Causal Side Information

We begin with our main result on the random-parameter quantum channel with strictly-causal CSI. That is, at time i, Alice has access to the parameters of the *past*, S^{i-1}. Define the rate-distortion region

$$\mathcal{R}_{\text{s-c}}(\mathcal{N}) \triangleq \bigcup \left\{ (R,D) : \begin{array}{ll} R & \leq I(Z,X;B)_\rho - I(Z;S|X) \\ D & \geq \sum\limits_{s,\hat{s},x,z} q(s)p_X(x)p_{Z|X,S}(z|x,s) \\ & \quad \cdot \mathrm{Tr}(\Gamma^{\hat{s}}_{B|x,z}\rho_B^{s,x})d(s,\hat{s}) \end{array} \right\} \tag{26}$$

where the union is over the set of all distributions $p_X(x)p_{Z|X,S}(z|x,s)$, state collection $\{\theta_A^x\}$, and set of POVMs $\{\Gamma^{\hat{s}}_{B|x,z}\}$, with

$$\rho_B^{s,x} = \mathcal{N}^{(s)}_{A\to B}(\theta_A^x) \tag{27}$$

$$\rho_{SZXB} = \sum_{s\in\mathcal{S}}\sum_{z\in\mathcal{Z}}\sum_{x\in\mathcal{X}} q(s)p_X(x)p_{Z|X,S}(z|x,s)|s\rangle\langle s| \otimes |z\rangle\langle z|\otimes|x\rangle\langle x|\otimes\rho_B^{s,x}. \tag{28}$$

Before we state the capacity-distortion theorem, we give the following lemma. In principle, one may use the property below in order to compute the region $\mathcal{R}_{\text{s-c}}(\mathcal{N})$ for a given channel.

Lemma 2: The union in (26) can be restricted to pure states $\theta_A^x = |\phi_A^x\rangle\langle\phi_A^x|$, with $|\mathcal{X}| \leq |\mathcal{H}_A|^2+1$ and $|\mathcal{Z}| \leq |\mathcal{H}_A|^2+|\mathcal{S}|$.

The restriction to pure states follows by state purification, and the cardinality bounds are based on the Fenchel-Eggleston-Carathéodory lemma [69], using similar arguments as in [70]. The details are given in Appendix B.

Our main result is given below.

Theorem 3:

1) The capacity-distortion region of a random-parameter quantum channel $\mathcal{N}_{SA\to B}$ with strictly-causal CSI at the encoder is given by
$$\mathbb{C}_{\text{s-c}}(\mathcal{N}) = \bigcup_{k=1}^{\infty}\frac{1}{k}\mathcal{R}_{\text{s-c}}(\mathcal{N}^{\otimes k}). \tag{29}$$
2) For a random-parameter measurement channel $\mathcal{M}_{SA\to Y}$,
$$\mathbb{C}_{\text{s-c}}(\mathcal{M}) = \mathcal{R}_{\text{s-c}}(\mathcal{M}). \tag{30}$$

The proof of Theorem 3 is given in Appendix C. To prove achievability, we extend the classical block Markov coding to the quantum setting, and then apply the quantum packing lemma for decoding the message, and the classical covering lemma for the reconstruction of the parameter sequence. The gentle measurement lemma [60] alleviates the proof, as it guarantees that multiple decoding measurements can be performed without "destroying" the quantum state, *i.e.* such that the output state after each measurement is almost the same.

Remark 4: Observe that the bound on the rate in the definition of $\mathcal{R}_{\text{s-c}}(\mathcal{N})$ in (26) can also be written as

$$R \leq I(X;B)_\rho - [I(Z;S|X) - I(Z;B|X)_\rho], \tag{31}$$

by the mutual information chain rule. As the expression in the square brackets above is nonnegative, by the data processing inequality, it follows that the rate is bounded by the capacity of the channel without CSI, *i.e.* the Holevo information (see Subsection III-A). We will come back to this point when we consider the capacity when the receiver is not required to estimate the channel parameters in Subsection IV-E.

Remark 5: The expression in the square brackets in (31) can be interpreted as the penalty that the encoder pays for the transmission to contain a (partial) representation of the parameter sequence. Luo and Devetak [36] have considered the source compression setting with SSI. In their setting, a classical memoryless source S^n is compressed, and then reconstructed at the decoder with distortion D, using quantum side information systems B^n. Based on their results, the rate-distortion function $r(D,\rho_{SB})$ for this source compression setting is given by $r(D,\rho_{SB})=\lim_{k\to\infty}\frac{1}{k}\mathsf{r}(D,\rho_{SB}^{\otimes k})$, where

$$\mathsf{r}(D,\rho_{SB})= \min_{\substack{p_{\tilde{Z}|S}(z|s),\ \{\Gamma^{\hat{s}}_{B|z}\}:\\ \sum_{s,\hat{s},z} q(s)p_{\tilde{Z}|S}(z|s)\mathrm{Tr}(\Gamma^{\hat{s}}_{B|z}\rho^{s}_{B})d(s,\hat{s})\le D}} [I(Z';S)-I(Z';B)_\rho] \tag{32}$$

(see Theorem 4.2 in [36]). Thereby, in our setting, the encoder's penalty can be interpreted as the average compression rate of the parameter sequence with the channel output as the decoder's SSI. The more Z is correlated with the channel parameter S, the parameter estimation will be better, *i.e.* with a lower distortion. Yet, the penalty may be larger, resulting in a lower communication rate.

We illustrate our results with the following example. We will use the dephasing channel as a running example, and come back to it in the next sections as well.

Example 1: Consider a random-parameter dephasing channel that is specified by

$$\mathcal{N}^{(0)}(\rho)=\rho \tag{33}$$

$$\mathcal{N}^{(1)}(\rho)=\mathsf{Z}\rho\mathsf{Z} \tag{34}$$

with a binary random parameter $S\sim$ Bernoulli(ε), where $\varepsilon\in[0,1]$ is a given constant, *i.e.* $q(1)=1-q(0)=\varepsilon$. In other words, given a parameter sequence S^n, the parameter S_i acts as a switch that controls the phase flip operation at time i. Observe that without CSI, the decoder receives the average output of the standard dephasing channel, *i.e.* $\overline{\mathcal{N}}(\rho)\equiv(1-\varepsilon)\rho+\varepsilon\mathsf{Z}\rho\mathsf{Z}$. Given CSI at the encoder, however, the input state is correlated with the channel parameters. Furthermore, our decoder needs to recover the message and estimate whether there was a phase flip at each time. The natural measure for the distortion between the binary parameter sequence and its reconstruction is the following: $d(s,\hat{s})=s+\hat{s}\mod 2$. Namely, $d(s,\hat{s})=1$ if $\hat{s}\neq s$, and $d(s,\hat{s})=0$ if $\hat{s}=s$.

Clearly, if the encoder sends a constant transmission $|+\rangle\otimes\cdots\otimes|+\rangle$, then the decoder can determine whether there was a phase flip at each instance and recover the parameters without distortion. Yet, the rate is zero as well. On the other hand, by restricting the transmission to the computational basis, we can communicate without error, since the states $|0\rangle$ and $|1\rangle$ are unaffected by the phase flips. Thereby, if one is not interested in parameter estimation, the rate $R=1$ can achieved. Based on Theorem 3, we show that the following rate-distortion region is achievable for the random-parameter dephasing channel with strictly-causal CSI at the encoder,

$$\mathbb{C}_{\text{s-c}}(\mathcal{N})\supseteq \bigcup_{0\le\alpha\le\frac{1}{2}}\left\{(R,D):\begin{array}{ll}R & \le 1-[h(\alpha*\varepsilon)-h(\alpha)]\\ D & \ge\alpha\end{array}\right\}, \tag{35}$$

where $a*b=(1-a)b+a(1-b)$ denotes the binary convolution operation, and $h(x)=-x\log(x)-(1-x)\log(1-x)$ is the binary entropy function. Here, we see the tradeoff between the communication rate and the distortion. Taking $\alpha=\frac{1}{2}$, we obtain the maximal rate $R=1$, but the distortion $D=\frac{1}{2}$ is that of guessing by a coin flip. On the other hand, for $\alpha=0$, the channel parameters are recovered without distortion, while the communication rate is bounded by $R=1-h(\varepsilon)$. To obtain the achievable region above from Theorem 3, consider $k=1$. Set the distribution of the input ensemble as $X\sim$ Bernoulli$\left(\frac{1}{2}\right)$ over the state collection $\{|0\rangle,\ |1\rangle\}$. Define $Z=X+S+V\mod 2$ with $V\sim$ Bernoulli(α), such that V, X, and S are statistically independent. Given X and Z, the decoder estimates the channel parameter by $\hat{S}=X+Z\mod 2=S+V\mod 2$. This yields $I(X,Z;B)_\rho=H(B)_\rho-H(B|X,Z)_\rho=1-0=1$, $I(Z;S|X)=H(Z|X)-H(V)=h(\alpha*\varepsilon)-h(\alpha)$, and $\mathbb{E}d(S,\hat{S})=\Pr(V=1)=\alpha$.

Equivalently, we can characterize the capacity-distortion function.

Corollary 4:

1) The capacity-distortion function of a random-parameter quantum channel $\mathcal{N}_{SA\to B}$ with strictly-causal CSI at the encoder is given by

$$C_{\text{s-c}}(\mathcal{N},D)=\lim_{k\to\infty}\frac{1}{k}\max_{\substack{p_{X^k}(x^k)p_{Z^k|X^k,S^k}(z^k|x^k,s^k),\\ |\phi^{x^k}_{A^k}\rangle,\ \{\Gamma^{\hat{s}^k}_{B^k|x^k,z^k}\}:\ \mathbb{E}d^k(S^k,\hat{S}^k)\le D}} [I(Z^k,X^k;B^k)_\rho-I(Z^k;S^k|X^k)] \tag{36}$$

with

$$\begin{aligned}\rho_{S^kZ^kX^kB^k}=&\sum_{s^k,x^k,z^k}q^k(s^k)p_{X^k}(x^k)\\ &\cdot p_{Z^k|X^k,S^k}(z^k|x^k,s^k)|s^k\rangle\langle s^k|\\ &\otimes|z^k\rangle\langle z^k|\otimes|x^k\rangle\langle x^k|\\ &\otimes\mathcal{N}^{(s^k)}_{A^k\to B^k}(|\phi^{x^k}_{A^k}\rangle\langle\phi^{x^k}_{A^k}|).\end{aligned} \tag{37}$$

2) For a random-parameter measurement channel $\mathcal{M}_{SA\to Y}$,

$$C_{\text{s-c}}(\mathcal{M},D)=\max_{\substack{p_X(x)p_{Z|X,S}(z|x,s),\\ |\phi^x_A\rangle,\ \{\Gamma^{\hat{s}}_{Y|x,z}\}:\ \mathbb{E}d(S,\hat{S})\le D}} [I(Z,X;Y)-I(Z;S|X)] \tag{38}$$

with $p_{Y|X,S}(y|x,s)=\langle\phi^x_A|\Lambda^{(s)}_y|\phi^x_A\rangle$.

The corollary follows from Lemma 2 and Theorem 3. For example, following the derivation in Example 1, the capacity-distortion function of the random-parameter dephasing channel is bounded from below by

$$C_{\text{s-c}}(\mathcal{N}, D) \geq 1 - [h(D * \varepsilon) - h(D)] \tag{39}$$

for $0 \leq D \leq \frac{1}{2}$.

B. Causal Side Information

Next, we consider the random-parameter quantum channel with causal CSI, where Alice has access to the *past and present* random parameters, *i.e.* S^{i-1} and S_i. Define the rate-distortion region

$$\mathcal{R}_{\text{caus}}(\mathcal{N}, D) \triangleq \bigcup \left\{ \begin{array}{lll} (R, D) : & R & \leq I(Z, X; B)_\rho - I(Z; S|X) \\ & D & \geq \sum_{s,\hat{s},x,z} q(s) p_X(x) p_{Z|X,S}(z|x,s) \\ & & \cdot \text{Tr}(\Gamma^{\hat{s}}_{B|x,z} \rho_B^{s,x}) d(s, \hat{s}) \end{array} \right\} \tag{40}$$

where the union is over the set of all distributions $p_X(x)p_{Z|X,S}(z|x,s)$, states $\{\theta_G^x\}$, quantum channels $\mathcal{F}^{(s)}_{G\to A}$, and set of POVMs $\{\Gamma^{\hat{s}}_{B|x,z}\}$, with

$$\eta_A^{x,s} = \mathcal{F}^{(s)}_{G\to A}(\theta_G^x) \tag{41}$$

$$\rho_B^{s,x} = \mathcal{N}^{(s)}_{A\to B}(\eta_A^{s,x}) \tag{42}$$

$$\rho_{SZXB} = \sum_{s\in\mathcal{S}} \sum_{z\in\mathcal{Z}} \sum_{x\in\mathcal{X}} q(s) p_X(x) p_{Z|X,S}(z|x,s) |s\rangle\langle s| \otimes |z\rangle\langle z| \otimes |x\rangle\langle x| \otimes \rho_B^{s,x}. \tag{43}$$

The union in (26) can also be restricted to pure states $\theta_G^{z,x} = |\phi_G^{z,x}\rangle\langle\phi_G^{z,x}|$ based on the same arguments as in the proof of Lemma 2.

Observe that the difference between the characterizations with strictly-causal and causal CSI, in $\mathcal{R}_{\text{s-c}}(\mathcal{N})$ and $\mathcal{R}_{\text{caus}}(\mathcal{N})$, respectively, is that the channel input is θ_A^x in the former, and $\mathcal{F}^{(s)}_{G\to A}(\theta_G^x)$ in the latter (*cf.* (27) and (41)). That is, the input state here depends on the random parameter through the auxiliary channel $\mathcal{F}^{(s)}_{G\to A}$ in (41). Further interpretation and intuition for the role of this auxiliary channel will be given in Remark 6 and Examples 2 and 3. Now, we give our main result on the random-parameter quantum channel with causal CSI.

Theorem 5:

1) The capacity-distortion region of a random-parameter quantum channel $\mathcal{N}_{SA\to B}$ with causal CSI at the encoder is given by

$$\mathbb{C}_{\text{caus}}(\mathcal{N}) = \bigcup_{k=1}^{\infty} \frac{1}{k} \mathcal{R}_{\text{caus}}(\mathcal{N}^{\otimes k}). \tag{44}$$

2) For a random-parameter measurement channel $\mathcal{M}_{SA\to Y}$,

$$\mathbb{C}_{\text{caus}}(\mathcal{M}) = \mathcal{R}_{\text{caus}}(\mathcal{M}). \tag{45}$$

To prove achievability, we apply the coding techniques from the proof of Theorem 3 to the virtual channel $\mathcal{V}^{(s)}_{G\to B}$, defined by

$$\mathcal{V}^{(s)}_{G\to B}(\rho_G) = \mathcal{N}^{(s)}_{A\to B}\left(\mathcal{F}^{(s)}_{G\to A}(\rho_G)\right). \tag{46}$$

The proof outline for Theorem 5 is given in Appendix D.

Remark 6: Shannon [62] has shown that the capacity of a random-parameter classical channel $\mathcal{W}_{SX\to Y}$ with causal CSI at the encoder is given by

$$C_{\text{caus}}(\mathcal{W}, d_{\max}) = \max_{p_T} I(T; Y) \tag{47}$$

with $X = T(S)$, where $T : \mathcal{S} \to \mathcal{X}$ is an auxiliary function, which is commonly referred to as a *Shannon strategy*. The maximization in the formula above is over the distribution of the Shannon strategy. In Shannon's achievability scheme, the strategy maps a parameter value $S_i = s$ to a classical input $X_i = T(s)$ [71, Remark 7.6]. The auxiliary random-parameter channel $\mathcal{F}^{(s)}$ in (41) can be viewed as the quantum counterpart of the classical Shannon-strategy (see discussions in previous work by the author [34, Section IV.D] on further relations between Shannon strategies and quantum channels with causal CSI). The effect of the quantum Shannon strategy is demonstrated in the examples below.

In a similar manner as in the previous subsection, we can equivalently characterize the capacity-distortion function, $C_{\text{caus}}(\mathcal{N}, D)$. We omit this characterization to save space. Causal CSI may lead to a significant advantage compared to strictly-causal CSI, as demonstrated in the example below.

Example 2: Consider the random-parameter qubit dephasing channel from Example 1. Observe that knowing the current parameter S_i, at time i, the encoder can revert the dephasing using the following strategy: Perform $\mathcal{F}^{(0)}(\rho) = \mathcal{N}^{(0)}(\rho) = \rho$ and $\mathcal{F}^{(1)}(\rho) = \mathcal{N}^{(1)}(\rho) = \mathsf{Z}\rho\mathsf{Z}$. By Theorem 5, the capacity-distortion region of the random-parameter dephasing channel with causal CSI is given by

$$\mathbb{C}_{\text{caus}}(\mathcal{N}) = \left\{ \begin{array}{lll} (R, D) : & R & \leq 1 \\ & D & \geq 0 \end{array} \right\}. \tag{48}$$

The converse part is immediate, since the classical transmission rate over a qubit channel is always bounded by 1. To show the direct part using Theorem 5, consider $k = 1$. Set $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$ over the input ensemble $\{|\pm\rangle\}$, and $Z \equiv 0$. The decoder performs a measurement in the $\pm$-basis. Given X and a measurement outcome Y, choose $\hat{S} = Y + X \mod 2 = S$.

Example 3: Consider a random-parameter qubit depolarizing channel that is specified by

$$\mathcal{N}^{(0)}(\rho) = \rho \tag{49}$$

$$\mathcal{N}^{(1)}(\rho) = \mathsf{X}\rho\mathsf{X} \tag{50}$$

$$\mathcal{N}^{(2)}(\rho) = \mathsf{Y}\rho\mathsf{Y} \tag{51}$$

$$\mathcal{N}^{(3)}(\rho) = \mathsf{Z}\rho\mathsf{Z} \tag{52}$$

with the following parameter distribution,

$$q(0) = 1 - \varepsilon,\ q(1) = q(2) = q(3) = \frac{\varepsilon}{3} \tag{53}$$

where $\varepsilon \in [0, \frac{3}{8}]$ is a given constant. In other words, the parameter S_i chooses a Pauli operator that is applied to the ith

input system. We note that without CSI, the average channel is the same as the standard depolarizing channel, *i.e.*

$$\begin{aligned}\overline{\mathcal{N}}_{A\to B}(\rho) &\equiv \sum_s q(s)\mathcal{N}^{(s)}(\rho)\\ &= (1-\varepsilon)\rho + \frac{\varepsilon}{3}\left(\mathsf{X}\rho\mathsf{X}+\mathsf{Y}\rho\mathsf{Y}+\mathsf{Z}\rho\mathsf{Z}\right)\\ &= (1-p)\rho + p\pi \end{aligned} \tag{54}$$

where $\pi = \frac{\mathbb{1}}{2}$ is the maximally mixed state, and $p \equiv \frac{4\varepsilon}{3}$ is interpreted as the probability of depolarization (see [27, Section 4.7.4]). Here, the decoder needs to recover the message and estimate which Pauli operator was applied. For the distortion to be measured by the Hamming distance between the parameter sequence and its reconstruction, let $d(s,\hat{s}) = 1$ if $\hat{s} \neq s$, and $d(s,\hat{s}) = 0$ if $\hat{s} = s$.

Knowing the current parameter S_i, at time i, the encoder can revert the operation of the channel using the following strategy: Perform $\mathcal{F}^{(0)}(\rho) = \mathcal{N}^{(0)}(\rho) = \rho$, $\mathcal{F}^{(1)}(\rho) = \mathcal{N}^{(1)}(\rho) = \mathsf{X}\rho\mathsf{X}$, $\mathcal{F}^{(2)}(\rho) = \mathcal{N}^{(2)}(\rho) = \mathsf{Y}\rho\mathsf{Y}$, and $\mathcal{F}^{(3)}(\rho) = \mathcal{N}^{(3)}(\rho) = \mathsf{Z}\rho\mathsf{Z}$. Therefore, if one ignores the parameter estimation requirement, then the rate $R = 1$ can be achieved. By Theorem 5, the following region is achievable for the random-parameter depolarizing channel with causal CSI,

$$\mathbb{C}_{\text{caus}}(\mathcal{N}) \supseteq \bigcup_{0\le\alpha\le\varepsilon}\left\{ (R,D) : \begin{array}{rl} R & \le 1 - [H(1-\varepsilon,\frac{\varepsilon}{3},\frac{\varepsilon}{3},\frac{\varepsilon}{3}) \\ & \quad -H(1-\alpha,\frac{\alpha}{3},\frac{\alpha}{3},\frac{\alpha}{3})] \\ D & \ge \alpha \end{array} \right\}. \tag{55}$$

Once more, we see the tradeoff between the communication rate and the distortion. If we want the transmission to describe the parameter sequence without distortion, then this costs $H(S) = H(1-\varepsilon,\frac{\varepsilon}{3},\frac{\varepsilon}{3},\frac{\varepsilon}{3})$. Thereby, taking $\alpha = 0$, we achieve $R = 1 - H(1-\varepsilon,\frac{\varepsilon}{3},\frac{\varepsilon}{3},\frac{\varepsilon}{3})$ and $D = 0$. At the other extreme, taking $\alpha = \varepsilon$, we obtain the maximal rate $R = 1$, but the distortion $D = \varepsilon$ is that of ignorantly guessing '$0, 0, \ldots, 0$'.

To show achievability of the region above by Theorem 5, set the distribution of the input ensemble as $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$ over a qubit basis, and apply $\mathcal{F}^{(s)}$ to the basis vectors as specified above. Let $\widetilde{S} \sim (1-\alpha,\frac{\alpha}{3},\frac{\alpha}{3},\frac{\alpha}{3})$ and T be statistically independent random variables such that $S = \widetilde{S} + T \mod 4$, for some $0 \le \alpha \le \varepsilon$. Define $Z = X + S + \widetilde{S} \mod 4$. Given X and Z, the decoder chooses $\hat{S} = Z - X \mod 4 = S + \widetilde{S} \mod 4$. This yields $I(X,Z;B)_\rho = H(B)_\rho - H(B|X, S+\widetilde{S})_\rho = 1 - 0 = 1$, $I(Z;S|X) = H(S) - H(S|T) = H(S) - H(\widetilde{S}) = H(1-\varepsilon,\frac{\varepsilon}{3},\frac{\varepsilon}{3},\frac{\varepsilon}{3}) - H(1-\alpha,\frac{\alpha}{3},\frac{\alpha}{3},\frac{\alpha}{3})$, and $\mathbb{E}d(S,\hat{S}) = 1 - \Pr(Z - X - S \mod 4 = 0) = 1 - \Pr(\widetilde{S} = 0) = \alpha$.

C. Non-Causal Side Information

We consider the random-parameter quantum channel with non-causal CSI. Define the rate-distortion region

$$\mathcal{R}_{\text{n-c}}(\mathcal{N}) \triangleq \bigcup\left\{ (R,D) : \begin{array}{rl} R & \le I(X;B)_\rho - I(X;S) \\ D & \ge \sum_{s,\hat{s},x} q(s)p_{X|S}(x|s) \\ & \quad \cdot\text{Tr}(\Gamma^{\hat{s}}_{B|x}\rho^{s,x}_B)d(s,\hat{s}) \end{array} \right\} \tag{56}$$

where the union is over the set of all distributions $p_{X|S}(x|s)$, states $\{\theta^x_A\}$, and set of POVMs $\{\Gamma^{\hat{s}}_{B|x}\}$, with

$$\rho^{s,x}_B = \mathcal{N}^{(s)}_{A\to B}(\theta^x_A) \tag{57}$$

$$\rho_{SXB} = \sum_{s\in\mathcal{S}}\sum_{x\in\mathcal{X}} q(s)p_{X|S}(x|s)|s\rangle\langle s| \otimes |x\rangle\langle x| \otimes \rho^{s,x}_B. \tag{58}$$

We note that here, as opposed to the previous characterizations, the auxiliary variable X is allowed to depend on the random parameter S (*cf.* (28), (43), and (58)). Our main result on the random-parameter quantum channel with non-causal CSI is given below.

Theorem 6: The capacity-distortion region of the random-parameter quantum channel $\mathcal{N}_{SA\to B}$ with non-causal CSI at the encoder is given by

$$\mathbb{C}_{\text{n-c}}(\mathcal{N}) = \bigcup_{k=1}^{\infty}\frac{1}{k}\mathcal{R}_{\text{n-c}}(\mathcal{N}^{\otimes k}). \tag{59}$$

The proof of Theorem 6 is given in Appendix E. To prove achievability, we use an extension of the classical binning technique [72] to the quantum setting, and then apply the quantum packing lemma and the classical covering lemma. We note that even for a classical channel, a single-letter characterization of the optimal region with non-causal CSI is an open problem [14]. As in Subsection IV-A, we can write an equivalent characterization in terms of the capacity-distortion function, $C_{\text{n-c}}(\mathcal{N}, D)$. We omit this to save space. We illustrate the results with a simple example below. Theorem 6 is also the basis for the bosonic dirty-paper analysis in Section V.

Example 4: Consider the following random-parameter qubit channel,

$$\mathcal{N}^{(0)}(\rho) = \rho \tag{60}$$

$$\mathcal{N}^{(1)}(\rho) = |\psi\rangle\langle\psi| \tag{61}$$

with $S \sim \text{Bernoulli}(\varepsilon)$, a Hamming distortion function as in the previous examples, and a given state $|\psi\rangle$, in the same qubit space. In other words, the parameter S_i chooses whether the ith input system is projected onto $|\psi\rangle$. Ignoring the CSI at the encoder, the model resembles the quantum erasure channel [73] (see also [27, Section 20.4.3]), except that the "erasure state" is orthogonal to the qubit space, while $|\psi\rangle$ in the present example is in the same qubit space. Nonetheless, we note that if the decoder knows the locations where the state is projected, then this model is equivalent to the quantum erasure channel. Without this knowledge at the decoder, it is less obvious.

By Theorem 6, the following region is achievable for the random-parameter channel above with non-causal CSI,

$$\mathbb{C}_{\text{n-c}}(\mathcal{N}) \supseteq \bigcup_{0\le\alpha\le\frac{1}{2}}\left\{ (R,D) : \begin{array}{rl} R & \le (1-\varepsilon)h(\alpha) \\ D & \ge (1-\varepsilon)\alpha \end{array} \right\}. \tag{62}$$

Again, we can see the tradeoff between the communication rate and the distortion. Let $|\psi_\perp\rangle$ be an orthogonal state with respect to $|\psi\rangle$. Clearly, if the encoder transmits $|\psi\rangle$ when $S_i = 1$, and $|\psi_\perp\rangle$ when $S_i = 0$, then the decoder

can recover the parameters without distortion, by performing a measurement in the corresponding basis, *i.e.* $\{|\psi\rangle, |\psi_\perp\rangle\}$. Indeed, for $\alpha = 0$, we achieve $(R, D) = (0, 0)$. On the other hand, taking $\alpha = \frac{1}{2}$, we obtain the maximal rate $R = 1 - \varepsilon$, which is also the capacity of the quantum erasure channel.

To show this, note that the bound on the rate on the RHS of (56) can also be expressed as

$$R \leq H(X|S) - H(X|B)_\rho. \tag{63}$$

Given non-causal CSI at the encoder, we can choose an auxiliary X that depends on the channel parameter S. Let the input ensemble be the basis $\{|\psi\rangle, |\psi_\perp\rangle\}$. The input distribution is chosen as follows. Let $V \sim$ Bernoulli(α) be statistically independent of S. If $S = 0$, set $X = V+1 \mod 2$. Otherwise, if $S = 1$, then $X = 0$. As for the decoder, given X, set $\hat{S} = X+1 \mod 2$. This yields $H(X|B)_\rho = 0$ and $H(X|S) = (1-\varepsilon)H(V) = (1-\varepsilon)h(\alpha)$, and $\mathbb{E}d(S, \hat{S}) = \Pr(\hat{S} \neq S) = \Pr(S = 0, V = 1) = (1-\varepsilon)\alpha$.

D. In the Absence of Side Information

Consider the case where Alice does not have access to the parameter sequence, yet Bob is required to estimate the sequence with limited distortion. Given our previous analysis, the proof of a regularized formula in this case is straightforward. However, here we obtain a single letter formula not just for measurement channels, but for the whole class of entanglement-breaking channels. As opposed to Shor [26], we do not show additivity (see Subsection III-C). Instead, we prove the converse part in a more direct manner using the observations that we have presented in Subsection III-D, which extend the methods by Wang *et al.* [58].

We give our capacity-distortion theorem for the random-parameter quantum channel without CSI. Define

$$\mathcal{R}(\mathcal{N}) \triangleq \bigcup \left\{ (R, D) : \begin{array}{ll} R & \leq I(X;B)_\rho \\ D & \geq \sum\limits_{s,\hat{s},x} q(s)p_X(x) \\ & \cdot \mathrm{Tr}(\Gamma^{\hat{s}}_{B|x}\rho^{s,x}_B)d(s,\hat{s}) \end{array} \right\} \tag{64}$$

where the union is over the set of all distributions $p_X(x)$, states $\{\theta^x_A\}$, and set of POVMs $\{\Gamma^{\hat{s}}_{B|x}\}$, with

$$\rho^{s,x}_B = \mathcal{N}^{(s)}_{A\to B}(\theta^x_A) \tag{65}$$

$$\rho_{SXB} = \sum_{s\in\mathcal{S}}\sum_{x\in\mathcal{X}} q(s)p_X(x)|s\rangle\langle s| \otimes |x\rangle\langle x| \otimes \rho^{s,x}_B. \tag{66}$$

We note that the union in (56) can be restricted to pure states $\theta^x_A = |\phi^x_A\rangle\langle\phi^x_A|$ with $|\mathcal{X}| \leq |\mathcal{H}_A|^2 + 1$, based on the same arguments as in the proof of Lemma 2.

Theorem 7:

1) The capacity-distortion region of a random-parameter quantum channel $\mathcal{N}_{SA\to B}$ without CSI is given by

$$\mathbb{C}(\mathcal{N}) = \bigcup_{k=1}^{\infty} \frac{1}{k}\mathcal{R}(\mathcal{N}^{\otimes k}). \tag{67}$$

2) If $\mathcal{N}_{SA\to B}$ is entanglement-breaking, then

$$\mathbb{C}(\mathcal{N}) = \mathcal{R}(\mathcal{N}). \tag{68}$$

The proof of Theorem 7 is given in Appendix F. The proof of the first part follows by similar arguments as for the previous results, while the proof of the second part is based on our observations in Subsection III-D. The characterization of the capacity-distortion function $C(\mathcal{N}, D)$ follows as before.

We revisit our running examples of dephasing and depolarizing channels.

Example 5: Consider the random-parameter qubit dephasing channel from Examples 1 and 2. By Theorem 7, the capacity-distortion region of the random-parameter dephasing channel without CSI is given by

$$\mathbb{C}(\mathcal{N}) = \left\{ (R, D) : \begin{array}{ll} R & \leq 1 - h(\varepsilon) \\ D & \geq 0 \end{array} \right\}. \tag{69}$$

To show achievability, set the distribution of the input ensemble and measure the parameter estimate as before. Observe that the region above is the same as in Example 1. That is, for this channel, the capacity-distortion region with strictly-causal CSI is the same as without CSI.

Example 6: Consider the random-parameter depolarizing channel from Example 3. The capacity-distortion region of the random-parameter depolarizing channel without CSI is bounded by

$$\mathbb{C}(\mathcal{N}) \supseteq \left\{ (R, D) : \begin{array}{ll} R & \leq 1 - h(\frac{2\varepsilon}{3}) \\ D & \geq \frac{2\varepsilon}{3} \end{array} \right\}. \tag{70}$$

To derive this achievable region, set the distribution of the input ensemble as $X \sim$ Bernoulli $\left(\frac{1}{2}\right)$ over the state collection $\{|+\rangle, \ |-\rangle\}$. The channel output is then the same as that of a dephasing channel, as in the previous example. As the dephasing corresponds to the Pauli operators Y and Z, the "dephasing probability" is $\varepsilon_0 \equiv \frac{2\varepsilon}{3}$. The decoder performs a measurement in the $\pm$-basis. Denote the measurement outcome by Y. If $X \neq Y$, then the decoder knows that the channel operator was either Y or Z. Then, the decoder chooses $\hat{S}$ to be either 2 or 3 with equal probability. If $X = Y$, then the decoder knows that the channel operator is not a dephasing one, *i.e.* either $\mathbb{1}$ or X, and chooses $\hat{S} = 0$. Thus, the average distortion is $\mathbb{E}d(S, \hat{S}) = (1-\varepsilon)\cdot 0 + \frac{\varepsilon}{3}\cdot 1 + \frac{2\varepsilon}{3}\cdot\frac{1}{2} = \frac{2\varepsilon}{3}$.

E. Without Parameter Estimation

We obtain the following results as direct consequences of Corollary 4 and Theorems 5-6. As mentioned, the standard definition of the capacity, *i.e.* when parameter estimation is not required at the decoder, is equivalent to the capacity-distortion function for $D = d_{\max}$. Henceforth, we use the term 'the capacity of $\mathcal{N}_{SA\to B}$ without CSI' referring to $C(\mathcal{N}, d_{max})$. Similarly, $C_{\text{s-c}}(\mathcal{N}, d_{max})$, $C_{\text{caus}}(\mathcal{N}, d_{max})$, and $C_{\text{n-c}}(\mathcal{N}, d_{max})$ are the capacities with strictly-causal, causal, and non-causal CSI, respectively (see Figure 3).

The next corollaries generalize the results of Boche *et al.* [1] on classical-quantum channels with CSI.

Corollary 8: The capacity of a random-parameter quantum channel $\mathcal{N}_{SA\to B}$ with strictly-causal CSI at the encoder is the same as without CSI, *i.e.* $C_{\text{s-c}}(\mathcal{N}, d_{max}) = C(\mathcal{N}, d_{max}) = \chi(\mathcal{N})$.

The direct part is immediate, since the encoder can simply ignore the CSI, while the converse part follows from Remark 4 and the HSW capacity Theorem, Theorem 1, where CSI is not available.

Next, we consider causal CSI.

Corollary 9:

1) The capacity of a random-parameter quantum channel $\mathcal{N}_{SA\to B}$ with causal CSI at the encoder is given by

$$C_{\text{caus}}(\mathcal{N}, d_{\max}) = \lim_{k\to\infty} \frac{1}{k} \sup_{\substack{p_{X^k}(x^k),\\ |\phi^{x^k}_{G^k}\rangle,\, \mathcal{F}^{(s^k)}_{G^k\to A^k}}} I(X^k; B^k)_\rho \tag{71}$$

with

$$\rho_{S^kX^kB^k} = \sum_{s^k,x^k} q^k(s^k) p_{X^k}(x^k) |s^k\rangle\langle s^k| \otimes |x^k\rangle\langle x^k| \otimes \mathcal{N}^{(s^k)}_{A^k\to B^k}\left(\mathcal{F}^{(s^k)}_{G^k\to A^k}(|\phi^{x^k}_{G^k}\rangle\langle\phi^{x^k}_{G^k}|)\right). \tag{72}$$

2) For a random-parameter measurement channel $\mathcal{M}_{SA\to Y}$,

$$C_{\text{caus}}(\mathcal{M}, d_{\max}) = \sup_{p_X(x),\, |\phi^x_G\rangle,\, \mathcal{F}^{(s)}_{G\to A}} I(X;Y) \tag{73}$$

with $p_{Y|X,Z,S}(y|x,z,s) = \operatorname{Tr}\left(\Lambda_y \mathcal{F}^{(s)}(|\phi^x_G\rangle\langle\phi^x_G|)\right)$.

The direct part follows by taking $Z = \emptyset$, and the converse part holds by the argument that we made in Remark 4 for strictly-causal CSI. In particular, for the depolarizing channel in Example 3, Corollary 8 and Corollary 9 imply $C_{\text{s-c}}(\mathcal{N}, d_{\max}) = 1 - h\left(\frac{p}{2}\right) = 1 - h\left(\frac{2\varepsilon}{3}\right)$ and $C_{\text{caus}}(\mathcal{N}, d_{\max}) = 1$. For the random-parameter quantum channel with non-causal CSI, we recover the result in [30, Corollary 1, part (c)].

Corollary 10 (See Also [30]): The capacity of a random-parameter quantum channel $\mathcal{N}_{SA\to B}$ with non-causal CSI at the encoder is given by

$$C_{\text{n-c}}(\mathcal{N}, d_{\max}) = \lim_{k\to\infty} \frac{1}{k} \sup_{p_{X^k|S^k}(x^k|s^k),\, \theta^{x^k,s^k}_{A^k}} [I(X^k; B^k)_\rho - I(X^k; S^k)] \tag{74}$$

with

$$\rho_{S^kX^kB^k} = \sum_{s^k,x^k} q^k(s^k) p_{X^k|S^k}(x^k|s^k) |s^k\rangle\langle s^k| \otimes |x^k\rangle\langle x^k| \otimes \mathcal{N}^{(s^k)}_{A^k\to B^k}\left(\theta^{x^k,s^k}_{A^k}\right). \tag{75}$$

The statement above immediately follows from Theorem 6 as the distortion constraint is inactive for $D = d_{\max}$. We will use the last corollary in the bosonic dirty-paper analysis in the next section.

V. Bosonic Dirty Paper Coding

In this section, we address the special case of a single-mode bosonic channel with classical interference in the modulation and with non-causal side information at the transmitter, without parameter estimation at the receiver, *i.e.* $D = d_{\max}$. In the analysis, we will use our result in Corollary 10.

A. Introduction

Consider a *classical* channel $W_{Y|X,S}$ with random parameters. Given non-causal CSI at the encoder, the channel is known as the Gel'fand-Pinsker model [72]. The capacity of this channel is given by [6]

$$C_{\text{n-c}}(W, d_{\max}) = \max_{p_{U,X|S}} [I(U;Y) - I(U;S)] \tag{76}$$

where U is an auxiliary random variable such that $U \multimap (X,S) \multimap Y$ form a Markov chain. The characterization above can also be obtained from Corollary 10.

A random-parameter Gaussian channel is specified by the input-output relation $Y = X + Z + S$, with a real-valued Gaussian noise $Z \sim \mathcal{N}_{\mathbb{R}}(0, \sigma^2_Z)$, an additive interference S known to the transmitter, and an input power constraint P. A well-known result by Costa [74] is that the capacity of the random-parameter Gaussian channel is the same as if the interference is not there, i.e. $C(W) = \frac{1}{2}\log\left(1 + \frac{P}{\sigma^2_Z}\right)$. Given that S^n is *not* known to the receiver, it is far from obvious that the interference can be canceled out without sacrificing transmission power. The trivial strategy is to send $X = U - S$, such that U represents the transmitted information and is uncorrelated with S, resulting in an intereference-free output, $Y = U + Z$. However, in the Gaussian case, the power constraint must be accounted for. Thereby, the trivial strategy above wastes transmission power and can only achieve a rate of $R \le \frac{1}{2}\log\left(1 + \frac{\max(P-\sigma^2_S, 0)}{\sigma^2_Z}\right)$, which is sub-optimal.

The derivation of the capacity of the random-parameter Gaussian channel with non-causal CSI requires a more gentle approach, and it is based on Costa's dirty-paper coding strategy [74]: Set

$$U = X + tS \tag{77}$$

such that X is statistically *independent* of S. The optimal choice of the coefficient t turns out to be the same as that of the minimum mean-square error (MMSE) estimator $\widehat{X} = t(X+Z)$ for X given the noisy observation $(X+Z)$ (see [2, Section 4.1]), namely,

$$t = \frac{P}{P + \sigma^2_Z}. \tag{78}$$

Explicit code constructions based on lattice codes were proposed in [75], [76] and references therein. Furthermore, efficient algorithms for practical implementation were presented, based on state-of-the art polar codes [77]–[79], LDPC codes [80]–[82], and so on.

The bosonic channel is a simple quantum-mechanical model for optical communication over free space or optical fibers [53], [54], and it can be viewed as the quantum counterpart of the classical channel with additive white Gaussian noise (AWGN). An optical communication system consists of a modulated source of photons, the optical channel, and an optical detector. For a single-mode bosonic channel, the channel input is an electromagnetic field mode with an annihilation operator $\hat{a}$, and the output is another mode with the annihilation operator $\hat{b}$. The input-output relation in the Heisenberg picture [55] is given by

$$\hat{b} = \sqrt{\eta}\,\hat{a} + \sqrt{1-\eta}\,\hat{e} \tag{79}$$

where $\hat{e}$ is associated with the environment noise and the parameter η is the transmissivity, $0 \leq \eta \leq 1$, which depends on the length of the optical fiber and its absorption length [56] (see Figure 1). For a lossy bosonic channel, the noise mode $\hat{e}$ is in a Gibbs thermal state $\tau(N_E)$ which consists of a mixture of coherent states, where

$$\tau(N) \equiv \int_{\mathbb{C}} d^2\alpha \frac{e^{-|\alpha|^2}}{\pi N} |\alpha\rangle\langle\alpha| \tag{80}$$

given an average photon number $N \geq 0$. Modulation is performed such that the unitary displacement operator $D(\alpha) = \exp(\alpha\hat{a}^\dagger - \alpha^*\hat{a})$ is applied to the vacuum state $|0\rangle\langle 0|$ [53].

B. Model and Results

We consider the single-mode lossy bosonic channel with a coherent-state protocol and a non-ideal displacement operation in the modulation process:

$$|\zeta\rangle = D(\alpha + s)|0\rangle = |\alpha + s\rangle. \tag{81}$$

This model can viewed as the quantum counterpart of the classical random-parameter Gaussian channel. Based on Costa's writing-on-dirty-paper result [74], the effect of the channel parameter can be canceled even when the decoder has no side information, and regardless of the input power constraint. For both homodyne and heterodyne detection with a coherent-state protocol, the model reduces to a classical channel with either real or complex-valued Gaussian noise. Thereby, by applying Costa's dirty-paper coding strategy, we observe that the effect of the classical interference can be canceled for those channels as well. Then, we consider the bosonic channel with joint detection, for which the classical results do not apply, and derive a dirty-paper coding lower bound. Furthermore, considering the special case of a pure-loss bosonic channel, we demonstrate that the optimal coefficient for dirty paper coding is not necessarily the MMSE estimator coefficient as in the classical setting. We denote the random-parameter bosonic channel by $\mathcal{B}$. To simplify the notation, we use the short notation $C_{\text{n-c}}(\mathcal{B}) \equiv C_{\text{n-c}}(\mathcal{B}, d_{\max})$ for the capacity of the random-parameter bosonic channel, without parameter estimation.

We begin with homodyne and heterodyne detection. A homodyne measurement of a quadrature observable is implemented in practice by combining the target quantum mode with an intense local oscillator at a 50:50 beam splitter, and measuring the photocurrent difference of the outgoing modes using two photodetectors [83]. When homodyne detection is used with a coherent-state protocol, the resulting channel $\mathcal{B}_{\text{hom}}$ is the random-parameter classical Gaussian channel

$$Y = \sqrt{\eta}(\alpha + S) + Z_{\text{hom}} \tag{82}$$

with a real-valued Gaussian parameter $S \sim \mathcal{N}_{\mathbb{R}}(0, N_S)$ and noise $Z_{\text{hom}} \sim \mathcal{N}_{\mathbb{R}}\left(0, \frac{1}{4}\left[2(1-\eta)N_E + 1\right]\right)$ [84]. Using the dirty-paper coding scheme, we take $\alpha \sim \mathcal{N}_{\mathbb{R}}(0, N_A)$ and $U = \alpha + t_0 S$ with $t_0 = \frac{N_A}{N_A + N_E}$, such that α and S are uncorrelated. The effect of the interference is thus removed, and the capacity is given by

$$C_{\text{n-c}}(\mathcal{B}_{\text{hom}}) = \frac{1}{2}\log\left(1 + \frac{4\eta N_A}{2(1-\eta)N_E + 1}\right) \tag{83}$$

as without interference.

In heterodyne detection, two quadratures are measured by combining the measured mode with a vacuum mode into a 50:50 beam splitter, and homodyning the quadratures of the outcome modes [53]. Heterodyne detection is described by a random-parameter channel $\mathcal{B}_{\text{het}}$ with complex-valued Gaussian noise, specified by

$$Y = \sqrt{\eta}(\alpha + S) + Z_{\text{het}} \tag{84}$$

with complex-valued circularly-symmetric Gaussian random parameter $S \sim \mathcal{N}_{\mathbb{C}}(0, \frac{1}{2}N_S)$ and noise $Z_{\text{het}} \sim \mathcal{N}_{\mathbb{C}}(0, \frac{1}{2}[(1-\eta)N_E + 1])$ [84]. Similarly, we use dirty-paper coding with $\alpha \sim \mathcal{N}_{\mathbb{C}}(0, \frac{1}{2}N_A)$ and $U = \alpha + t_0 S$, achieving the capacity

$$C_{\text{n-c}}(\mathcal{B}_{\text{het}}) = \log\left(1 + \frac{\eta N_A}{(1-\eta)N_E + 1}\right) \tag{85}$$

as without interference.

At last, we consider the case where the decoder can perform an arbitrary quantum measurement on the output systems $B_1, \ldots, B_n$ together. For joint detection [55], the channel does not have a classical description. Based on Corollary 10, the capacity of a random-parameter quantum channel $\mathcal{B}$ with CSI at the transmitter, is given by the regularized formula $C_{\text{n-c}}(\mathcal{B}) = \lim_{n\to\infty} \frac{1}{n}\mathsf{C}_{\text{n-c}}(\mathcal{B}^{\otimes n})$, with

$$\mathsf{C}_{\text{n-c}}(\mathcal{B}) = \sup_{p_{X|S},\, \theta_A^{x,s}} [I(X;B)_\rho - I(X;S)]. \tag{86}$$

Previously, we have assumed that the space dimensions are finite. Yet, this result is now extended to the bosonic channel with infinite-dimension Hilbert spaces following the discretization limiting argument by Guha *et al.* [85]. Using the dirty-paper coding strategy, we obtain the lower bound $C_{\text{n-c}}(\mathcal{B}_{\text{joint}}) \geq \mathsf{R}_{\text{DPC}}(t)$,

$$\begin{aligned}\mathsf{R}_{\text{DPC}}(t) &\equiv I(\gamma;B) - I(\gamma;S)\Big|_{\gamma=\alpha+tS} \\ &= g(\eta(N_A + N_S) + (1-\eta)N_E) \\ &\quad - g\left(\frac{\eta(1-t)^2 N_A N_S}{N_A + t^2 N_S} + (1-\eta)N_E\right) \\ &\quad - \log\left(\frac{N_A + t^2 N_S}{N_A}\right)\end{aligned} \tag{87}$$

where the subscript 'DPC' stands for 'dirty-paper coding', and $g(N)$ is the von Neumann entropy of the thermal state $\tau(N)$,

$$g(N) = \begin{cases} (N+1)\log(N+1) - N\log(N) & N > 0, \\ 0 & N = 0. \end{cases} \tag{88}$$

The second equality in (87) holds since the channel input is associated with $\zeta \equiv \alpha + S = \gamma + (1-t)S$ (see (81)), and the conditional variance of the channel parameter S given γ is

$$\text{var}(S|\gamma) = \left[1 - \frac{(\text{cov}(\gamma, S))^2}{\text{var}(S)\text{var}(\gamma)}\right]\text{var}(S) = \frac{N_A N_S}{N_A + t^2 N_S}.$$

In particular, consider the special case of a pure-loss bosonic channel, where $N_E = 0$. In this case,

$$\begin{aligned}\mathsf{R}_{\text{DPC}}(t) &= g(\eta(N_A + N_S)) - g\left(\frac{\eta(1-t)^2 N_A N_S}{N_A + t^2 N_S}\right) \\ &\quad - \log\left(\frac{N_A + t^2 N_S}{N_A}\right).\end{aligned} \tag{89}$$

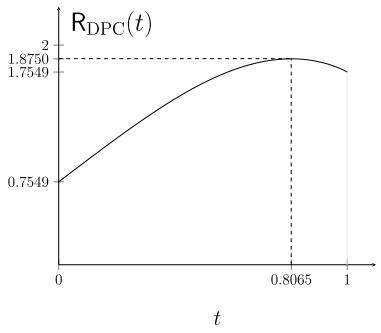

Fig. 4. The dirty-paper coding lower bound for the pure-loss bosonic channel with joint detection and a coherent-state protocol.

To demonstrate, suppose that $N_A = N_S = 2$ and $\eta = \frac{1}{2}$. Then, we have

$$\mathsf{R}_{\mathrm{DPC}}(t) = g(2) - g\left(\frac{(1-t)^2}{1+t^2}\right) - \log(1+t^2). \tag{90}$$

Ignoring the CSI, we obtain a rate $\mathsf{R}_{\mathrm{DPC}}(t=0) = g(2) - g(1) = 3\log 3 - 4 = 0.7549$. Whereas, using the dirty-paper coding scheme with the MMSE coefficient $t_0 = \frac{2}{2+0} = 1$, we obtain a better rate: $\mathsf{R}_{\mathrm{DPC}}(t=1) = 3\log 3 - 2 - \log 2 = 1.7549$. The optimal value for dirty-paper coding turns out to be $t_{\max} = 0.8065$, for which

$$\mathsf{R}_{\mathrm{DPC}}(t_{\max}) = 1.8750. \tag{91}$$

See Figure 4. The rate above is higher than the homodyne and heterodyne-detection capacities, $C(\mathcal{E}_{\mathrm{hom}}) = 0.5849$ and $C(\mathcal{E}_{\mathrm{het}}) = 1$, respectively. However, this rate is lower than the joint-detection capacity without interference ($N_S = 0$), which is given by $g(1) = 2$.

Our results can be further extended to other optical channels. In particular, the random-parameter thermal amplifier channel $\mathcal{A}$ with an amplification gain $\kappa > 1$ has the input-output relation $\hat{b} = \sqrt{\kappa}\,\hat{a}(s) + \sqrt{\kappa - 1}\,\hat{e}^{\dagger}$ [55], [86]. In a similar manner, we obtain the dirty-paper coding lower bound

$$\begin{aligned} C(\mathcal{A}) \geq \max_{t\in[0,1]} \Bigg[& g(\kappa(N_A+N_S)+(\kappa-1)N_E) \\ & - g\left(\frac{\kappa(1-t)^2 N_A N_S}{N_A + t^2 N_S} + (\kappa-1)N_E\right) \\ & - \log\left(\frac{N_A + t^2 N_S}{N_A}\right)\Bigg]. \end{aligned} \tag{92}$$

C. Concluding Remarks

We conclude with the following remarks on the comparison between the classical and quantum dirty-paper settings:

1) Costa [74] provided the intuitive analogy of 'writing on dirty paper'. When a writer (Alice) is given a dirty paper, she knows the location and intensity of the dirt spots before writing. On the other hand, the reader (Bob) sees a mixture of the written text (channel input) and the dirt (channel parameter) without prior knowledge. In our setting, the dirt is the interference s_i in the modulation displacement $D(\alpha_i + s_i)$. Alternatively, in the quantum watermarking scheme that we have described above, the dirt is the host covertext.
2) The classical capacity result can be derived using the dirty-paper coding strategy in (77)-(78) following the observation that $U - tY = X - t(X+Z)$ is the error of the MMSE estimation of X given $V = X + Z$, hence it is statistically independent of the observation. Thereby, $(U - tY)$ is jointly independent of (V, S). This, in turn, implies that $(U - tY)$ and $Y = V + S$ are statistically independent, leading to $H(U|Y) = H(U - tY) = H(X|V)$ which can be used in order to show that
$$I(U;Y) - I(U;S) = H(U|S) - H(U|Y) = I(X;V) \tag{93}$$
(see further details in [2] [71, Section 7.7]). For a bosonic channel with joint detection, we can also write the capacity in terms of $H(X|S) - H(X|B)_{\rho}$, with conditioning on the channel output. However, conditioning on a quantum system does not necessarily carry the meaning of an observation as in the classical setting [87].
3) While dirty-paper coding was originally introduced to treat a channel with random parameters [74], the technique is useful in multi-user setups of wireless communications as well, such as the multiple-input multiple-output (MIMO) broadcast fading channel [88]. It is only natural to apply and extend our results to multi-mode bosonic networks.

VI. Summary and Discussion

We have considered a quantum channel $\mathcal{N}^{(s)}_{A\to B}$ that depends on a classical random parameter $S \sim q(s)$, when the decoder is required to reconstruct the parameter sequence in a lossy manner, *i.e.* with limited distortion. This model can be viewed as the quantum analog of the classical rate-and-state (RnS) channel.

We consider two applications for this model: digital multicast using *quantum* communication channels, and classical watermarking with a quantum embedding. The first application is digital multicast, where the message represents digital control information that is multicast on top of an existing analog transmission, which is also estimated by the receiver. In the watermarking application, an authentication message m is mixed within classical host data S^n ("stegotext"), and this mixture is encoded into a quantum state that is sent to an authenticator. The estimation of the channel parameters at the decoder corresponds to a scenario where the host data S^n itself contains desirable information. Our model can also be interpreted as a form of quantum metrology [57], where the decoder performs measurements on the received (quantum) systems in order to estimate classical noise parameters, while exploiting the entanglement generated by the encoder.

The scenarios that we studied in the present work include either strictly-causal, causal, or non-causal channel side information (CSI) available at the encoder, as well as the case where CSI is not available. With strictly-causal CSI, Alice knows, the *past* random parameters S^{i-1}; given causal CSI, she knows the *past and present* parameters S^i; with non-causal CSI, the entire sequence S^n is available to her a priori; and without CSI, Alice is ignorant. In all of those cases, Bob is unaware of the random parameters, and he

has two tasks to perform. He is required to decode the message and to reconstruct the parameter sequence S^n with a limited distortion, D. We derived regularized formulas for the capacity-distortion tradeoff regions. In the special case of measurement channels, single-letter characterizations were established for the strictly-causal and causal settings. Furthermore, in the more general case of entanglement-breaking channels, a single-letter characterization was derived when CSI is not available. We also demonstrated the results in multiple examples, such as random-parameter dephasing channels and depolarizing channels.

While reviewing previous work in Section III, we reviewed single-letterization and regularization, additivity, and entanglement-breaking channels; and we compared between Shor's original approach for single-letterization, based on additivity, and the alternative argument that follows from [58]. Later, we used this alternative argument in the analysis for our setting. In particular, considering entanglement-breaking channels without CSI, we used a different approach from that of Shor [26]. As opposed to Shor [26], we did not show additivity of the capacity formula, but rather extended the methods of Wang *et al.* [58] to prove the converse part in a more direct manner. This more direct approach has less insight compared to Shor's additivity argument, and yet, we believe that it can be easier to extend to complex settings, as with parameter estimation at the decoder.

To prove achievability with strictly-causal CSI, we extended the classical block Markov coding method from [3] to the quantum setting, and then applied the quantum packing lemma for decoding the message, and the classical covering lemma for the reconstruction of the parameter sequence. The gentle measurement lemma alleviates the proof, as it guarantees that multiple decoding measurements can be performed without collapsing the quantum state and such that the output state after each measurement is almost the same. Thus, we can separate between measurements for recovering the message and for sequence reconstruction. Achievability with causal CSI was proved using similar techniques with the addition of a quantum "Shannon-strategy" encoding operation. To prove achievability with non-causal CSI, we used an extension of the classical binning technique [6] to the quantum setting.

Furthermore, we introduced bosonic dirty-paper coding. We considered the single-mode lossy bosonic channel with a coherent-state protocol and a non-ideal displacement operation in the modulation process. The channel parameters in our model represent classical interference in the transmission equipment, which the transmitter becomes aware of, while the receiver is not. Alternatively, this can be viewed as a watermarking model with a quantum embedding. Given a classical host data sequence $s_1, \ldots, s_n$, Alice encodes an authentication message m into a watermark $(\alpha_i(m, s_1, \ldots, s_n))_{i=1}^n$. Next, Alice performs a quantum embedding of the watermark; she prepares a *watermarked state* $|\zeta_1\zeta_2\cdots\zeta_n\rangle$ where $\zeta_i \equiv D(\alpha_i + s_i)|0\rangle = |\alpha_i + s_i\rangle$, and transmits it to the authenticator Bob through the optical fiber. The capacity of the random-parameter bosonic channel represents the optimal rate at which the authenticator can recover the messages with high fidelity.

First, we considered homodyne and heterodyne detection. Both of those settings reduce to a classical random-parameter channel with either real or complex-valued Gaussian noise. Thereby, we observed that based on Costa's dirty-paper solution, the effect of the classical interference can be canceled, and the capacity is the same regardless of the intensity of the interference. Then, we considered joint detection, in which case, the problem does not reduce to that of a classical description. We derived a dirty-paper coding lower bound based on the results above, using an auxiliary $\gamma = \alpha + tS$ with a general coefficient $t \in [0, 1]$, such that $\alpha \sim \mathcal{N}_{\mathbb{C}}(0, \frac{N_A}{2})$ is statistically independent of the channel parameter S. Considering the special case of a pure-loss bosonic channel, we showed that the optimal coefficient is not necessarily the MMSE value $t_0 = \frac{N_A}{N_A+N_E}$.

As a consequence of our main results, we obtained regularized formulas for the capacity of random-parameter quantum channels with strictly-causal, causal, or non-causal CSI, generalizing the previous results by Boche *et al.* [1] on classical-quantum channels. We believe that this could open the door to extend other important classical side-information models to quantum communication.

APPENDIX A
INFORMATION THEORETIC TOOLS

To derive our results, we use the quantum version of the method of types properties and techniques. The basic definitions and lemmas that are used in this paper are given below.

A. Classical Types

The type of a classical sequence x^n is defined as the empirical distribution $\hat{P}_{x^n}(a) = N(a|x^n)/n$ for $a \in \mathcal{X}$, where $N(a|x^n)$ is the number of occurrences of the symbol a in the sequence x^n. The set of all types over $\mathcal{X}$ is then denoted by $\mathcal{P}_n(\mathcal{X})$. The type class associated with a type $\hat{P} \in \mathcal{P}_n(\mathcal{X})$ is defined as the set of sequences of that type, *i.e.*

$$\mathcal{T}(\hat{P}) \equiv \left\{ x^n \in \mathcal{X}^n \,:\, \hat{P}_{x^n} = \hat{P} \right\}. \tag{94}$$

For a pair of sequences x^n and y^n, we give similar definitions in terms of the joint type $\hat{P}_{x^n,y^n}(a,b) = N(a,b|x^n,y^n)/n$ for $a \in \mathcal{X}$, $b \in \mathcal{Y}$, where $N(a,b|x^n,y^n)$ is the number of occurrences of the symbol pair (a,b) in the sequence $(x_i,y_i)_{i=1}^n$. Given a sequence $y^n \in \mathcal{Y}^n$, we further define the conditional type $\hat{P}_{x^n|y^n}(a|b) = N(a,b|x^n,y^n)/N(b|y^n)$ and the conditional type class

$$\mathcal{T}(\hat{P}|y^n) \equiv \left\{ x^n \in \mathcal{X}^n \,:\, \hat{P}_{x^n,y^n}(a,b) = \hat{P}_{y^n}(b)\hat{P}(a|b) \right\}. \tag{95}$$

Given a probability distribution $p_X \in \mathcal{P}(\mathcal{X})$, the δ-typical set is defined as

$$\mathcal{A}^\delta(p_X) \equiv \Big\{ x^n \in \mathcal{X}^n \,:\, \left| \hat{P}_{x^n}(a) - p_X(a) \right| \le \delta \;\text{ if }\; p_X(a) > 0, \;\; \hat{P}_{x^n}(a) = 0 \;\text{ if }\; p_X(a) = 0, \;\forall\, a \in \mathcal{X} \Big\} \tag{96}$$

The covering lemma is a powerful tool in classical information theory [89].

Lemma 11 (Classical Covering Lemma [89] [71, Lemma 3.3]): Let $X^n \sim \prod_{i=1}^n p_X(x_i)$, $\delta > 0$, and let $Z^n(m)$, $m \in [1 : 2^{nR}]$, be conditionally independent random sequences distributed according to $\prod_{i=1}^n p_Z(z_i)$. Suppose that the sequence X^n is pairwise independent of the sequences $Z^n(m)$, $m \in [1 : 2^{nR}]$. Then,

$$\Pr\left((Z^n(m), X^n) \notin \mathcal{A}^\delta(p_{Z,X}) \text{ for all } m \in [1 : 2^{nR}]\right) \leq \exp\left(-2^{n(R-I(Z;X)-\varepsilon_n(\delta)}\right) \quad (97)$$

where $\varepsilon_n(\delta)$ tends to zero as $n \to \infty$ and $\delta \to 0$.

Let $X^n \sim \prod_{i=1}^n p_X(x_i)$ be an information source sequence, encoded by an index m at compression rate R. Based on the covering lemma above, as long as the compression rate is higher than $I(Z;X)$, a set of random codewords, $Z^n(m) \sim \prod_{i=1}^n p_Z(z_i)$, contains with high probability at least one sequence that is jointly typical with the source sequence.

Though originally stated in the context of lossy source coding, the classical covering lemma is useful in a variety of scenarios [71], including the random-parameter channel with non-causal CSI. In this case, the parameter sequence $S^n \sim \prod_{i=1}^n q(s_i)$ plays the role of the "source sequence".

B. Quantum Typical Subspaces

Moving to the quantum method of types, suppose that the state of a system is generated from an ensemble $\{p_X(x), |x\rangle\}_{x\in\mathcal{X}}$, hence, the average density operator is

$$\rho = \sum_{x\in\mathcal{X}} p_X(x)|x\rangle\langle x|. \quad (98)$$

Consider the subspace spanned by the vectors $|x^n\rangle$, $x^n \in \mathcal{T}(\hat{P})$, for a given type $\hat{P} \in \mathcal{P}_n(\mathcal{X})$. Then, the projector onto the subspace is given by

$$\Pi_{A^n}(\hat{P}) \equiv \sum_{x^n\in\mathcal{T}(\hat{P})} |x^n\rangle\langle x^n|. \quad (99)$$

Note that the dimension of the subspace of type class $\hat{P}$ is given by $\mathrm{Tr}(\Pi_{A^n}(\hat{P})) = |\mathcal{T}(\hat{P})|$. By classical type properties [89, Lemma 2.3] (see also [27, Property 15.3.2]),

$$(n+1)^{|\mathcal{X}|}2^{nH(\rho)} \leq \mathrm{Tr}(\Pi_{A^n}(\hat{P})) \leq 2^{nH(\rho)}. \quad (100)$$

The projector onto the δ-typical subspace is defined as

$$\Pi^\delta(\rho) \equiv \sum_{x^n\in\mathcal{A}^\delta(p_X)} |x^n\rangle\langle x^n|. \quad (101)$$

Based on [90] [66, Theorem 12.5], for every $\varepsilon, \delta > 0$ and sufficiently large n, the δ-typical projector satisfies

$$\mathrm{Tr}(\Pi^\delta(\rho)\rho^{\otimes n}) \geq 1-\varepsilon, \quad (102)$$

$$2^{-n(H(\rho)+c\delta)}\Pi^\delta(\rho) \preceq \Pi^\delta(\rho)\,\rho^{\otimes n}\,\Pi^\delta(\rho) \preceq 2^{-n(H(\rho)-c\delta)}\Pi^\delta(\rho), \quad (103)$$

$$\mathrm{Tr}(\Pi^\delta(\rho)) \leq 2^{n(H(\rho)+c\delta)} \quad (104)$$

where $c > 0$ is a constant.

We will also need the conditional δ-typical subspace. Consider a state

$$\sigma = \sum_{x\in\mathcal{Y}} p_X(x)\rho_B^x \quad (105)$$

with

$$\rho_B^x = \sum_{y\in\mathcal{Y}} p_{Y|X}(y|x)|\psi^{x,y}\rangle\langle\psi^{x,y}|. \quad (106)$$

Given a fixed sequence $x^n \in \mathcal{X}^n$, divide the index set $[1:n]$ into the subsets $I_n(a) = \{i : x_i = a\}$, $a \in \mathcal{X}$, and define the conditional δ-typical subspace $\mathscr{S}^\delta(\sigma_B|x^n)$ as the span of the vectors $|\psi^{x^n,y^n}\rangle = \otimes_{i=1}^n |\psi^{x_i,y_i}\rangle$ such that

$$y^{I_n(a)} \in \mathcal{A}_\delta^{(|I_n(a)|)}(p_{Y|X=a}), \text{ for } a \in \mathcal{X}. \quad (107)$$

The projector onto the conditional δ-typical subspace is defined as

$$\Pi^\delta(\sigma_B|x^n) \equiv \sum_{|\psi^{x^n,y^n}\rangle\in\mathscr{S}^\delta(\sigma_B|x^n)} |\psi^{x^n,y^n}\rangle\langle\psi^{x^n,y^n}|. \quad (108)$$

Based on [90] [27, Section 15.2.4], for every $\varepsilon', \delta > 0$ and sufficiently large n,

$$\mathrm{Tr}(\Pi^\delta(\sigma_B|x^n)\rho_{B^n}^{x^n}) \geq 1-\varepsilon', \quad (109)$$

$$2^{-n(H(B|X')_\sigma+c'\delta)}\Pi^\delta(\sigma_B|x^n) \preceq \Pi^\delta(\sigma_B|x^n)\,\rho_{B^n}^{x^n}\,\Pi^\delta(\sigma_B|x^n) \preceq 2^{-n(H(B|X')_\sigma-c'\delta)}\Pi^\delta(\sigma_B|x^n), \quad (110)$$

$$\mathrm{Tr}(\Pi^\delta(\sigma_B|x^n)) \leq 2^{n(H(B|X')_\sigma+c'\delta)} \quad (111)$$

where $c' > 0$ is a constant, $\rho_{B^n}^{x^n} = \bigotimes_{i=1}^n \rho_{B_i}^{x_i}$, and the classical random variable X' is distributed according to the type of x^n. Furthermore, if $x^n \in \mathcal{A}^\delta(p_X)$, then

$$\mathrm{Tr}(\Pi^\delta(\sigma_B)\rho_{B^n}^{x^n}) \geq 1-\varepsilon'. \quad (112)$$

(see [27, Property 15.2.7]). We note that the conditional entropy in the bounds above can also be expressed as

$$H(B|X')_\sigma = \frac{1}{n}H(B^n|X^n = x^n)_\sigma \equiv \frac{1}{n}H(B^n)_{\rho^{x^n}}. \quad (113)$$

C. Quantum Packing Lemma

To prove achievability for the HSW Theorem (see Theorem 1), one may invoke the quantum packing lemma [27], [59]. Suppose that Alice employs a codebook that consists of 2^{nR} codewords $x^n(m)$, $m \in [1 : 2^{nR}]$, by which she chooses a quantum state from an ensemble $\{\rho_{x^n}\}_{x^n\in\mathcal{X}^n}$. The proof is based on random codebook generation, where the codewords are drawn at random according to an input distribution $p_X(x)$. To recover the transmitted message, Bob may perform the square-root measurement [22], [23] using a code projector Π and codeword projectors Π_{x^n}, $x^n \in \mathcal{X}^n$, which project onto subspaces of the Hilbert space $\mathcal{H}_{B^n}$.

The lemma below is a simplified, less general, version of the quantum packing lemma by Hsieh, Devetak, and Winter [59].

Lemma 12 (Quantum Packing Lemma [59, Lemma 2]): Let

$$\rho = \sum_{x\in\mathcal{X}} p_X(x)\rho_x \quad (114)$$

where $\{p_X(x), \rho_x\}_{x\in\mathcal{X}}$ is a given ensemble. Furthermore, suppose that there is a code projector Π and codeword projectors Π_{x^n}, $x^n \in \mathcal{A}^\delta(p_X)$, that satisfy for every $\alpha > 0$ and sufficiently large n,

$$\mathrm{Tr}(\Pi\rho_{x^n}) \geq 1-\alpha \tag{115}$$
$$\mathrm{Tr}(\Pi_{x^n}\rho_{x^n}) \geq 1-\alpha \tag{116}$$
$$\mathrm{Tr}(\Pi_{x^n}) \leq 2^{ne_0} \tag{117}$$
$$\Pi\rho^{\otimes n}\Pi \preceq 2^{-n(E_0-\alpha)}\Pi \tag{118}$$

for some $0 < e_0 < E_0$ with $\rho_{x^n} \equiv \bigotimes_{i=1}^n \rho_{x_i}$. Then, there exist codewords $x^n(m)$, $m \in [1 : 2^{nR}]$, and a POVM $\{\Lambda_m\}_{m\in[1:2^{nR}]}$ such that

$$\mathrm{Tr}\left(\Lambda_m\rho_{x^n(m)}\right) \geq 1 - 2^{-n[E_0-e_0-R-\varepsilon_n(\alpha)]} \tag{119}$$

for all $m \in [1 : 2^{nR}]$, where $\varepsilon_n(\alpha)$ tends to zero as $n \to \infty$ and $\alpha \to 0$.

In our analysis, where there is CSI at the encoder, we apply the packing lemma such that the quantum ensemble encodes both the message m and a compressed representation of the parameter sequence s^n.

D. Gentle Measurement

The gentle measurement lemma is a useful tool. As will be seen, it guarantees that we can perform multiple measurements such that the state of the system remains almost the same after each measurement.

Lemma 13 (See [60], [61]): Let ρ be a density operator. Suppose that Λ is a meaurement operator such that $0 \preceq \Lambda \preceq \mathbb{1}$. If

$$\mathrm{Tr}(\Lambda\rho) \geq 1-\varepsilon \tag{120}$$

for some $0 \leq \varepsilon \leq 1$, then the post-measurement state $\rho' \equiv \frac{\sqrt{\Lambda}\rho\sqrt{\Lambda}}{\mathrm{Tr}(\Lambda\rho)}$ is $2\sqrt{\varepsilon}$-close to the original state in trace distance, *i.e.*

$$\|\rho-\rho'\|_1 \leq 2\sqrt{\varepsilon}. \tag{121}$$

The lemma is particularly useful in our analysis since the POVM operators in the quantum packing lemma satisfy the conditions of the lemma for large n (see (119)).

APPENDIX B
PROOF OF LEMMA 2

Consider the region $\mathcal{R}_{\text{s-c}}(\mathcal{N})$ as defined in (26).

A. Purification

To prove that a union over pure states is sufficient, we show that for every rate R_0 that can be achieved with distortion D, there exists a rate $R_1 \geq R_0$ that can be achieved with pure states and the same distortion. Fix $p_X(x)p_{Z|X,S}(z|x,s)$, $\{\theta_A^x\}$, and $\{\Gamma^{\hat{s}}_{B|x,z}\}$. Let

$$R_0 = I(X,Z;B)_\rho - I(Z;S|X) \tag{122}$$
$$D_0 = \sum_{s\in\mathcal{S}}\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}}\sum_{\hat{s}\in\widehat{S}} q(s)p_X(x)p_{Z|X,S}(z|x,s) \cdot \mathrm{Tr}(\Gamma^{\hat{s}}_{B|x,z}\rho_B^{s,x})d(s,\hat{s}) \tag{123}$$

and consider the spectral decomposition,

$$\theta_A^x = \sum_{w\in\mathcal{W}} p_{W|X}(w|x)\phi_A^{x,w} \tag{124}$$

where $P_{W|X}(w|x)$ is a conditional probability distribution, and $\phi_A^{x,w}$ are pure. Consider the extended state

$$\rho_{SXZWA} = \sum_{s,x,z,w} q(s)p_X(x)p_{Z|X,S}(z|x,s) \cdot p_{W|X}(w|x)|s\rangle\langle s| \otimes |x\rangle\langle x| \otimes |z\rangle\langle z| \otimes |w\rangle\langle w| \otimes \phi_A^{x,w}. \tag{125}$$

Now, observe that the union in the RHS of (26) includes the rate-distortion pair (R_1, D_1) that is given by

$$R_1 = I(X,W,Z;B)_\rho - I(Z;S|X,W) \tag{126}$$
$$D_1 = \sum_{s,x,z,w,\hat{s}} q(s)p_X(x)p_{Z|X,S}(z|x,s)p_{W|X}(w|x) \cdot \mathrm{Tr}\left(\Gamma^{\hat{s}}_{B|x,z}\mathcal{N}^{(s)}(\phi_A^{x,w})\right) d(s,\hat{s}) \tag{127}$$

which is obtained by plugging $X' = (X,W)$ instead of X, and the pure states $\phi_A^{x,w}$ instead of θ_A^x. That is, $(R_1, D_1) \in \mathcal{R}_{\text{s-c}}(\mathcal{N})$. According to (125), the random variables $S \multimap (X,Z) \multimap W$ form a Markov chain, thus $I(W;S|X,Z) = 0$. By the chain rule, it follows that $I(Z;S|X,W) = I(Z;S|X) + I(W;S|X,Z) - I(W;S|X) = I(Z;S|X)$, hence $R_1 = I(X,W,Z;B)_\rho - I(Z;S|X) \geq I(X,Z;B)_\rho - I(Z;S|X) = R_0$. As for the distortion level, we have by linearity that

$$D_1 = \sum_{s,x,z,\hat{s}} q(s)p_X(x)p_{Z|X,S}(z|x,s) \cdot \mathrm{Tr}\left(\Gamma^{\hat{s}}_{B|x,z}\mathcal{N}^{(s)}\left(\sum_w p_{W|X}(w|x)\phi_A^{x,w}\right)\right) d(s,\hat{s}) = D_0 \tag{128}$$

where the last equality is due to (123) and (124). Thereby, the union can be restricted to pure states.

B. Cardinality Bounds

To bound the alphabet size of the random variables X and Z, we use the Fenchel-Eggleston-Carathéodory lemma [69] and similar arguments as in [70]. Let

$$L_0 = |\mathcal{H}_A|^2 + 1 \tag{129}$$
$$L_1 = |\mathcal{H}_A|^2 + |\mathcal{S}|. \tag{130}$$

First, fix $q(s)$ and $p_{Z|X,S}(z|x,s)$, and consider the ensemble $\{p_X(x)p_{Z|X,S}(z|x,s), \theta_A^x\}$. An Hermitian matrix can be specified by $|\mathcal{H}_A|$ real values for the diagonal and $\frac{1}{2}|\mathcal{H}_A|(|\mathcal{H}|_A-1)$ complex numbers for the non-diagonal entries, or, $|\mathcal{H}_A|^2$ real parameters in total. Since a density matrix is Hermitian and also has a unit trace, every quantum state θ_A has a unique parametric representation $u(\theta_A)$ of dimension $|\mathcal{H}_A|^2-1$. Then, define a map $f_0 : \mathcal{X} \to \mathbb{R}^{L_0}$ by

$$f_0(x) = \Big(u(\theta_A^x),\ -H(B|X=x,Z)_\rho + H(S|X=x,Z), \mathbb{E}[d(S,\hat{S})|X=x]\Big). \tag{131}$$

The map f_0 can be extended to probability distributions as follows,

$$F_0 \,:\, p_X \mapsto \sum_{x\in\mathcal{X}} p_X(x) f_0(x) = \Big(u(\theta_A),\ -H(B|X,Z)_\rho + H(S|X,Z),\ \mathbb{E}d(S,\hat{S})\Big) \quad (132)$$

where $\theta_A = \sum_x p_X(x)\theta_A^x$. According to the Fenchel-Eggleston-Carathéodory lemma [69], any point in the convex closure of a connected compact set within $\mathbb{R}^d$ belongs to the convex hull of d points in the set. Since the map F_0 is linear, it maps the set of distributions on $\mathcal{X}$ to a connected compact set in $\mathbb{R}^{L_0}$. Thus, for every p_X, there exists a probability distribution $p_{\bar{X}}$ on a subset $\overline{\mathcal{X}} \subseteq \mathcal{X}$ of size L_0, such that $F_0(p_{\bar{X}}) = F_0(p_X)$. We deduce that alphabet size can be restricted to $|\mathcal{X}| \leq L_0$, while preserving θ_A and $\rho_B \equiv \sum_s q(s)\mathcal{N}^{(s)}(\theta_A)$; $I(X,Z;B)_\rho - I(Z;S|X) = H(B)_\rho - H(B|X,Z)_\rho + H(S|X,Z) - H(S)$; and $\mathbb{E}d(S,\hat{S})$.

We move to the alphabet size of Z. Fix $p_{X,S|Z}$, where

$$p_{X,S|Z}(x,s|z) \equiv \frac{q(s)p_X(x)p_{Z|X,S}(z|x,s)}{\sum_{s'\in\mathcal{S}} q(s') \sum_{x'\in\mathcal{S}} p_X(x')p_{Z|X,S}(z|x',s')}. \quad (133)$$

Define the map $f_1 : \mathcal{Z} \to \mathbb{R}^{L_1}$ by

$$f_1(z) = \Big(p_{S|Z}(\cdot|z),\ -H(B|X,Z=z)_\rho + H(S|X,Z=z),\ \mathbb{E}[d(S,\hat{S})|Z=z]\Big). \quad (134)$$

Now, the extended map is

$$F_1 \,:\, p_Z \mapsto \sum_{z\in\mathcal{Z}} p_Z(z) f_1(z) = \Big(q(\cdot),\ -H(B|X,Z)_\rho + H(S|X,Z),\ \mathbb{E}d(S,\hat{S})\Big). \quad (135)$$

By the Fenchel-Eggleston-Carathéodory lemma [69], for every p_Z, there exists $p_{\bar{Z}}$ on a subset $\overline{\mathcal{Z}} \subseteq \mathcal{Z}$ of size L_1, such that $F_1(p_{\bar{Z}}) = F_1(p_Z)$. We deduce that alphabet size can be restricted to $|\mathcal{Z}| \leq L_1$, while preserving $q(s)$, ρ_B, $I(X,Z;B)_\rho - I(Z;S|X)$, and $\mathbb{E}d(S,\hat{S})$. □

APPENDIX C
PROOF OF THEOREM 3

Consider a random-parameter quantum channel $\mathcal{N}_{SA\to B}$ with strictly-causal CSI.

Part 1:

A. Achievability Proof

We show that for every $\zeta_0, \varepsilon_0, \delta_0 > 0$, there exists a $(2^{n(R-\zeta_0)}, n, \varepsilon_0, D+\delta_0)$ code for $\mathcal{N}_{SA\to B}$ with strictly-causal CSI, provided that $(R,D) \in \mathcal{R}_{\text{s-c}}(\mathcal{N})$. To prove achievability, we extend the classical block Markov coding to the quantum setting, and then apply the quantum packing lemma and the classical covering lemma. We use the gentle measurement lemma [60], which guarantees that multiple decoding measurements can be performed without "destroying" the output state.

Recall that with strictly-causal CSI, the encoder has access to the sequence of *past* parameters $s_1, s_2, \ldots, s_{i-1}$. Let $\{p_X(x)p_{Z|X,S}(z|x,s), \theta_A^x\}$ be a given ensemble, and fix a set of POVMs $\{\Gamma^{\hat{s}}_{B|x,z}\}$ such that

$$\sum_{s,\hat{s},x,z} q(s)p_X(x)p_{Z|X,S}(z|x,s)\mathrm{Tr}(\Gamma^{\hat{s}}_{B|x,z}\mathcal{N}^{(s)}(\theta_A^x))d(s,\hat{s}) \leq D. \quad (136)$$

Define the average states

$$\rho_B^x = \sum_{s\in\mathcal{S}} q(s)\mathcal{N}^{(s)}(\theta_A^x), \text{ for } x \in \mathcal{X}, \quad (137)$$

$$\rho_B = \sum_{x\in\mathcal{X}} p_X(x)\rho_B^x \quad (138)$$

We will also consider the a posteriori probability distribution, conditioning on $Z=z$:

$$\hat{p}_{X,S|Z}(x,s|z) = \frac{q(s)p_X(x)p_{Z|X,S}(z|x,s)}{\sum_{x'\in\mathcal{X}}\sum_{s'\in\mathcal{S}} q(s')p_X(x')p_{Z|X,S}(z|x',s')}. \quad (139)$$

Then, the corresponding output state is

$$\sigma_B^{x,z} = \sum_{s\in\mathcal{S}} \hat{p}_{S|Z,X}(s|z,x)\mathcal{N}^{(s)}(\theta_A^x). \quad (140)$$

We use T transmission blocks, where each block consists of n input systems. In particular, with strictly-causal CSI, the encoder has access to the parameter sequences from the previous blocks. In effect, the j^{th} transmission block encodes a message $m_j \in [1 : 2^{nR}]$ and a compression of the parameter sequence s_{j-1}^n from the previous block, for $j \in [2:T]$.

The code construction, encoding and decoding procedures are described below.

1) Classical Code Construction: Let $\delta > 0$, $R_s > 0$, and $\widetilde{R}_s > 0$ such that $R_s < \widetilde{R}_s$. For every $j \in [2:T]$, select $2^{n(R+R_s)}$ independent sequences $x_j^n(m_j, \ell_{j-1})$, $m_j \in [1 : 2^{nR}]$, $\ell_{j-1} \in [1 : 2^{nR_s}]$, at random according to $\prod_{i=1}^n p_X(x_{j,i})$. For every $m_j \in [1 : 2^{nR}]$ and $\ell_{j-1} \in [1 : 2^{nR_s}]$, select $2^{n\widetilde{R}_s}$ conditionally independent sequences $z_j^n(k_j|m_j,\ell_{j-1})$, $k_j \in [1 : 2^{n\widetilde{R}_s}]$, at random according to $\prod_{i=1}^n p_{Z|X}(z_{j,i}|x_{j,i}(m_j,\ell_{j-1}))$. For $j=1$, set $\ell_0 \equiv 1$, and select $x_1^n(m_1,1)$ and $z_1^n(k_1|m_1,1)$ in the same manner, for $(m_1,k_1) \in [1:2^{nR}] \times [1:2^{n\widetilde{R}_s}]$. We have thus defined the classical codebooks

$$\mathscr{B}(j) = \{(x_j^n(m_j,\ell_{j-1}), z_j^n(k_j|m_j,\ell_{j-1}))\},\ j \in [1:T] \quad (141)$$

with $m_j \in [1:2^{nR}]$, $\ell_{j-1} \in [1:2^{nR_s}]$, $k_j \in [1:2^{n\widetilde{R}_s}]$. Partition the set of indices $[1:2^{n\widetilde{R}_s}]$ into bins $\mathcal{K}(\ell_j) = [(\ell_j-1)2^{n(\widetilde{R}_s-R_s)}+1 : \ell_j 2^{n(\widetilde{R}_s-R_s)}]$ of equal size $2^{n(\widetilde{R}_s-R_s)}$.

2) Encoding and Decoding: To send the messages (m_j), given the parameter sequences $(s_1^n,\ldots,s_{j-1}^n)$, Alice performs the following.

(i) At the end of block j, find an index $k_j \in [1:2^{n\widetilde{R}_s}]$ such that $(s_j^n, z_j^n(k_j|m_j,\ell_{j-1}), x_j^n(m_j,\ell_{j-1})) \in \mathcal{A}^\delta(p_{S,X,Z})$, where $p_{S,X,Z}(s,x,z) = q(s)p_X(x)p_{Z|X,S}(z|x,s)$. If there is none, select k_j arbitrarily, and if there is more

than one such index, choose the smallest. Set ℓ_j to be the bin index of k_j, *i.e.* such that $k_j \in \mathcal{K}(\ell_j)$.

(ii) In block $j+1$, prepare $\rho_{A_{j+1}^n} = \bigotimes_{i=1}^n \theta_A^{x_{j+1,i}(m_{j+1},\ell_j)}$ and send the block A_{j+1}^n.

Bob receives the systems $B_1^n, \ldots, B_T^n$ in the state

$$\rho_{B^{Tn}} = \bigotimes_{j=1}^{T} \bigotimes_{i=1}^{n} \rho_B^{x_{j+1,i}(m_{j+1},\ell_j)} \tag{142}$$

and decodes as follows.

(i) At the end of block $j+1$, decode $(\hat{m}_{j+1}, \hat{\ell}_j)$ by applying a POVM $\{\Lambda^1_{m_{j+1},\ell_j}\}_{(m_{j+1},\ell_j)\in[1:2^{nR}]\times[1:2^{nR_s}]}$, which will be specified later, to the systems B_{j+1}^n, for $j = 0, 1, \ldots, T-1$.

(ii) Decode $\hat{k}_j$ by applying a second POVM $\{\Lambda^2_{k_j|x^n(\hat{m}_{j+1},\hat{\ell}_j)}\}_{k_j\in\mathcal{K}(\hat{\ell}_j)}$, which will also be specified later, to the systems B_j^n.

(iii) Reconstruct the parameter sequence by applying the POVM $\Gamma^{\hat{s}_{j,i}}_{B|x_{j,i},z_{j,i}}$ to the system $B_{j,i}$ with $x_{j,i} \equiv x_{j,i}(\hat{m}_j, \hat{\ell}_{j-1})$ and $z_{j,i} \equiv z_{j,i}(\hat{k}_j|\hat{m}_j, \hat{\ell}_{j-1})$, for $j \in [1:T]$ and $i \in [1:n]$.

3) Analysis of Probability of Error and Distortion: By symmetry, we may assume without loss of generality that Alice sends the message $M_j = 1$ using $L_j = L_{j-1} = 1$, for $j \in [1:T]$. Consider the following events,

$$\mathscr{E}_1(j) = \{(S^n, X^n(1,1), Z^n(k_j|1,1)) \notin \mathcal{A}^{\delta_1}(p_{S,X,Z}), \text{ for all } k_j \in [1:2^{n\widetilde{R}_s}]\} \tag{143}$$

$$\mathscr{E}_2(j) = \{(\hat{M}_j, \hat{L}_{j-1}) \neq (1,1)\} \tag{144}$$

$$\mathscr{E}_3(j) = \{\hat{K}_j \neq K_j\} \tag{145}$$

$$\mathscr{E}_4(j) = \{d^n(S_j^n, \hat{S}_j^n) > D + \frac{1}{2}\delta_0\} \tag{146}$$

with $\delta_1 \equiv \delta/(2|\mathcal{S}||\mathcal{Z}|)$. By the union of events bound, the probability of error is bounded by

$$\begin{aligned} &P_{e|m=1}^{(Tn)}(\rho_{A^{Tn}}, \Lambda_{B^{Tn}}) \\ &\leq \sum_{j=1}^{T} \Pr(\mathscr{E}_1(j)) + \sum_{j=0}^{T-1} \Pr(\mathscr{E}_2(j+1) \mid \mathscr{E}_1^c(j) \cap \mathscr{E}_1^c(j+1)) \\ &+ \sum_{j=0}^{T-1} \Pr(\mathscr{E}_3(j+1) \mid \mathscr{E}_1^c(j) \cap \mathscr{E}_1^c(j+1) \cap \mathscr{E}_2^c(j+1)) \\ &+ \sum_{j=0}^{T-1} \Pr\big(\mathscr{E}_4(j+1) \mid \mathscr{E}_1^c(j) \cap \mathscr{E}_1^c(j+1) \\ &\cap \mathscr{E}_2^c(j+1) \cap \mathscr{E}_3^c(j+1)\big) \end{aligned} \tag{147}$$

where the conditioning on $M_j = L_j = L_{j-1} = 1$ is omitted for convenience of notation. By the classical covering lemma, the probability terms $\Pr(\mathscr{E}_1(j))$ tend to zero as $n \to \infty$ for

$$\widetilde{R}_s > I(X,Z;S) + \varepsilon_1(\delta) = I(Z;S|X) + \varepsilon_1(\delta) \tag{148}$$

where the last equality holds since the random variables X and S are statistically independent, using the notation $\varepsilon_i(\delta)$ for terms that tend to zero as $\delta \to 0$.

To bound the second sum, we use the quantum packing lemma. Given $\mathscr{E}_1^c(j)$, we have that $X^n(1,1) \in \mathcal{A}^{\delta/2}(p_X)$. Now, observe that

$$\Pi^\delta(\rho_B)\rho_{B^n}\Pi^\delta(\rho_B) \preceq 2^{-n(H(B)_\rho - \varepsilon_2(\delta))}\Pi^\delta(\rho_B) \tag{149}$$

$$\operatorname{Tr}\left[\Pi^\delta(\rho_B|x^n)\rho_{B^n}^{x^n}\right] \geq 1 - \varepsilon_2(\delta) \tag{150}$$

$$\operatorname{Tr}\left[\Pi^\delta(\rho_B|x^n)\right] \leq 2^{n(H(B|X)_\rho + \varepsilon_2(\delta))} \tag{151}$$

$$\operatorname{Tr}\left[\Pi^\delta(\rho_B)\rho_{B^n}^{x^n}\right] \geq 1 - \varepsilon_2(\delta) \tag{152}$$

for all $x^n \in \mathcal{A}^{\delta_1}(p_X)$, by (103), (109), (111), and (112), respectively. Since the codebooks are statistically independent of each other, we have by Lemma 12 that there exists a POVM $\Lambda^1_{m_{j+1},\ell_j}$ such that $\Pr(\mathscr{E}_2(j+1) \mid \mathscr{E}_1^c(j) \cap \mathscr{E}_1^c(j+1)) \leq 2^{-n(I(X;B)_\rho - (R+R_s) - \varepsilon_3(\delta))}$, which tends to zero as $n \to \infty$, provided that

$$R < I(X;B)_\rho - R_s - \varepsilon_3(\delta). \tag{153}$$

Moving to the third sum in the RHS of (147), suppose that $\mathscr{E}_2^c(j+1)$ occurred, namely the decoder measured the correct M_{j+1} and L_j. Denote the state of the systems B_j^n after this measurement by $\rho'_{B_j^n}$. Then, observe that due to the packing lemma inequality (119), Lemma 13 (the gentle measurement lemma) implies that the post-measurement state is close to the original state in the sense that

$$\frac{1}{2}\left\|\rho'_{B_j^n} - \rho_{B_j^n}\right\|_1 \leq 2^{-n\frac{1}{2}(I(X;B)_\rho - (R+R_s) - \varepsilon_4(\delta))} \leq \varepsilon_5(\delta) \tag{154}$$

for sufficiently large n and rates as in (153). Therefore, the distribution of measurement outcomes when $\rho'_{B_j^n}$ is measured is roughly the same as if the POVM $\Lambda^1_{m_{j+1},\ell_j}$ was never performed. To be precise, the difference between the probability of a measurement outcome $\widehat{k}_j$ when $\rho'_{B_j^n}$ is measured and the probability when $\rho_{B_j^n}$ is measured is bounded by $\varepsilon_5(\delta)$ in absolute value [27, Lemma 9.11]. Furthermore,

$$\operatorname{Tr}\left[\Pi^\delta(\rho_B|x^n, z^n)\sigma_{B^n}^{x^n,z^n}\right] \geq 1 - \varepsilon_6(\delta) \tag{155}$$

$$\Pi^\delta(\rho_B|x^n)\rho_{B^n}^{x^n}\Pi^\delta(\rho_B|x^n) \preceq 2^{-n(H(B|X)_\rho - \varepsilon_6(\delta))}\Pi^\delta(\rho_B|x^n) \tag{156}$$

$$\operatorname{Tr}\left[\Pi^\delta(\rho_B|x^n, z^n)\right] \leq 2^{n(H(B|X,Z)_\rho + \varepsilon_6(\delta))} \tag{157}$$

$$\operatorname{Tr}\left[\Pi^\delta(\rho_B|x^n)\sigma_{B^n}^{x^n,z^n}\right] \geq 1 - \varepsilon_6(\delta) \tag{158}$$

for all $(x^n, z^n) \in \mathcal{A}^{\delta/2}(p_X p_{Z|X})$, by (109), (110), (111), and (112), respectively. Therefore, we have by the packing lemma that there exists a POVM $\Lambda^2_{k_j|x^n}$ such that

$$\begin{aligned} &\Pr(\mathscr{E}_3(j+1) \mid \mathscr{E}_1^c(j) \cap \mathscr{E}_1^c(j+1) \cap \mathscr{E}_2^c(j+1)) \\ &\leq 2^{-n(I(Z;B|X)_\rho - (\widetilde{R}_s - R_s) - \varepsilon_7(\delta))} \end{aligned} \tag{159}$$

which tends to zero as $n \to \infty$, provided that

$$R_s > \widetilde{R}_s - I(Z;B|X)_\rho + \varepsilon_7(\delta). \tag{160}$$

It remains to verify that the distortion requirement is satisfied. Suppose that $\mathscr{E}_3^c(j+1)$ occurred, namely the decoder measured the correct K_{j+1}. Denote the post-measurement

state by $\rho''_{B_j^n}$. As before, the gentle measurement lemma guarantees that the difference between the probability of a measurement outcome $\widehat{s}$ when $\rho''_{B_j^n}$ is measured and the probability when $\rho'_{B_j^n}$ is measured is bounded by $\varepsilon_5(\delta)$ in absolute value. Therefore, given $\mathscr{E}_1^c(j)\cap\mathscr{E}_1^c(j+1)\cap\mathscr{E}_2^c(j+1)\cap\mathscr{E}_3^c(j+1)$, the parameter sequence S_{j+1}^n and the reconstruction $\hat{S}_{j+1}^n$ have a product distribution that is $2\varepsilon_5(\delta)$-close to

$$\Pr\left(S=s,\hat{S}=\hat{s}\right)=q(s)\sum_{x,z}p_X(x)p_{Z|X}(z|x)\cdot\operatorname{Tr}(\Gamma^{\hat{s}}_{B|x,z}\mathcal{N}^{(s)}(\theta^x_A)). \tag{161}$$

By (136), the distribution above satisfies $\mathbb{E}d(S,\hat{S})\leq D$, hence the last term tends to zero as $n\to\infty$ by the law of large numbers. By the law of total expectation,

$$\mathbb{E}d^{Tn}(S^{Tn},\hat{S}^{Tn})\leq\sum_{j=1}^{T}\Pr\left(\mathscr{E}_1(j)\cup\mathscr{E}_2(j)\cup\mathscr{E}_3(j)\cup\mathscr{E}_4(j)\right)d_{\max}+D+\frac{1}{2}\delta_0. \tag{162}$$

Thereby, the asymptotic average distortion is bounded by $(D+\delta_0)$ and the probability of error tends to zero as $n\to\infty$ for rates that satisfy (148), (153), and (160), which requires

$$\begin{aligned}R&<I(X;B)_\rho-(I(Z;S|X)-I(Z;B|X)_\rho\\&\quad+\varepsilon_1(\delta)+\varepsilon_7(\delta))-\varepsilon_3(\delta)\\&=I(X,Z;B)_\rho-I(Z;S|X)-\varepsilon_8(\delta).\end{aligned} \tag{163}$$

To show that rate-distortion pairs in $\frac{1}{\kappa}\mathcal{R}_{\text{s-c}}(\mathcal{N}^{\otimes\kappa})$ are achievable as well, one may employ the coding scheme above for the product channel $\mathcal{N}^{\otimes\kappa}$, where κ is arbitrarily large. This completes the proof of the direct part.

B. Converse Proof

Consider the converse part for the regularized capacity formula. As can be seen below, a regularized converse is straightforward. Let M be a uniformly distributed message. Suppose that at time $i\in[1:n]$, Alice sends $\rho_{A_i}^{m,s^{i-1}}$ over the channel. After Alice sends the systems A^n through the channel, Bob receives the systems B^n in the state $\rho_{B^n}=\frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}}\sum_{s^n\in\mathcal{S}^n}q^n(s^n)\mathcal{N}^{(s^n)}(\overline{\mathcal{F}}(m,s^n))$ with

$$\overline{\mathcal{F}}_{M,S^n\to A^n}=\bigotimes_{i=1}^{n}\mathcal{F}^{(i)}_{M,S^{i-1}\to A_i}. \tag{164}$$

Then, Bob performs a decoding POVM $\Lambda_{B^n}^{m,\hat{s}^n}$.

Consider a sequence of codes $(\overline{\mathcal{F}}_{MS^n\to A^n},\Lambda_{B^n})$ such that the average probability of error tends to zero and the distortion requirement holds. That is,

$$\Pr\left(\hat{M}\neq M\right)\leq\alpha_n, \tag{165}$$

and

$$\Delta^{(n)}(\mathcal{F},\Lambda)\leq D. \tag{166}$$

By Fano's inequality, (165) implies that $H(M|\hat{M})\leq n\varepsilon_n$, where ε_n tends to zero as $n\to\infty$. Hence,

$$nR=H(M)\leq I(M;\hat{M})_\rho+n\varepsilon_n\leq I(M;B^n)_\rho+n\varepsilon_n \tag{167}$$

where the last inequality follows from the Holevo bound [66, Theorem 12.1]. Since M and S^n are statistically independent, we can write the last bound as

$$\begin{aligned}R&\leq\frac{1}{n}[I(M;B^n)_\rho-I(M;S^n)]+\varepsilon_n\\&=\frac{1}{n}[I(X^n,Z^n;B^n)_\rho-I(X^n,Z^n;S^n)]+\varepsilon_n\end{aligned} \tag{168}$$

for $X^n=f(M)$ and $Z^n=\emptyset$, where f is an arbitrary one-to-one map from $[1:2^{nR}]$ to $\mathcal{X}^n$.

As for the distortion requirement,

$$\begin{aligned}D&\geq\Delta^{(n)}(\mathcal{E},\Lambda)=\mathbb{E}d^n(S^n,\hat{S}^n)\\&=P_e^{(n)}(\rho_{A^n},\Lambda_{B^n})\mathbb{E}[d^n(S^n,\hat{S}^n)\mid\hat{M}\neq M]\\&\quad+(1-P_e^{(n)}(\rho_{A^n},\Lambda_{B^n}))\mathbb{E}[d^n(S^n,\hat{S}^n)\mid\hat{M}=M]\\&\geq(1-P_e^{(n)}(\rho_{A^n},\Lambda_{B^n}))\mathbb{E}[d^n(S^n,\hat{S}^n)\mid\hat{M}=M]\\&\geq(1-\alpha_n)\mathbb{E}[d^n(S^n,\hat{S}^n)\mid\hat{M}=M]\end{aligned} \tag{169}$$

where we have used the law of total expectation in the second line, and (165) in the last line. Thus,

$$\begin{aligned}D&\geq\mathbb{E}[d^n(S^n,\hat{S}^n)\mid\hat{M}=M]-\alpha_n d_{\max}\\&=\sum_{s^n\in\mathcal{S}^n}\sum_{\hat{s}^n\in\hat{\mathcal{S}}^n}d^n(s^n,\hat{s}^n)q^n(s^n)\\&\quad\cdot\frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}}\operatorname{Tr}\left[\Lambda_{B^n}^{m,\hat{s}^n}\mathcal{N}^{(s^n)}_{A^n\to B^n}(\overline{\rho}^{m,s^n}_{A^n})\right]-\alpha_n d_{\max}\qquad(170)\\&=\sum_{s^n\in\mathcal{S}^n}\sum_{x^n\in\mathcal{X}^n}\sum_{\hat{s}^n\in\widehat{S}^n}q^n(s^n)p_{X^n}(x^n)\\&\quad\cdot\operatorname{Tr}(\Gamma^{\hat{s}^n}_{B^n|x^n}\rho_{B^n}^{s^n,x^n})d^n(s^n,\hat{s}^n)\qquad(171)\end{aligned}$$

with $\Gamma^{\hat{s}^n}_{B^n|f(m)}\equiv\Lambda_{B^n}^{m,\hat{s}^n}$. This concludes the converse proof for part 1.

Part 2: Now, we consider the quantum-classical special case of a measurement channel $\mathcal{M}_{SA\to Y}$. The direct part follows from part 1, taking $\kappa=1$. It remains to prove the converse part, which we show by extending the methods of Choudhuri *et al.* [3].

By (167) and the chain rule for classical mutual information, we have

$$nR\leq I(M;Y^n)+n\varepsilon_n=\sum_{i=1}^{n}I(M;Y_i|Y_{i+1}^n)+n\varepsilon_n. \tag{172}$$

We can rewrite the bound above as

$$\begin{aligned}&R-\varepsilon_n\\&\leq\frac{1}{n}\sum_{i=1}^{n}[I(M,S^{i-1};Y_i|Y_{i+1}^n)-I(S^{i-1};Y_i|M,Y_{i+1}^n)]\end{aligned}$$

$$= \frac{1}{n}\sum_{i=1}^{n}[I(M,S^{i-1};Y_i|Y_{i+1}^n) - I(Y_{i+1}^n;S_i|M,S^{i-1})]$$
$$\leq \frac{1}{n}\sum_{i=1}^{n}[I(M,S^{i-1},Y_{i+1}^n;Y_i) - I(Y_{i+1}^n;S_i|M,S^{i-1})] \quad (173)$$

where the equality follows from the Csiszár sum identity [71, Section 2.3]. Since the pair (M,S^{i-1}) is statistically independent of S_i, we have $I(Y_{i+1}^n;S_i|M,S^{i-1}) = I(M,S^{i-1},Y_{i+1}^n;S_i)$, hence

$$R-\varepsilon_n \leq \frac{1}{n}\sum_{i=1}^{n}[I(X_i,Z_i;Y_i) - I(X_i,Z_i;S_i)] \quad (174)$$

where we have defined $X_i=(M,S^{i-1})$ and $Z_i = Y_{i+1}^n$. Thus,

$$R-\varepsilon_n \leq I(X,Z;Y|J) - I(X,Z;S|J) \quad (175)$$

with

$$X \equiv X_J,\ S=S_J,\ Y=Y_J,\ \hat{S}=\hat{S}_J \quad (176)$$

where J is uniformly distributed over $[1:n]$, and independent of (M,S^n). Then, defining $X'=(X,J)$, we have that $I(X,Z;Y|J) \leq I(X',Z;Y)$ and $I(X,Z;S|J) = I(X',Z;S) = I(Z;S|X')$, hence $R-\varepsilon_n \leq I(X',Z;Y) - I(Z;S|X')$.

As for the distortion level,

$$D \geq \mathbb{E}d^n(S^n,\hat{S}^n) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}d(S_i,\hat{S}_i) = \mathbb{E}d(S,\hat{S}) \quad (177)$$

where the first equality holds since the distortion measure is additive (see (11)), and the second follows from the definition of S and $\hat{S}$ in (176). This completes the proof of Theorem 3. □

APPENDIX D
PROOF OF THEOREM 5

Since the proof is similar to that of Theorem 3 in Appendix C, we only give an outline. The converse proof in part 1 follows the same arguments, and it is thus omitted. Moving to the achievability proof, we need to show that for every $\zeta_0,\varepsilon_0,\delta_0>0$, there exists a $(2^{n(R-\zeta_0)},n,\varepsilon_0,D+\delta_0)$ code for $\mathcal{N}_{SA\to B}$ with causal CSI, provided that $(R,D)\in\mathcal{R}_{\text{caus}}(\mathcal{N})$. Here, the encoder has access to the sequence of past *and present* parameters $s_1,s_2,\ldots,s_i$. Let $\{p_X(x)p_{Z|X}(z|x),\theta_G^x\}$ be a given ensemble, and fix the channels $\mathcal{F}_{G\to A}^{(s)}$ and set of POVMs $\{\Gamma_{B|x,z}^{\hat{s}}\}$ such that

$$\sum_{s,\hat{s},x,z} q(s)p_X(x)p_{Z|X,S}(z|x,s)\text{Tr}(\Gamma_{B|x,z}^{\hat{s}}\mathcal{V}^{(s)}(\theta_G^x))d(s,\hat{s}) \leq D. \quad (178)$$

where the channel $\mathcal{V}_{G\to B}^{(s)}$ is defined by

$$\mathcal{V}^{(s)}(\rho_G) = \mathcal{N}^{(s)}(\mathcal{F}^{(s)}(\rho_G)). \quad (179)$$

Then, define the average states

$$\rho_B^x = \sum_{s\in\mathcal{S}} q(s)\mathcal{V}^{(s)}(\rho_G^x) \quad (180)$$

$$\rho_B = \sum_{x\in\mathcal{X}} p_X(x)\rho_B^x \quad (181)$$

and

$$\sigma_B^{x,z} = \sum_{s\in\mathcal{S}} p_{S|X,Z}(s|x,z)\mathcal{V}^{(s)}(\theta_G^x) \quad (182)$$

for $x\in\mathcal{X}$ and $z\in\mathcal{Z}$.

We use T transmission blocks, where each block consists of n input systems. The code construction, encoding and decoding procedures are described below.

Classical Code Construction: The classical code construction is the same as in the previous proof: Select i.i.d sequences $x_j^n(m_j,\ell_{j-1})$ according to p_X, and then for every (m_j,ℓ_{j-1}), select conditionally independent sequences $z_j^n(k_j|m_j,\ell_{j-1})$ according to $p_{Z|X}$, where $m_j\in[1:2^{nR}]$, $\ell_{j-1}\in[1:2^{nR_s}]$, $k_j\in[1:2^{n\widetilde{R}_s}]$. Partition the set of indices $[1:2^{n\widetilde{R}_s}]$ into bins $\mathcal{K}(\ell_j)$ of equal size $2^{n(\widetilde{R}_s-R_s)}$.

Encoding and Decoding: To send the messages (m_j), given the parameter sequences $(s_1^n,\ldots,s_j^n)$, Alice performs the following.

(i) Let $\ell_0=1$. At the end of block j, find an index $k_j\in[1:2^{n\widetilde{R}_s}]$ such that $(s_j^n,\ z_j^n(k_j|m_j,\ell_{j-1}),\ x_j^n(m_j,\ell_{j-1}))\in\mathcal{A}^{\delta}(p_{S,Z,X})$, where $p_{S,X,Z}(s,x,z) = q(s)p_X(x)\ \cdot p_{Z|X,S}(z|x,s)$. If there is none, select k_j arbitrarily, and if there is more than one, choose the smallest. Set ℓ_j to be the bin index of k_j, *i.e.* such that $k_j\in\mathcal{K}(\ell_j)$.

(ii) In block $j+1$, prepare $\rho_{A_{j+1}^n} = \bigotimes_{i=1}^n \mathcal{F}^{(s_{j+1,i})}(\theta_G^{x_{j+1,i}(m_{j+1},\ell_j)})$ and send the block A_{j+1}^n.

Bob receives the systems $B_1^n,\ldots,B_T^n$ in the state

$$\rho_{B^{Tn}} = \bigotimes_{j=1}^{T}\bigotimes_{i=1}^{n}\rho_B^{x_{j+1,i}(m_{j+1},\ell_j)}. \quad (183)$$

Observe that this is the same state as in (142) where the channel $\mathcal{N}^{(s)}$ is replaced by $\mathcal{V}^{(s)}$. Thus, Bob can decode reliably and satisfy the distortion requirement, provided that

$$R < I(X,Z;B)_\rho - I(X,Z;S) - \varepsilon(\delta). \quad (184)$$

This concludes the proof of part 1.

Part 2 also follows from a similar derivation as in Appendix C, except that now, the state of the input system A_i depends on $(m,s^{i-1},s_i)=(x_i,s_i)$. Hence, we choose the system G_i to be classical, with $\theta_{G_i}^{x_i}\equiv|x_i\rangle\langle x_i|$, and then we define the channel $\mathcal{F}_{G_i\to A_i}^{s_i}$ as a preparation channel. Specifically, given the knowledge of $x_i=(m,s^{i-1})$ from the state of G_i, the channel $\mathcal{F}_{G_i\to A_i}^{s_i}$ prepares the state $\rho_{A_i}^{m,s^{i-1},s_i} = \rho_{A_i}^{x_i,s_i}$. □

APPENDIX E
PROOF OF THEOREM 6

Consider a random-parameter quantum channel $\mathcal{N}_{SA\to B}$ with non-causal CSI at the encoder. We show that for every $\zeta_0,\varepsilon_0,\delta_0>0$, there exists a $(2^{n(R-\zeta_0)},n,\varepsilon_0,D+\delta_0)$ code for $\mathcal{N}_{SA\to B}$ with non-causal CSI, provided that $(R,D)\in\mathcal{R}_{\text{n-c}}(\mathcal{N})$. To prove achievability, we use an extension of the classical binning technique to the quantum setting, and then

apply the quantum packing lemma and the classical covering lemma.

Recall that with non-causal CSI, the encoder has access to the entire sequence of parameters $s_1, s_2, \ldots, s_n$ a priori. Let $\{p_{X|S}(x|s), \theta_A^x\}$ be a given ensemble, and fix a set of POVMs $\{\Gamma_{B|x}^{\hat{s}}\}$ such that

$$\sum_{s,\hat{s},x,z} q(s)p_{X|S}(x|s)\mathrm{Tr}(\Gamma_{B|x}^{\hat{s}}\mathcal{N}^{(s)}(\theta_A^x))d(s,\hat{s}) \leq D. \quad (185)$$

Define the average states

$$\rho_B^x = \sum_{s\in\mathcal{S}} q(s)\mathcal{N}^{(s)}(\rho_A^x) = \sum_{s\in\mathcal{S}} q(s)\sigma_B^{x,s} \quad (186)$$

with

$$\sigma_B^{x,s} = \mathcal{N}^{(s)}(\theta_A^{x,s}) \quad (187)$$

for $x \in \mathcal{X}$.

The code construction, encoding and decoding procedures are described below.

Classical Code Construction: Let $\delta > 0$ and $\widetilde{R}_s > 0$. Select $2^{n(R+\widetilde{R}_s)}$ independent sequences $x^n(m,\ell)$, $m \in [1 : 2^{nR}]$, $\ell \in [1 : 2^{n\widetilde{R}_s}]$, at random according to $\prod_{i=1}^n p_X(x_i)$.

Encoding and Decoding: To send the message m, given the parameter sequence s^n, Alice performs the following.

(i) Find an index $\ell \in [1 : 2^{n\widetilde{R}_s}]$ such that $(s^n, x^n(m,\ell)) \in \mathcal{A}^\delta(p_{S,X})$, where $p_{S,X}(s,x) = q(s)p_{X|S}(x|s)$. If there is none, select ℓ arbitrarily, and if there is more than one such index, choose the smallest.

(ii) Transmit $\rho_{A^n}^{m,\ell} = \bigotimes_{i=1}^n \theta_A^{x_i(m,\ell)}$

Bob receives the systems B^n in the state

$$\rho_{B^n} = \bigotimes_{i=1}^n \rho_B^{x_i(m,\ell)} \quad (188)$$

and decodes as follows.

(i) Decode $(\hat{m}, \hat{\ell})$ by applying a POVM $\{\Lambda_{m,\ell}\}_{(m,\ell)\in[1:2^{nR}]\times[1:2^{n\widetilde{R}_s}]}$, which will be specified later.

(ii) Reconstruct the parameter sequence by applying the POVM $\Gamma_{B|x_i}^{\hat{s}_i}$ to the system B_i with $x_i \equiv x_i(\hat{m}, \hat{\ell})$, for $i \in [1 : n]$.

Analysis of Probability of Error and Distortion: By symmetry, we may assume without loss of generality that Alice sends the message $M = 1$ using L. Consider the following events,

$$\mathscr{E}_1 = \{(S^n, X^n(1,\ell)) \notin \mathcal{A}^\delta(p_{S,X}), \text{ for all } \ell \in [1 : 2^{n\widetilde{R}_s}]\} \quad (189)$$

$$\mathscr{E}_2 = \{(\hat{M}, \hat{L}) \neq (1, L)\} \quad (190)$$

$$\mathscr{E}_3 = \{d^n(S^n, \hat{S}^n) > D + \frac{1}{2}\delta_0\}. \quad (191)$$

By the union of events bound, the probability of error is bounded by

$$P_{e|m=1}^{(n)}(\rho_{A^n}, \Lambda_{B^n}) \leq \Pr(\mathscr{E}_1) + \Pr(\mathscr{E}_2 \mid \mathscr{E}_1^c) + \Pr(\mathscr{E}_3 \mid \mathscr{E}_1^c \cap \mathscr{E}_2^c) \quad (192)$$

where the conditioning on $M = 1$ is omitted for convenience of notation. By the classical covering lemma, the first term tends to zero as $n \to \infty$ for

$$\widetilde{R}_s > I(X;S) + \varepsilon_1(\delta). \quad (193)$$

To bound the second term, we use the quantum packing lemma. Given $\mathscr{E}_1^c$, we have that $X^n(1,L) \in \mathcal{A}^{\delta_1}(p_X)$, with $\delta_1 \triangleq \delta|\mathcal{S}||\mathcal{Z}|$. Now, observe that

$$\Pi^\delta(\rho_B)\rho_{B^n}\Pi^\delta(\rho_B) \preceq 2^{-n(H(B)_\rho - \varepsilon_2(\delta))}\Pi^\delta(\rho_B) \quad (194)$$

$$\mathrm{Tr}\left[\Pi^\delta(\rho_B|x^n)\rho_{B^n}^{x^n}\right] \geq 1 - \varepsilon_2(\delta) \quad (195)$$

$$\mathrm{Tr}\left[\Pi^\delta(\rho_B|x^n)\right] \leq 2^{n(H(B|X)_\rho + \varepsilon_2(\delta))} \quad (196)$$

$$\mathrm{Tr}\left[\Pi^\delta(\rho_B)\rho_{B^n}^{x^n}\right] \geq 1 - \varepsilon_2(\delta) \quad (197)$$

for all $x^n \in \mathcal{A}^{\delta_1}(p_X)$, by (103), (109), (111), and (112), respectively. Thus, by Lemma 12, there exists a POVM $\Lambda_{m,\ell}$ such that $\Pr(\mathscr{E}_2 \mid \mathscr{E}_1^c) \leq 2^{-n(I(X;B)_\rho - (R+\widetilde{R}_s) - \varepsilon_3(\delta))}$, which tends to zero as $n \to \infty$, provided that

$$R < I(X;B)_\rho - \widetilde{R}_s - \varepsilon_3(\delta). \quad (198)$$

Moving to the third sum in the RHS of (192), suppose that $\mathscr{E}_2^c$ occurred, *i.e.* the decoder measured the correct M and L. Then, due to the packing lemma inequality (119) and Lemma 13 (the gentle measurement lemma), the post-measurement state ρ'_{B^n} is close to the original state ρ_{B^n} in the sense that

$$\frac{1}{2}\|\rho'_{B^n} - \rho_{B^n}\|_1 \leq 2^{-n\frac{1}{2}(I(X;B)_\rho - (R+R_s) - \varepsilon_4(\delta))} \leq \varepsilon_5(\delta) \quad (199)$$

for sufficiently large n and rates as in (198). Thus, the difference between the probability of a measurement outcome $\hat{s}$ when ρ'_{B^n} is measured and the probability when ρ_{B^n} is measured is bounded by $\varepsilon_5(\delta)$ in absolute value [27, Lemma 9.11].

Therefore, given $\mathscr{E}_1^c \cap \mathscr{E}_2^c \cap \mathscr{E}_3^c$, the parameter sequence S^n and the reconstruction $\hat{S}^n$ have a product distribution according to

$$\Pr\left(S = s, \hat{S} = \hat{s}\right) = q(s)\sum_{x,z} p_X(x)p_{Z|X}(z|x)\mathrm{Tr}(\Gamma_{B|x,z}^{\hat{s}}\rho_B^s) \pm \varepsilon_5(\delta). \quad (200)$$

By (185), the distribution above satisfies $\mathbb{E}d(S,\hat{S}) \leq D$, hence the last term, $\Pr(\mathscr{E}_3 \mid \mathscr{E}_1^c \cap \mathscr{E}_2^c)$, tends to zero as $n \to \infty$ by the law of large numbers. It follows by the law of total expectation,

$$\mathbb{E}d^n(S^n, \hat{S}^n) \leq \Pr(\mathscr{E}_1 \cup \mathscr{E}_2 \cup \mathscr{E}_3)\, d_{\max} + D + \frac{1}{2}\delta_0. \quad (201)$$

Thereby, the asymptotic average distortion is bounded by $(D + \delta_0)$ and the probability of error tends to zero as $n \to \infty$ for rates that satisfy (193) and (198), which requires

$$R < I(X;B)_\rho - I(X;S) - \varepsilon_1(\delta) - \varepsilon_3(\delta). \quad (202)$$

To show that rate-distortion pairs in $\frac{1}{\kappa}\mathcal{R}_{\text{n-c}}(\mathcal{N}^{\otimes\kappa})$ are achievable as well, one may employ the coding scheme above

for the product channel $\mathcal{N}^{\otimes\kappa}$, where κ is arbitrarily large. This completes the proof of the direct part.

The converse part follows by the same arguments as in the previous proofs, and it is thus omitted. □

APPENDIX F
PROOF OF THEOREM 7

Consider a random-parameter quantum channel $\mathcal{N}_{SA\to B}$ without CSI.

Part 1: Given our previous analysis, the proof of part 1 is straightforward. Achievability is shown using the coding scheme in the proof of Theorem 6 in Appendix E with the following modifications. The random variable X is statistically independent of the random parameter, *i.e.* $p_{X|S}$ is replaced by p_X. The input state does not depend on the random parameter, hence $\theta_A^{x,s}$ is replaced by θ_A^x. Set $\widetilde{R}_s \to 0$. Hence, $\ell \equiv 1$, the encoding stage (i) is no longer necessary, and the error event $\mathscr{E}_1$ can be ignored. Then, by the same considerations as in Appendix E, we have that the asymptotic average distortion is bounded by $(D+\delta_0)$ and the probability of error tends to zero as $n\to\infty$, provided that

$$R < I(X;B)_\rho - \varepsilon_3(\delta). \tag{203}$$

To show that rate-distortion pairs in $\frac{1}{\kappa}\mathcal{R}(\mathcal{N}^{\otimes\kappa})$ are achievable as well, employ this coding scheme for the product channel $\mathcal{N}^{\otimes\kappa}$. The details are omitted.

The converse proof also follows similar arguments as in the previous proofs. Suppose that Alice sends $\rho_{A^n}^m$ over the channel. Bob receives the systems B^n in the state $\rho_{B^n} = \frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}}\sum_{s^n\in\mathcal{S}^n} q^n(s^n)\rho_{B^n}^{m,s^n}$, with

$$\rho_{B^n}^{m,s^n} \equiv \mathcal{N}^{(s^n)}(\rho_{A^n}^m) \tag{204}$$

and then, performs a decoding POVM $\Lambda_{B^n}^{m,\hat{s}^n}$. Now, consider a sequence of codes $(\mathcal{E}_{M\to A^n},\ \Lambda_{B^n})$ such that the average probability of error tends to zero and the distortion requirement holds. That is,

$$P_e^{(n)}(\mathcal{E},\Lambda) \le \alpha_n, \tag{205}$$

and

$$\Delta^{(n)}(\mathcal{E},\Lambda) \le D. \tag{206}$$

By Fano's inequality, (205) implies that $H(M|\hat{M}) \le n\varepsilon_n$, where ε_n tends to zero as $n\to\infty$. Hence,

$$nR = H(M) \le I(M;\hat{M})_\rho + n\varepsilon_n \le I(M;B^n)_\rho + n\varepsilon_n \tag{207}$$

where the last inequality follows from the Holevo bound [66, Theorem 12.1]. Thus,

$$R \le \frac{1}{n}I(M;B^n)_\rho + \varepsilon_n = \frac{1}{n}I(X^n;B^n)_\rho + \varepsilon_n \tag{208}$$

for $X^n = f(M)$ where f is an arbitrary one-to-one map from $[1:2^{nR}]$ to $\mathcal{X}^n$.

As for the distortion requirement,

$$\begin{aligned} D \ge \Delta^{(n)}(\mathcal{E},\Lambda) &= \mathbb{E}d^n(S^n,\hat{S}^n) \\ &= P_e^{(n)}(\rho_{A^n},\Lambda_{B^n})\mathbb{E}[d^n(S^n,\hat{S}^n)\mid \hat{M}\neq M] \\ &\quad + (1-P_e^{(n)}(\rho_{A^n},\Lambda_{B^n}))\mathbb{E}[d^n(S^n,\hat{S}^n)\mid \hat{M}=M] \\ &\ge (1-P_e^{(n)}(\rho_{A^n},\Lambda_{B^n}))\mathbb{E}[d^n(S^n,\hat{S}^n)\mid \hat{M}=M] \\ &\ge (1-\alpha_n)\mathbb{E}[d^n(S^n,\hat{S}^n)\mid \hat{M}=M] \end{aligned} \tag{209}$$

where we have used the law of total expectation in the second line, and (205) in the last line. Thus,

$$\begin{aligned} D &\ge \mathbb{E}[d^n(S^n,\hat{S}^n)\mid \hat{M}=M] - \alpha_n d_{\max} \\ &= \sum_{s^n\in\mathcal{S}^n}\sum_{\hat{s}^n\in\widehat{\mathcal{S}}^n} d^n(s^n,\hat{s}^n)q^n(s^n) \\ &\quad\cdot \frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}} \mathrm{Tr}\Big[\Lambda_{B^n}^{m,\hat{s}^n}\mathcal{N}_{A^n\to B^n}^{(s^n)}(\overline{\rho}_{A^n}^m)\Big] - \alpha_n d_{\max} \qquad (210) \\ &= \sum_{s^n\in\mathcal{S}^n}\sum_{x^n\in\mathcal{X}^n}\sum_{\hat{s}^n\in\widehat{\mathcal{S}}^n} q^n(s^n)p_{X^n}(x^n) \\ &\quad\cdot \mathrm{Tr}(\Gamma_{B^n|x^n}^{\hat{s}^n}\rho_{B^n}^{s^n,x^n})d^n(s^n,\hat{s}^n) \qquad (211) \end{aligned}$$

with $\Gamma_{B^n|f(m)}^{\hat{s}^n} \equiv \Lambda_{B^n}^{m,\hat{s}^n}$.

The union in (56) is restricted to pure states $\theta_A^x = |\phi_A^x\rangle\langle\phi_A^x|$ with $|\mathcal{X}| \le |\mathcal{H}_A|^2+1$, based on the same arguments as in the proof of Lemma 2 in Appendix B. This concludes the converse proof for part 1.

Part 2: Now, we consider an entanglement-breaking channel. The direct part follows from part 1, taking $\kappa = 1$. It remains to prove the converse part, which we show by extending the methods of Wang *et al.* [58]. By (207), we have

$$\begin{aligned} R &\le \frac{1}{n}I(M;B^n)_\rho + \varepsilon_n \\ &= \frac{1}{n}\sum_{i=1}^n I(M;B_i|B^{i-1})_\rho \\ &\le \frac{1}{n}\sum_{i=1}^n I(M,B^{i-1};B_i)_\rho \end{aligned} \tag{212}$$

by the chain rule. Without CSI, the channel input systems A^n have no correlation with the channel parameter sequence S^n. As the channel is memoryless, it follows that B_i and S^{i-1} are in a product state. Then,

$$\begin{aligned} I(M,B^{i-1};B_i)_\rho &\le I(M,B^{i-1},S^{i-1};B_i)_\rho \\ &= I(M,B^{i-1};B_i|S^{i-1})_\rho + I(S^{i-1};B_i)_\rho \\ &= I(M,B^{i-1};B_i|S^{i-1})_\rho \end{aligned} \tag{213}$$

where the last equality holds since $I(S^{i-1};B_i)_\rho = 0$.

If the random-parameter quantum channel is entanglement breaking, then $\mathcal{N}_{A\to B}^{(s)}$ can be presented as a concatenation of a measurement channel, followed by a state-preparation channel, i.e.

$$\mathcal{N}_{A\to B}^{(s)} = \mathcal{P}_{Y_s\to B}^{(s)} \circ \mathcal{M}_{A\to Y_s}^{(s)}$$

where Y_s is classical, for $s\in\mathcal{S}$ (see Subsection II-B). Therefore, by the quantum data processing theorem due to Schumacher and Nielsen [68] [27, Theorem 11.9.4],

$$\begin{aligned} I(M,B^{i-1};B_i|S^{i-1}=s^{i-1})_\rho \le \\ I(M,Y_{s^{i-1}}^{i-1};B_i|S^{i-1}=s^{i-1})_\rho. \end{aligned} \tag{214}$$

By (212)- (214), we have

$$R-\varepsilon_n \le \frac{1}{n}\sum_{i=1}^{n} I(M, Y_{S^{i-1}}^{i-1}; B_i|S^{i-1})_\rho \le \frac{1}{n}\sum_{i=1}^{n} I(X_i; B_i)_\rho \quad (215)$$

with $X_i \equiv (M, Y_{S^{i-1}}^{i-1}, S^{i-1})$. Define

$$X \equiv X_J,\ S \equiv S_J,\ \hat{S} \equiv \hat{S}_J \quad (216)$$

where J is a classical time-sharing variable that is uniformly distributed over $[1:n]$. Observe that for this choice of X, we have

$$\rho_{XB} \equiv \sum_x p_X(x)|x\rangle\langle x| \otimes \mathcal{N}(\phi_A^x) = \frac{1}{n}\sum_{i=1}^{n}\sum_{x_i} p_{X_i}(x_i)|x_i\rangle\langle x_i| \otimes \mathcal{N}(\sigma_{A_i}^{x_i}) \quad (217)$$

where $\sigma_{A_i}^{m,y_{s^{i-1}}^{i-1},s^{i-1}} = \rho_{A_i}^m$. Thus, by (215),

$$R \le I(X;B|J)_\rho + \varepsilon_n \le I(X';B) + \varepsilon_n \quad (218)$$

with $X' \equiv (X, J)$.

As for the distortion,

$$D \ge \mathbb{E}d^n(S^n, \hat{S}^n) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}d(S_i, \hat{S}_i) = \mathbb{E}d(S, \hat{S}) \quad (219)$$

where the first equality holds since the distortion measure is additive (see (11)), and the second follows from the definition of S and $\hat{S}$ in (216). This completes the proof of Theorem 7. □

ACKNOWLEDGMENT

The author would like to thank Roberto Ferrara (Technical University of Munich) for useful discussions.

REFERENCES

[1] H. Boche, N. Cai, and J. Nötzel, "The classical-quantum channel with random state parameters known to the sender," *J. Phys. A, Math. Theor.*, vol. 49, no. 19, Apr. 2016, Art. no. 195302.

[2] G. Keshet, Y. Steinberg, and N. Merhav, "Channel coding in the presence of side information," *Found. Trends Commun. Inf. Theory*, vol. 4, no. 6, pp. 445–586, Jan. 2007.

[3] C. Choudhuri, Y.-H. Kim, and U. Mitra, "Causal state communication," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3709–3719, Jun. 2013.

[4] U. Pereg and Y. Steinberg, "The arbitrarily varying channel under constraints with side information at the encoder," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 861–887, Feb. 2019.

[5] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proc. IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.

[6] C. Heegard and A. El Gamal, "On the capacity of computer memory with defects," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 5, pp. 731–739, Sep. 1983.

[7] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 563–593, Mar. 2003.

[8] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[9] S. Sedghi, M. Khademi, and N. Cvejic, "Analysis of channel capacity of spread spectrum audio watermarking system," in *Proc. Int. Symp. Intell. Sig. Process. Commun.*, Dec. 2006, pp. 175–178.

[10] A. Sutivong, M. Chiang, T. M. Cover, and Y.-H. Kim, "Channel capacity and state estimation for state-dependent Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1486–1495, Apr. 2005.

[11] W. Zhang, S. Vedantam, and U. Mitra, "A constrained channel coding approach to joint communication and channel estimation," in *Proc. IEEE Int. Symp. Inf. Theory*, Toronto, ON, Canada, Jul. 2008, pp. 930–934.

[12] W. Zhang, S. Vedantam, and U. Mitra, "A constrained channel coding approach to joint transmission and state estimation problem," in *Proc. IEEE Int. Symp. Inf. Theory*, Toronto, ON, Canada, Aug. 2008, pp. 930–934.

[13] A. Sutivong, "Channel capacity and state estimation for state-dependent channel," Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ., Palo Alto, CA, USA, 2003.

[14] S. I. Bross and A. Lapidoth, "The rate- and-state capacity with feedback," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1893–1918, Mar. 2018.

[15] J. P. Dowling and G. J. Milburn, "Quantum technology: The second quantum revolution," *Philos. Trans. Royal Soc. A, Math., Phys. Eng. Sci.*, vol. 361, no. 1809, pp. 1655–1674, Jun. 2003.

[16] P. Jouguet, S. Kunz-Jacques, A. Leverrier, P. Grangier, and E. Diamanti, "Experimental demonstration of long-distance continuous-variable quantum key distribution," *Nature Photon.*, vol. 7, no. 5, p. 378, 2013.

[17] A. Orieux and E. Diamanti, "Recent advances on integrated quantum communications," *J. Opt.*, vol. 18, no. 8, Aug. 2016, Art. no. 083002.

[18] F. Flamini, N. Spagnolo, and F. Sciarrino, "Photonic quantum information processing: A review," *Rep. Prog. Phys.*, vol. 82, no. 1, Nov. 2018, Art. no. 016001.

[19] L. Petit *et al.*, "Universal quantum logic in hot silicon qubits," *Nature*, vol. 580, no. 7803, pp. 355–359, Apr. 2020.

[20] L. Gyongyosi, S. Imre, and H. V. Nguyen, "A survey on quantum channel capacities," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 1149–1205, 2nd Quart., 2018.

[21] G. Smith and J. T. Yard, "Quantum communication with zero-capacity channels," *Science*, vol. 321, pp. 1812–1815, Sep. 2008.

[22] A. S. Holevo, "The capacity of the quantum channel with general signal states," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 269–273, Jan. 1998.

[23] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels," *Phys. Rev. A, Gen. Phys.*, vol. 56, p. 131, Jul. 1997.

[24] M. B. Hastings, "Superadditivity of communication capacity using entangled inputs," *Nature Phys.*, vol. 5, no. 4, p. 255, Mar. 2009.

[25] A. S. Holevo, *Quantum Systems, Channels, Information: A Mathematical Introduction*. Berlin, Germany: De Gruyter, 2012.

[26] P. W. Shor, "Additivity of the classical capacity of entanglement-breaking quantum channels," *J. Math. Phys.*, vol. 43, no. 9, pp. 4334–4340, May 2002.

[27] M. M. Wilde, *Quantum Information Theory*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[28] I. Devetak, "The private classical capacity and quantum capacity of a quantum channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 44–55, Jan. 2005.

[29] N. A. Warsi and J. P. Coon, "Coding for classical-quantum channels with rate limited side information at the encoder: Information-spectrum approach," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 3322–3331, May 2017.

[30] A. Anshu, M. Hayashi, and N. A. Warsi, "Secure communication over fully quantum gel-fand-pinsker wiretap channel," *IEEE Trans. Inf. Theory*, vol. 66, no. 9, pp. 5548–5566, Jun. 2020.

[31] F. Dupuis, "Coding for quantum channels with side information at the transmitter," 2008, *arXiv:0805.3352*.

[32] F. Dupuis, "The capacity of quantum channels with side information at the transmitter," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2009, pp. 948–952.

[33] U. Pereg, "Entanglement-assisted capacity of quantum channels with side information," in *Proc. Int. Zürich Seminar Inf. Commun. (IZS)*, Zürich, Switzerland, Feb. 2020, pp. 106–110.

[34] U. Pereg, "Entanglement-assisted capacity of quantum channels with side information," 2019, *arXiv:1909.09992*.

[35] A. Anshu, R. Jain, and N. A. Warsi, "On the near-optimality of one-shot classical communication over quantum channels," *J. Math. Phys.*, vol. 60, no. 1, Jan. 2019, Art. no. 012204.

[36] Z. Luo and I. Devetak, "Channel simulation with quantum side information," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1331–1342, Mar. 2009.

[37] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.
[38] N. Datta, C. Hirche, and A. Winter, "Convexity and operational interpretation of the quantum information bottleneck function," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 1157–1161.
[39] H.-C. Cheng, E. P. Hanson, N. Datta, and M.-H. Hsieh, "Duality between source coding with quantum side information and c-q channel coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Paris, France, Jul. 2019, pp. 1142–1146.
[40] Z. B. Khanian and A. Winter, "Distributed compression of correlated classical-quantum sources or: The price of ignorance," *IEEE Trans. Inf. Theory*, vol. 66, no. 9, pp. 5620–5633, Mar. 2020.
[41] J. A. Smolin, F. Verstraete, and A. Winter, "Entanglement of assistance and multipartite state distillation," *Phys. Rev. A, Gen. Phys.*, vol. 72, Nov. 2005, Art. no. 052317.
[42] A. Winter, "On environment-assisted capacities of quantum channels," 2005, *arXiv:quant-ph/0507045*.
[43] S. K. Oskouei, S. Mancini, and A. Winter, "Capacities of Gaussian quantum channels with passive environment assistance," 2021, *arXiv:2101.00602*.
[44] U. Pereg, C. Deppe, and H. Boche, "Quantum channel state masking," *IEEE Trans. Inf. Theory*, vol. 67, no. 4, pp. 2245–2268, Apr. 2021.
[45] I. Savov, M. M. Wilde, and M. Vu, "Partial decode-forward for quantum relay channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Cambridge, MA, USA, Jul. 2012, pp. 731–735.
[46] D. Ding, H. Gharibyan, P. Hayden, and M. Walter, "A quantum multiparty packing lemma and the relay channel," *IEEE Trans. Inf. Theory*, vol. 66, no. 6, pp. 3500–3519, Jun. 2020.
[47] A. Fujiwara, "Quantum channel identification problem," in *Asymptotic Theory of Quantum Statistical Inference*. Singapore: World Scientific, 2005, pp. 487–493.
[48] Z. Ji, G. Wang, R. Duan, Y. Feng, and M. Ying, "Parameter estimation of quantum channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5172–5185, Nov. 2008.
[49] M. Zorzi, F. Ticozzi, and A. Ferrante, "On quantum channel estimation with minimal resources," 2011, *arXiv:1106.2105*.
[50] B. R. Bardhan and J. H. Shapiro, "Ultimate capacity of a linear time-invariant bosonic channel," *Phys. Rev. A, Gen. Phys.*, vol. 93, no. 3, Mar. 2016, Art. no. 032342.
[51] I. Savov, "Network information theory for classical-quantum channels," Ph.D. dissertation, McGill Univ., Montreal, QC, Canada, 2012.
[52] S. Kumar and M. J. Deen, *Fiber Optic Communications: Fundamentals and Applications*. Hoboken, NJ, USA: Wiley, 2014.
[53] C. Weedbrook *et al.*, "Gaussian quantum information," *Rev. Modern Phys.*, vol. 84, no. 2, pp. 621–669, May 2012.
[54] M. M. Wilde, P. Hayden, and S. Guha, "Quantum trade-off coding for bosonic communication," *Phys. Rev. A, Gen. Phys.*, vol. 86, no. 6, Dec. 2012, Art. no. 062306.
[55] A. S. Holevo and R. F. Werner, "Evaluating capacities of bosonic Gaussian channels," *Phys. Rev. A, Gen. Phys.*, vol. 63, Feb. 2001, Art. no. 032312.
[56] J. Eisert and M. Wolf, "Gaussian quantum channels," in *Quantum Information with Continuous Variables of Atoms and Light*. Singapore: World Scientific, 2007, pp. 23–42.
[57] V. Giovannetti, S. Lloyd, and L. Maccone, "Advances in quantum metrology," *Nature Photon.*, vol. 5, no. 4, p. 222, 2011.
[58] Q. Wang, S. Das, and M. M. Wilde, "Hadamard quantum broadcast channels," *Quantum Inf. Process.*, vol. 16, no. 10, p. 248, Oct. 2017.
[59] M.-H. Hsieh, I. Devetak, and A. Winter, "Entanglement-assisted capacity of quantum multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3078–3090, Jul. 2008.
[60] A. Winter, "Coding theorem and strong converse for quantum channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2481–2485, Nov. 1999.
[61] T. Ogawa and H. Nagaoka, "Making good codes for classical-quantum channel coding via quantum hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2261–2266, Jun. 2007.
[62] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, no. 4, pp. 289–293, Oct. 1958.
[63] N. Ramakrishnan, R. Iten, V. Scholz, and M. Berta, "Quantum blahut-arimoto algorithms," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 1909–1914.
[64] J. Körner, "The concept of single-letterization in information theory," in *Open Problems in Communication and Computation*. New York, NY, USA: Springer, 1987, pp. 35–36.
[65] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem," *IEEE Trans. Inf. Theory*, vol. 48, no. 10, pp. 2637–2655, Oct. 2002.
[66] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*. Cambridge, U.K.: Cambridge Univ. Press, 2002.
[67] M. Fukuda and M. M. Wolf, "Simplifying additivity problems using direct sum constructions," *J. Math. Phys.*, vol. 48, no. 7, Jul. 2007, Art. no. 072101.
[68] B. Schumacher and M. A. Nielsen, "Quantum data processing and error correction," *Phys. Rev. A, Gen. Phys.*, vol. 54, no. 4, p. 2629, 1996.
[69] H. G. Eggleston, "Convexity," *J. London Math. Soc.*, vol. 1, no. 1, pp. 183–186, 1966.
[70] J. Yard, P. Hayden, and I. Devetak, "Capacity theorems for quantum multiple-access channels: Classical-quantum and quantum-quantum capacity regions," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3091–3113, Jul. 2008.
[71] A. El Gamal and Y. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[72] S. Gel'fand and M. Pinsker, "Coding for channel with random parameters," *Problems Control Inf. Theory*, vol. 9, no. 1, pp. 19–31, Jan. 1980.
[73] C. H. Bennett, D. P. DiVincenzo, and J. A. Smolin, "Capacities of quantum erasure channels," *Phys. Rev. Lett.*, vol. 78, no. 16, p. 3217, Apr. 1997.
[74] M. H. M. Costa, "Writing on dirty paper (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 5, pp. 439–441, May 1983.
[75] T. Philosof, U. Erez, and R. Zamir, "Combined shaping and precoding for interference cancellation at low SNR," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2003, p. 68.
[76] S. T. Brink and U. Erez, "A close-to-capacity dirty paper coding scheme," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3417–3432, Oct. 2005.
[77] J. Sima and W. Chen, "Polar codes for broadcast channels with receiver message side information and noncausal state available at the encoder," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 993–997.
[78] B. Beilin and D. Burshtein, "On polar coding for side information channels," *IEEE Trans. Inf. Theory*, vol. 67, no. 2, pp. 673–685, Feb. 2021.
[79] M. Y. Sener, R. Bohnke, W. Xu, and G. Kramer, "Dirty paper coding based on polar codes and probabilistic shaping," *IEEE Commun. Lett.*, early access, Sep. 17, 2021, doi: 10.1109/LCOMM.2021.3113722.
[80] Q. Wang and C. He, "Practical dirty paper coding with nested binary LDGM-LDPC codes," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2009, pp. 1–6.
[81] K. M. Rege, K. Balachandran, J. H. Kang, and M. K. Karakayali, "Practical dirty paper coding with sum codes," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 441–455, Feb. 2016.
[82] G. Böcherer, D. Lentner, A. Cirino, and F. Steiner, "Probabilistic parity shaping for linear codes," 2019, *arXiv:1902.10648*.
[83] S. L. Braunstein and P. van Loock, "Quantum information with continuous variables," *Rev. Mod. Phys.*, vol. 77, no. 2, pp. 513–577, Jun. 2005.
[84] S. Guha, "Multiple-user quantum information theory for optical communication channels," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep., 2008.
[85] S. Guha, J. H. Shapiro, and B. I. Erkmen, "Classical capacity of bosonic broadcast communication and a minimum output entropy conjecture," *Phys. Rev. A, Gen. Phys.*, vol. 76, no. 3, Sep. 2007, Art. no. 032303.
[86] V. Giovannetti, R. García-Patrón, N. J. Cerf, and A. S. Holevo, "Ultimate classical communication rates of quantum optical channels," *Nature Photon.*, vol. 8, pp. 796–800, Sep. 2014.
[87] M. Horodecki, J. Oppenheim, and A. Winter, "Partial quantum information," *Nature*, vol. 436, pp. 673–676, Aug. 2005.
[88] H. Weingarten, Y. Steinberg, and S. S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.
[89] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[90] B. Schumacher, "Quantum coding," *Phys. Rev. A, Gen. Phys.*, vol. 51, no. 4, p. 2738, 1995.

Uzi Pereg (Member, IEEE) received the B.Sc. *(summa cum laude)* degree in electrical engineering from the Azrieli College of Engineering, Jerusalem, Israel, in 2011, and the M.Sc. and Ph.D. degrees from the Technion—Israel Institute of Technology, Haifa, Israel, in 2015 and 2019, respectively. He is currently a Post-Doctoral Researcher with the Institute for Communications Engineering, Technical University of Munich, Munich, Germany. In 2020, he joined the Theory Group of the German Federal Government (BMBF) Project for the design and analysis of quantum communication and repeater systems. His research interests are in the areas of quantum communications, information theory, and coding theory. He was a recipient of the 2018 Pearl Award for outstanding research work in the field of communications, 2018 KLA-Tencor Award for Excellent Conference Paper, the (2018–2019) Viterbi Fellowship, and the (2020–2021) Israel CHE Fellowship for Quantum Science and Technology.