Louisiana State University

LSU Digital Commons

1-1-2014

Identifying the information gain of a quantum measurement

Mario Berta
Institute for Quantum Information and Matter

Joseph M. Renes
ETH Zürich

Mark M. Wilde
Louisiana State University

Follow this and additional works at: https://digitalcommons.lsu.edu/physics_astronomy_pubs

Recommended Citation

Identifying the Information Gain of a Quantum Measurement

Mario Berta[*] and Joseph M. Renes

Institute for Theoretical Physics, ETH Zurich, 8093 Zürich, Switzerland

Mark M. Wilde

School of Computer Science, McGill University, Montréal, Québec, Canada

We show that quantum-to-classical channels, i.e., quantum measurements, can be asymptotically simulated by an amount of classical communication equal to the quantum mutual information of the measurement, if sufficient shared randomness is available. This result generalizes Winter's measurement compression theorem for fixed independent and identically distributed inputs [Winter, CMP 244 (157), 2004] to arbitrary inputs, and more importantly, it identifies the quantum mutual information of a measurement as the information gained by performing it, independent of the input state on which it is performed. Our result is a generalization of the classical reverse Shannon theorem to quantum-to-classical channels. In this sense, it can be seen as a quantum reverse Shannon theorem for quantum-to-classical channels, but with the entanglement assistance and quantum communication replaced by shared randomness and classical communication, respectively. The proof is based on a novel one-shot state merging protocol for "classically coherent states" as well as the post-selection technique for quantum channels, and it uses techniques developed for the quantum reverse Shannon theorem [Berta *et al.*, CMP 306 (579), 2011].

I. INTRODUCTION

Measurement is an integral part of quantum theory. It is the means by which we gather information about a quantum system. Although the classical notion of a measurement is rather straightforward, the quantum notion of measurement has been the subject of much thought and debate [1]. One interpretation is that the act of measurement on a quantum system causes it to abruptly jump or "collapse" into one of several possible states with some probability, an evolution seemingly different from the smooth, unitary transitions resulting from Schrödinger's wave equation. Some have advocated for a measurement postulate in quantum theory [20], while others have advocated that our understanding of quantum measurement should follow from other postulates [62].

---

[*] berta@phys.ethz.ch

In spite of the aforementioned difficulties in understanding and interpreting quantum measurement, there is a precise question that one can formulate concerning it:

How much information is gained by performing a given quantum measurement?

This question has a rather long history, which to our knowledge begins with the work of Groenewold [24]. In 1971, Groenewold argued on intuitive grounds for the following "entropy reduction" to quantify the information gained by performing a quantum measurement:

$$H(\rho) - \sum_x p_x H(\rho_x), \tag{1}$$

where ρ is the initial state before the measurement occurs, $\{p_x, \rho_x\}$ is the post-measurement ensemble induced by the measurement, and $H(\sigma) \equiv -\text{tr}[\sigma \log \sigma]$ is the von Neumann entropy of a state σ. The intuition behind this measure is that it quantifies the reduction in uncertainty after performing a quantum measurement on a quantum system in state ρ, and its form is certainly reminiscent of a Holevo-like quantity [26], although the classical data in the above Groenewold quantity appears at the *output* of the process rather than at the *input* as in the case of the Holevo quantity. Groenewold left open the question of whether this quantity is non-negative for all measurements, and Lindblad proved that non-negativity holds whenever the measurement is of the von Neumann-Lüders kind (projecting onto an eigenspace of an observable) [38]. Ozawa then settled the matter by proving that the above quantity is non-negative if and only if the post-measurement states are of the form

$$\rho_x = \frac{M_x \rho M_x^\dagger}{\text{tr}[M_x^\dagger M_x \rho]}, \tag{2}$$

for some operators $\{M_x\}$ such that $\sum_x M_x^\dagger M_x = \mathbb{1}$ [43]. Such measurements are termed "efficient", and differ from general measurements as the latter may have several operators $M_{x,s}$ corresponding to the result x [23].

The fact that the quantity in (1) can become negative for some quantum measurements excludes it from being a generally appealing measure of information gain. To remedy this situation, Buscemi *et al.* later advocated for the following measure to characterize the information gain of a quantum measurement when acting upon a particular state ρ [9, 39, 50, 57]:

$$I(X : R)_\omega, \tag{3}$$

where $I(X : R)_\omega \equiv H(X)_\omega + H(R)_\omega - H(XR)_\omega$ is the quantum mutual information of the following state:

$$\omega_{XR} \equiv \sum_x |x\rangle\langle x|_X \otimes \text{tr}_A\{(\mathcal{M}_x \otimes \mathcal{I}_R)(|\rho\rangle\langle\rho|_{AR})\}. \tag{4}$$

The register X is a classical register containing the outcome of the measurement, $\mathcal{M} \equiv \{\mathcal{M}_x\}$ is a collection of completely positive, trace non-increasing maps characterizing the measurement (for which the sum map $\sum_x \mathcal{M}_x$ is trace preserving), $\mathcal{I}$ is the identity map, and $|\rho\rangle_{AR}$ is a purification of the initial state ρ on system A to a purifying system R. The advantages of the measure of information gain in (3) are as follows:

- It is non-negative.

- It reduces to Groenewold's quantity in (1) for the special case of measurements of the form in (2) [9].

- It characterizes the trade-off between information and disturbance in quantum measurements [9].

- It has an operational interpretation in Winter's measurement compression protocol as the optimal rate at which a measurement gathers information [60].

This last advantage is the most compelling one from the perspective of quantum information theory—one cannot really justify a measure as an information measure unless it corresponds to a meaningful information processing task. Indeed, when reading the first few paragraphs of Groenewold's paper [24], it becomes evident that his original motivation was information theoretic in nature, and with this in mind, Winter's measure in (3) is clearly the one Groenewold was seeking after all.

In spite of the above arguments in favor of the information measure in (3) as a measure of information gain, it is still lacking in one aspect: it is dependent on the state on which the quantum measurement $\mathcal{M}$ acts in addition to the measurement itself. A final requirement that one should impose for a measure of information gain by a measurement is that it should depend only on the measurement itself. A simple way to remedy this problem is to maximize the quantity in (3) over all possible input states, leading to the following characterization of information gain:

$$I(\mathcal{M}) \equiv \max_{\rho_{AR}} I(X:R)_\omega \,, \tag{5}$$

for ω_{RX} as in (4). The quantity above has already been identified and studied by previous authors as an important information quantity, being labeled as the "purification capacity" of a measurement [33, 34] or the "information capacity of a quantum observable" [29]. The above quantity also admits an operational interpretation as the entanglement-assisted capacity of a quantum measurement for transmitting classical information [3, 28, 29], though it is our opinion that this particular
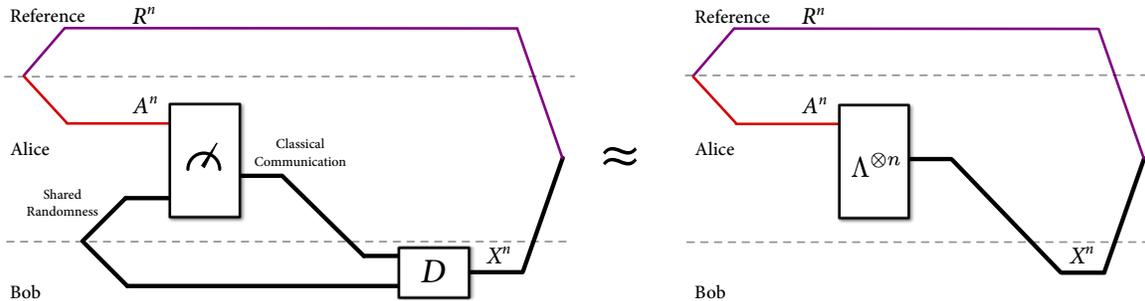
FIG. 1. Simulation (left) of the measurement $\Lambda^{\otimes n}$ (right). In the simulation, Alice uses shared randomness to perform a new measurement, whose result she communicates to Bob, such that Bob can recover the actual measurement output X^n using the message and the shared randomness. If the simulation scheme works for any input, we can associate the amount of communication with the information gained by the measurement.

operational interpretation is not sufficiently compelling such that we should associate the measure in (5) with the notion of information gain. The main aim of this paper is to address this issue by providing a compelling operational interpretation of the measure in (5).

II.   SUMMARY OF RESULTS

In this paper, our main contribution is to show that $I(\mathcal{M})$ is the optimal rate at which a measurement gains information when many identical instances of it act on an arbitrary input state. In our opinion, this new result establishes (5) as *the* information-theoretic measure of information gain of a quantum measurement. In more detail, let A denote the input Hilbert space for a given measurement $\mathcal{M}$. We suppose that a third party prepares an arbitrary quantum state on a Hilbert space $A^{\otimes n}$, which is equivalent to n identical copies of the original Hilbert space A, where n is a large positive number. A sender and receiver can then exploit some amount of shared random bits and classical communication to *simulate* the action of n instances of the measurement $\mathcal{M}$ (denoted by $\mathcal{M}^{\otimes n}$) on the chosen input state, in such a way that it becomes physically impossible for the third party, to whom the receiver passes along the measurement outcomes, to distinguish between the simulation and the ideal measurement $\mathcal{M}^{\otimes n}$ as n becomes large (the third party can even keep the purifying system of a purification of the chosen input state in order to help with the distinguishing task). By design, the information gained by the measurement is that relayed by the classical communication. Following [60], we call this task *universal measurement compression*. We prove that the optimal rate of classical communication is equal to $I(\mathcal{M})$, if sufficient shared randomness is available.

The information-theoretic task outlined above is also known as *channel simulation* (depicted in Figure 1), and it has been well studied for the case of fully classical channels (with classical inputs and classical outputs) [3, 13, 14] and fully quantum channels (with quantum inputs and quantum outputs) [2, 3, 7]. The "in-between" case of channels with quantum inputs and classical outputs (i.e., measurements) has been studied as well [60] (see also [57]), but as mentioned above, the problem of simulating many instances of a quantum measurement on an *arbitrary* input state has not been studied before this paper. Beyond its intrinsic interest as an information-processing task, channel simulation has two known concrete applications: in establishing a strong converse rate for a channel coding task [2–5] and in rate distortion coding (lossy data compression) [16–18, 59].

Our paper also features some related results of interest. We characterize the optimal rate region consisting of the rates of shared randomness and classical communication that are both necessary and sufficient for the existence of a measurement simulation, whenever both the sender and receiver are required to obtain the measurement outcomes (this is known as a feedback simulation since the sender also obtains the measurement outcomes). We also characterize the optimal rate region of shared randomness and classical communication for a non-feedback simulation, in which the sender is not required to obtain the measurement outcomes. Note that if sufficient shared randomness is available and we are only interested in quantifying the rate of classical communication, then there is no advantage of a non-feedback simulation over a feedback one—the optimal rate of classical communication is given by (5).

Our proof technique in this paper exploits ideas from the approach in [7] for proving the fully quantum reverse Shannon theorem. In fact, one can think of our approach here as a "classicalized" or "dephased" version of that approach. In particular, we begin by establishing a protocol known as "classically coherent state merging," which is a variation of the well-known state merging protocol [30, 31] specialized to classically coherent states (see Section III for definition). We then show how time-reversing this protocol and exchanging the roles of Alice and Bob leads to a protocol known as "classically coherent state splitting." It suffices for our purposes for this protocol to use shared randomness and classical communication rather than entanglement and quantum communication, respectively. Generalizing this last protocol then leads to a one-shot state-and-channel simulation which is essentially optimal when acting on a single copy of a *known* state. Finally, we exploit the post-selection technique for quantum channels [10] and the aforementioned state splitting protocol to show that it suffices to simulate many instances of a measurement on a purification of a particular de Finetti quantum input state in order to guarantee that the simulation is asymptotically perfect when acting on an *arbitrary* quantum state. We then show that applying very similar reasoning

as above along with randomness recycling [2] solves the non-feedback case.

We organize this paper as follows. In Section III, we introduce our notation and review preliminary concepts such as states, distance measures, channels, isometries, entropies, smooth entropies, and classically coherent states. Section IV then introduces one-shot protocols for state merging and state splitting of classically coherent states (the classical state splitting protocol turns out to be the most important tool for proving our main result). Section V provides a proof of our main results for the case of feedback and non-feedback simulations, and we shortly comment on possible extensions and applications in Section VI. We finally conclude in Section VII by summarizing our results and stating some directions for future research.

III. PRELIMINARIES

States, Distance Measures, Channels, Isometries. Let $A, B, C, \ldots$ denote finite dimensional Hilbert spaces and let $|A|$ denote the dimension of A. We establish notation for several sets: $\mathcal{L}(A)$ linear operators on A, $\mathcal{P}(A)$ non-negative linear operators on A, $\mathcal{S}_{\leq}(A) = \{\rho_A \in \mathcal{P}(A) \,|\, \mathrm{tr}[\rho] \leq 1\}$ subnormalized states on A, $\mathcal{S}(A) = \{\rho_A \in \mathcal{P}(A) \,|\, \mathrm{tr}[\rho] = 1\}$ density operators or states on A, and $\mathcal{V}(A) = \{\rho_A \in \mathcal{S}(A) \,|\, \mathrm{tr}[\rho^2] = 1\}$ pure-state density operators on A. We define the purified distance $P(\rho_A, \sigma_A) = \sqrt{1 - \bar{F}^2(\rho_A, \sigma_A)}$ for $\rho_A, \sigma_A \in \mathcal{S}_{\leq}(A)$, where $\bar{F}(\rho_A, \sigma_A) = F(\rho_A, \sigma_A) + \sqrt{(1 - \mathrm{tr}[\rho_A])(1 - \mathrm{tr}[\sigma_A])}$, and the quantum fidelity $F(\rho_A, \sigma_A) = \|\sqrt{\rho_A}\sqrt{\sigma_A}\|_1$ with $\|\Gamma_A\|_1 = \mathrm{tr}[\sqrt{\Gamma_A \Gamma_A^\dagger}]$ for $\Gamma_A \in \mathcal{L}(A)$. We use the notation $\rho_A \approx_\varepsilon \sigma_A$ to indicate that ρ_A and σ_A are ε-close with respect to the purified distance: $P(\rho_A, \sigma_A) \leq \varepsilon$. We define the ε-ball around ρ_A as $\mathcal{B}^\varepsilon(\rho_A) = \{\tilde{\rho}_A \in \mathcal{S}_{\leq}(A) : \tilde{\rho}_A \approx_\varepsilon \rho_A\}$. The tensor product of two Hilbert spaces A and B is denoted by $AB \equiv A \otimes B$. Given a multipartite operator $\rho_{AB} \in \mathcal{P}(AB)$, we unambiguously write $\rho_A = \mathrm{tr}_B[\rho_{AB}]$ for the corresponding reduced operator. For $M_A \in \mathcal{L}(A)$, we write $M_A \equiv M_A \otimes \mathbb{1}_B$ for the enlargement on any joint Hilbert space AB, where $\mathbb{1}_B$ denotes the identity operator acting on $\mathcal{L}(B)$. Isometries from A to B are denoted by $V_{A \to B}$. For Hilbert spaces A, B with orthonormal bases $\{|i\rangle_A\}_{i=1}^{|A|}$, $\{|i\rangle_B\}_{i=1}^{|B|}$ and $|A| = |B|$, the canonical identity mapping from $\mathcal{L}(A)$ to $\mathcal{L}(B)$ with respect to these bases is denoted by $\mathcal{I}_{A \to B}$, i.e., $\mathcal{I}_{A \to B}(|i\rangle\langle j|_A) = |i\rangle\langle j|_B$. A linear map $\mathcal{E}_{A \to B} : \mathcal{L}(A) \to \mathcal{L}(B)$ is positive if $\mathcal{E}_{A \to B}(\rho_A) \in \mathcal{P}(B)$ for all $\rho_A \in \mathcal{P}(A)$. It is completely positive if the map $(\mathcal{E}_{A \to B} \otimes \mathcal{I}_{C \to C})$ is positive for all C. Completely positive and trace preserving maps are called quantum channels. The support of $\rho_A \in \mathcal{P}(A)$ is denoted by $\mathrm{supp}(\rho_A)$, the projector onto $\mathrm{supp}(\rho_A)$ is denoted by ρ_A^0 and $\mathrm{tr}[\rho_A^0] = \mathrm{rank}(\rho_A)$, the rank of ρ_A. For $\rho_A \in \mathcal{P}(A)$ we write $\|\rho_A\|_\infty$ for the operator norm of ρ_A, which is equal to the maximum eigenvalue of ρ_A.

Diamond Norm. We will need a distance measure for quantum channels. We use a norm on the set of quantum channels which measures the bias in distinguishing two such mappings. In quantum information theory, this norm is known as the diamond norm [37]. Here, we present it in a formulation which highlights that it is dual to the well-known completely bounded (cb) norm [44].

Definition 1. *Let $\mathcal{E}_A : \mathcal{L}(A) \mapsto \mathcal{L}(B)$ be a linear map. The diamond norm of $\mathcal{E}_A$ is defined as*

$$\|\mathcal{E}_A\|_\diamond = \sup_{k \in \mathbb{N}} \|\mathcal{E}_A \otimes \mathcal{I}_k\|_1 \ , \tag{6}$$

where $\|\mathcal{F}_A\|_1 = \sup_{\sigma \in \mathcal{S}_{\leq}(A)} \|\mathcal{F}_A(\sigma_A)\|_1$ and $\mathcal{I}_k$ denotes the identity map on states of a k-dimensional quantum system.

The supremum in Definition 1 is reached for $k = |A|$ [37, 44]. Two quantum channels $\mathcal{E}$ and $\mathcal{F}$ are called ε-close if they are ε-close in the metric induced by the diamond norm.

Classically Coherent States. We say that a pure state $|\psi\rangle\langle\psi|_{X_A X_B R} \in \mathcal{V}(X_A X_B R)$ is *classically coherent* with respect to systems $X_A X_B$ if there is an orthonormal basis $\{|x\rangle\}$ such that $|\psi\rangle$ can be written in the following form:

$$|\psi\rangle_{X_A X_B R} = \sum_x \sqrt{p_x} |xx\rangle_{X_A X_B} \otimes |\psi_x\rangle_R, \tag{7}$$

for some probability distribution p_x and states $|\psi_x\rangle_R$. Harrow realized the importance of classically coherent states for quantum communication tasks [25], while Refs. [22, 51] recently exploited this notion in devising a "decoupling approach" to the Holevo-Schumacher-Westmoreland coding theorem [27, 49] that is useful for our purposes here. Classically coherent states are also related to Zurek's approach to decoherence [61], in which classicality arises from an inaccessible environment possessing an "imprint" of a classical state in superposition (as in the above state if we think of X_B as an environment).

Entropies. Recall the following standard definitions. The von Neumann entropy of $\rho_A \in \mathcal{S}(A)$ is defined as[1]

$$H(A)_\rho = -\text{tr}[\rho_A \log \rho_A] \ . \tag{8}$$

The quantum relative entropy of $\rho_A \in \mathcal{S}_{\leq}(A)$ with respect to $\sigma_A \in \mathcal{P}(A)$ is given by

$$D(\rho_A \| \sigma_A) = \text{tr}[\rho_A \log \rho_A] - \text{tr}[\rho_A \log \sigma_A], \tag{9}$$

---

[1] All logarithms in this paper are taken to base 2.

if $\mathrm{supp}(\rho_A) \subseteq \mathrm{supp}(\sigma_A)$ and ∞ otherwise. The conditional von Neumann entropy of A given B for $\rho_{AB} \in \mathcal{S}(AB)$ is defined as

$$H(A|B)_\rho = -D(\rho_{AB}\|\mathbb{1}_A \otimes \rho_B) \ . \tag{10}$$

The mutual information between A and B for $\rho_{AB} \in \mathcal{S}(AB)$ is given by

$$I(A:B)_\rho = D(\rho_{AB}\|\rho_A \otimes \rho_B) \ . \tag{11}$$

Note that we can also write

$$H(A|B)_\rho = -\inf_{\sigma_B \in \mathcal{S}(B)} D(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B), \tag{12}$$

$$I(A:B)_\rho = \inf_{\sigma_B \in \mathcal{S}(B)} D(\rho_{AB}\|\rho_A \otimes \sigma_B) \ . \tag{13}$$

Smooth Entropies. We now give the definitions of the smooth entropy measures that we need in this work. We define the max-relative entropy of $\rho_A \in \mathcal{S}_\leq(A)$ with respect to $\sigma_A \in \mathcal{P}(A)$ as [15]

$$D_{\max}(\rho_A\|\sigma_A) = \inf\{\lambda \in \mathbb{R} : 2^\lambda \cdot \sigma_A \geq \rho_A\} \ . \tag{14}$$

The conditional min-entropy of A given B for $\rho_{AB} \in \mathcal{S}_\leq(AB)$ is defined as

$$H_{\min}(A|B)_\rho = -\inf_{\sigma_B \in \mathcal{S}(B)} D_{\max}(\rho_{AB}\|\mathbb{1}_A \otimes \sigma_B) \ . \tag{15}$$

In the special case where B is trivial, we get $H_{\min}(A)_\rho = -\log\|\rho_A\|_\infty$. The max-information that B has about A for $\rho_{AB} \in \mathcal{S}_\leq(AB)$ is defined as [7]

$$I_{\max}(A:B)_\rho = \inf_{\sigma_B \in \mathcal{S}(B)} D_{\max}(\rho_{AB}\|\rho_A \otimes \sigma_B) \ . \tag{16}$$

Note that, unlike the mutual information, the max-information is not symmetric in its arguments.[2]

Smooth entropy measures are defined by extremizing the non-smooth measures over a set of nearby states, where our notion of "nearby" is expressed in terms of the purified distance. The smooth max-information that B has about A for $\rho_{AB} \in \mathcal{S}_\leq(AB)$ is defined as

$$I_{\max}^\varepsilon(A:B)_\rho = \inf_{\bar{\rho}_{AB} \in \mathcal{B}^\varepsilon(\rho_{AB})} I_{\max}(A:B)_{\bar{\rho}} \ . \tag{17}$$

In contrast to the non-smooth case, the smooth max-information is approximately symmetric in its arguments.

---

[2] For a further discussion of max-based measures for *mutual information*, see [11].

Lemma 1. *[11, Corollary 4.2.4] Let $\varepsilon \geq 0$, $\varepsilon' > 0$, and $\rho_{AB} \in \mathcal{S}(AB)$. Then, we have that*

$$I_{\max}^{\varepsilon+2\varepsilon'}(B:A)_\rho \leq I_{\max}^\varepsilon(A:B)_\rho + \log\left(\frac{2}{(\varepsilon')^2} + 2\right), \tag{18}$$

and the same holds for A and B interchanged.

For technical reasons, we also need the following entropic quantities. For $\varepsilon \geq 0$, and $\rho_A \in \mathcal{S}_{\leq}(A)$, the max-entropy and its smooth version are defined as

$$H_{\max}(A)_\rho = 2 \log \text{tr}\left[\rho_A^{1/2}\right], \tag{19}$$

$$H_{\max}^\varepsilon(A)_\rho = \inf_{\bar{\rho}_A \in \mathcal{B}^\varepsilon(\rho_A)} H_{\max}(A)_{\bar{\rho}}. \tag{20}$$

Furthermore, the zero-Rényi entropy and its smooth version are defined as

$$H_0(A)_\rho = \log \text{rank}(\rho_A), \tag{21}$$

$$H_0^\varepsilon(A)_\rho = \inf_{\bar{\rho}_A \in \mathcal{B}^\varepsilon(\rho_A)} H_0(A)_{\bar{\rho}}. \tag{22}$$

Since all Hilbert spaces in this paper are assumed to be finite dimensional and the ball $\mathcal{B}^\varepsilon$ is convex and compact [52], we can replace the infima by minima and the suprema by maxima in all the definitions of this section. We will do so in what follows.

IV. CLASSICALLY COHERENT STATE MERGING AND STATE SPLITTING

We first establish "one-shot" protocols for state merging and state splitting of classically coherent quantum states. The classical state splitting protocol established in this section will then be the basis for the universal measurement compression protocol discussed in the next section.

Definition 2 (State Merging for Classically Coherent States)**.** *Consider a bipartite system with parties Alice and Bob. Let $\varepsilon > 0$, and $\rho_{X_A X_B B R} \in \mathcal{V}(X_A X_B B R)$ be classically coherent on $X_A X_B$ with respect to the basis $\{|x\rangle\}$, where Alice controls X_A, Bob $X_B B$, and R is a reference system. A quantum protocol $\mathcal{E}$ is called an ε-error state merging of $\rho_{X_A X_B B R}$ if it consists of applying local operations at Alice's side, sending q qubits from Alice to Bob, local operations at Bob's side, and it outputs a state $\omega_{X_{B'} X_B B R X_{A_1} B_1} = (\mathcal{E} \otimes \mathcal{I}_R)(\rho_{X_A X_B B R})$ such that*

$$\omega_{X_{B'} X_B B R X_{A_1} B_1} \approx_\varepsilon \mathcal{I}_{X_A \to X_{B'}}(\rho_{X_A X_B B R}) \otimes \phi_{X_{A_1} B_1}^E, \tag{23}$$

where $\phi_{X_{A_1} B_1}^E$ is a maximally entangled state of Schmidt rank E. The quantity q is called the quantum communication cost, and $e = \lfloor \log E \rfloor$ the entanglement gain.
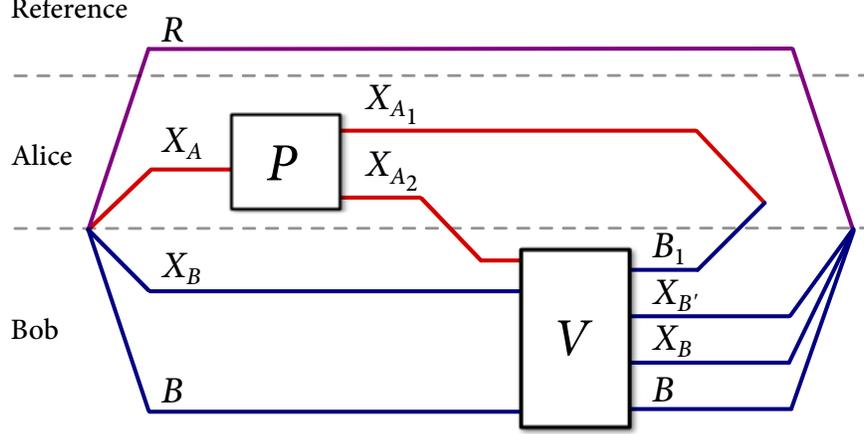
FIG. 2. The protocol from the proof of Lemma 2 for state merging of a classically coherent state on systems RX_AX_BB. The operation P is a permutation of states in the orthonormal basis $\{|x\rangle\}$ of X_A, and it also splits X_A into two subsystems. The operation V is an isometry guaranteed by Uhlmann's theorem to complete the merging task, while also generating entanglement between Alice and Bob.

Lemma 2. *Let $\varepsilon > 0$, and $\rho_{X_AX_BBR} \in \mathcal{V}(X_AX_BBR)$ be classically coherent on X_AX_B with respect to the basis $\{|x\rangle\}$. Then there exists an ε-error state merging protocol for $\rho_{X_AX_BBR}$ with quantum communication cost*

$$q = \left\lceil H_0(X_A)_\rho - H_{\min}(X_A|R)_\rho + 4 \cdot \log\frac{1}{\varepsilon} \right\rceil, \tag{24}$$

and entanglement gain

$$e = \left\lfloor H_{\min}(X_A|R)_\rho - 4 \cdot \log\frac{1}{\varepsilon} \right\rfloor . \tag{25}$$

Proof. The intuition is as follows. First Alice applies a particular permutation $P_{X_A \to X_{A_1}X_{A_2}}$ in the basis $\{|x\rangle\}_{x \in X_A}$; it also splits the output into two subsystems X_{A_1} and X_{A_2}. Then she sends X_{A_2} to Bob, who finally performs a local isometry $V_{X_{A_2}X_BB \to X_{B'}X_BBB_1}$. After Alice applies the permutation, the state on $X_{A_1}R$ is approximately given by $\frac{\mathbb{1}_{X_{A_1}}}{|X_{A_1}|} \otimes \rho_R$ and Bob holds a purification of this. But $\frac{\mathbb{1}_{X_{A_1}}}{|X_{A_1}|} \otimes \rho_R$ is the reduced state of $\rho_{X_{B'}X_BBR} \otimes \phi^E_{X_{A_1}B_1}$, and since all purifications are equivalent up to local isometries, there exists an isometry $V_{X_{A_2}X_BB \to X_{B'}X_BBB_1}$ on Bob's side that transforms the state into $\rho_{X_{B'}X_BBR} \otimes \phi^E_{X_{A_1}B_1}$. Figure 2 depicts this protocol.

More formally, let $X_A = X_{A_1}X_{A_2}$ with $\log|X_{A_2}| = \lceil \log|X_A| - H_{\min}(X_A|R)_\rho + 4 \cdot \log\frac{1}{\varepsilon} \rceil$. According to Proposition 27 concerning permutation based extractors, there exists a permutation

$P_{X_A \to X_{A_1} X_{A_2}}$ such that for $\sigma_{X_{A_1} X_{A_2} BR} = P_{X_A \to X_{A_1} X_{A_2}}(\rho_{X_A X_B BR})$,

$$\left\| \sigma_{X_{A_1} R} - \frac{\mathbb{1}_{X_{A_1}}}{|X_{A_1}|} \otimes \rho_R \right\|_1 \leq \varepsilon^2 \ . \tag{26}$$

By an upper bound of the purified distance in terms of the trace distance (Lemma 24), this implies $\sigma_{X_{A_1} R} \approx_\varepsilon \frac{\mathbb{1}_{X_{A_1}}}{|X_{A_1}|} \otimes \rho_R$. Alice applies this permutation $P_{X_A \to X_{A_1} X_{A_2}}$ and then sends X_{A_2} to Bob; therefore

$$q = \left\lceil \log |X_A| - H_{\min}(X_A | R)_\rho + 4 \cdot \log \frac{1}{\varepsilon} \right\rceil \ . \tag{27}$$

Uhlmann's theorem [36, 55] guarantees that there exists an isometry $V_{X_{A_2} X_B B \to X_{B'} X_B BB_1}$ such that

$$P\left(\sigma_{X_{A_1} R}, \frac{\mathbb{1}_{X_{A_1}}}{|X_{A_1}|} \otimes \rho_R \right) = P\left(V_{X_{A_2} X_B B \to X_{B'} X_B BB_1}(\sigma_{X_{A_1} X_{A_2} X_B BR}), \phi^E_{X_{A_1} B_1} \otimes \rho_{X_{B'} X_B BR} \right) \ . \tag{28}$$

Hence the entanglement gain is given by

$$e = \left\lfloor H_{\min}(X_A | R)_\rho - 4 \cdot \log \frac{1}{\varepsilon} \right\rfloor \ . \tag{29}$$

Now if ρ_{X_A} has full rank, this is already what we want. In general $\log \operatorname{tr}\left[\rho^0_{X_A} \right] = \log |X_{\hat{A}}| \leq \log |X_A|$. But in this case we can restrict X_A to the subspace $X_{\hat{A}}$ on which ρ_{X_A} has full rank, i.e. those x for which $p_x \neq 0$. $\qquad \square$

Definition 3 (State Splitting for Classically Coherent States). *Consider a bipartite scenario with parties Alice and Bob. Let $\varepsilon > 0$, and $\rho_{AX_A X_{A'} R} \in \mathcal{V}(AX_A X_{A'} R)$ be classically coherent on $X_A X_{A'}$ with respect to the basis $\{|x\rangle\}$, where Alice controls $AX_A X_{A'}$, and R is a reference system. Furthermore let $\phi^E_{A_1 B_1}$ be a maximally entangled state of Schmidt rank E shared between Alice and Bob. A quantum protocol $\mathcal{E}$ is called an ε-error state splitting of $\rho_{AX_A X_{A'} R}$ if it consists of applying local operations at Alice's side, sending q qubits from Alice to Bob, local operations at Bob's side, and it outputs a state $\omega_{AX_A X_B R} = (\mathcal{E} \otimes \mathcal{I}_R)(\rho_{AX_A X_{A'} R} \otimes \phi^E_{A_1 B_1})$ such that*

$$\omega_{AX_A X_B R} \approx_\varepsilon \mathcal{I}_{X_{A'} \to X_B}(\rho_{AX_A X_{A'} R}) \ . \tag{30}$$

The quantity q is called the quantum communication cost, and $e = \lceil \log E \rceil$ the entanglement cost.

Lemma 3. *Let $\varepsilon > 0$, and $\rho_{AX_A X_{A'} R} \in \mathcal{V}(AX_A X_{A'} R)$ be classically coherent on $X_A X_{A'}$ with respect to the basis $\{|x\rangle\}$. Then there exists an ε-error state splitting protocol for $\rho_{AX_A X_{A'} R}$ with quantum communication cost*

$$q = \left\lceil H_0(X_{A'})_\rho - H_{\min}(X_{A'} | R)_\rho + 4 \cdot \log \frac{1}{\varepsilon} \right\rceil, \tag{31}$$
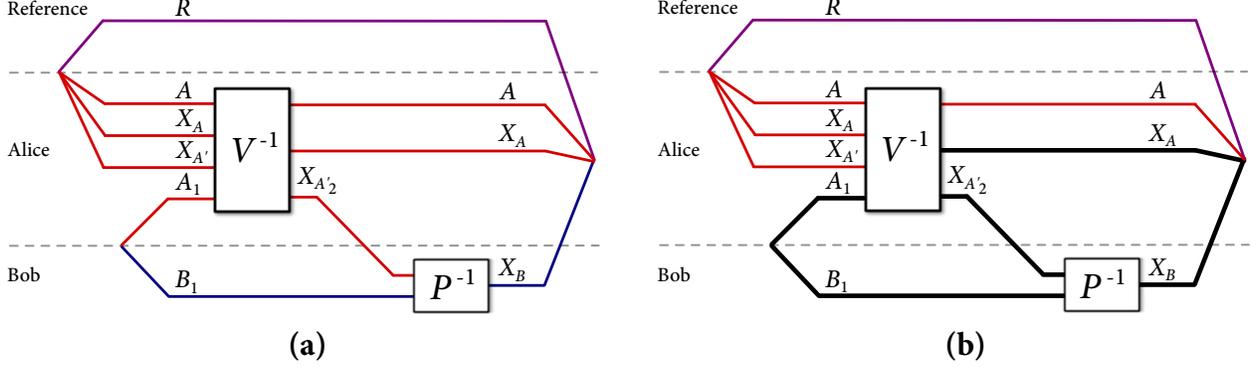
FIG. 3. **(a)** A simple protocol for state splitting obtained by time-reversing the state merging protocol of Lemma 2 and interchanging the roles of Alice and Bob. **(b)** If it is not necessary to maintain the quantum coherence of the X systems (if they can be dephased to classical registers), then the state splitting protocol can exploit shared randomness and classical communication instead of entanglement and quantum communication, respectively.

and entanglement cost

$$e = \left\lfloor H_{\min}(X_{A'}|R)_\rho - 4 \cdot \log \frac{1}{\varepsilon} \right\rfloor . \tag{32}$$

Proof. We get the desired state splitting protocol by time-reversing the state merging protocol of Lemma 2 and interchanging the roles of Alice and Bob. Figure 3(a) depicts the state splitting protocol for classically coherent states. More precisely, we first define an isometry $V_{X_{A'_2} X_A A \to X_{A'} X_A A A_1}$, analogously to $V_{X_{A_2} X_B B \to X_{B'} X_B B B_1}$ of (28) in the state merging protocol. Because all isometries are injective, we can define an inverse of V acting on the image of V (which we denote by $\mathrm{Im}(V)$). The inverse is again an isometry and we denote it by $V^{-1}_{\mathrm{Im}(V) \to X_{A'_2} X_A A}$. The protocol starts by measuring the $A X_A X_{A'} A_1$ systems to decide whether $\rho_{A X_A X_{A'}} \otimes \phi^E_{A_1} \in \mathrm{Im}(V)$ or not. If so, the protocol proceeds by applying the isometry $V^{-1}_{\mathrm{Im}(W) \to X_{A'_2} X_A A}$, but otherwise the state is discarded and replaced with $|0\rangle\langle 0|_{X_{A'_2} X_A A}$. This step is necessary because the output of merging is not exactly $\rho_{A X_A X_{A'} R}$. The next step is to send $X_{A'_2}$ to Bob, who then applies the permutation $P^{-1}_{X_{A'_2} B_1 \to X_B}$ defined analogously to $P_{X_A \to X_{A_1} X_{A_2}}$ in (26). By the monotonicity of the purified distance, we get a state that is ε-close to $\mathcal{I}_{X_{A'} \to X_B}(\rho_{A X_A X_{A'} R})$. $\qquad \square$

If we are not concerned with the coherence of the registers X_A and X_B shared between Alice and Bob, then the protocol given above (Lemma 3) also works if the entanglement assistance and the quantum communication are replaced by the same amount of shared randomness assistance and classical communication, respectively. More precisely, we define:

Definition 4 (Classical State Splitting of Classically Coherent States). *Consider a bipartite system with parties Alice and Bob. Let $\varepsilon > 0$, and $\rho_{AX_AX_{A'}R} \in \mathcal{V}(AX_AX_{A'}R)$ be classically coherent on $X_AX_{A'}$ with respect to the basis $\{|x\rangle\}$, where Alice controls $AX_AX_{A'}$, and R is a reference system. Furthermore let $\overline{\phi}^S_{X_{A_1}X_{B_1}}$ denote S bits of shared randomness shared between Alice and Bob. A quantum protocol $\mathcal{E}$ is called an ε-error classical state splitting of $\rho_{AX_AX_{A'}R}$ if it consists of applying local operations at Alice's side, sending c bits from Alice to Bob, local operations at Bob's side, and it outputs a state $\omega_{AX_AX_BR} = (\mathcal{E} \otimes \mathcal{I}_R)(\rho_{AX_AX_{A'}R} \otimes \overline{\phi}^S_{X_{A_1}X_{B_1}})$ such that*

$$\omega_{AX_AX_BR} \approx_\varepsilon \sum_x \langle x|\rho_{AX_AX_{A'}R}|x\rangle_{X_{A'}} \otimes |x\rangle\langle x|_{X_B} . \tag{33}$$

The quantity c is called the classical communication cost, and $s = \lceil \log S \rceil$ shared randomness cost.

Using the achievability of state splitting of classically coherent states (Lemma 3) we get the following.

Corollary 4. *Let $\varepsilon > 0$, and $\rho_{AX_AX_{A'}R} \in \mathcal{V}(AX_AX_{A'}R)$ be classically coherent on $X_AX_{A'}$ with respect to the basis $\{|x\rangle\}$. Then there exists a classical ε-error state splitting protocol for $\rho_{AX_AX_{A'}R}$ with classical communication cost*

$$c = \left\lceil H_0(X_{A'})_\rho - H_{\min}(X_{A'}|R)_\rho + 4 \cdot \log \frac{1}{\varepsilon} \right\rceil \tag{34}$$

and shared randomness cost

$$s = \left\lfloor H_{\min}(X_{A'}|R)_\rho - 4 \cdot \log \frac{1}{\varepsilon} \right\rfloor . \tag{35}$$

Proof. Note that it is sufficient to find a protocol for state splitting of classically coherent states (as in Definition 3) that only works up to random phase flips on the X_B register. These random phase flips then commute with the action of the permutation that takes systems B_1 and $X_{A'_2}$ to X_B. Thus, if we use the protocol for state splitting of classically coherent states described before (Lemma 3), random phase flips on X_B are the same as random phase flips on $X_{A'_2}B_1$ before the permutation $P^{-1}_{X_{A'_2}B_1 \to X_B}$ is applied. Since random phase flips on B_1 just transform the maximally entangled state $\phi_{A_1B_1}$ to shared randomness $\overline{\phi}_{X_{A_1}X_{B_1}}$ of the same size (with the relabeling of A_1B_1 to $X_{A_1}X_{B_1}$), and they dephase the quantum system $X_{A'_2}$ to a classical system, the protocol of Lemma 3 also works for classical state splitting of classically coherent states. □

Note that the above idea is similar to how Hsieh *et al.* recovered the Holevo-Schumacher-Westmoreland coding theorem for classical communication from a protocol for entanglement-assisted classical communication [32], simply by dephasing shared entanglement to common randomness and replacing random unitaries with random permutations.

However, the classical communication cost of this protocol is not yet optimal (for the general one-shot case considered here). To improve this, we use an idea from a recent proof of the quantum reverse Shannon theorem, and Theorem 6 demonstrates that the rate found in terms of the smooth max-information is essentially optimal. The following lemma is the crucial ingredient for the proof of our main result: universal measurement compression (Theorem 7).

Theorem 5. *Let $\varepsilon > 0$, $\varepsilon' \geq 0$, and $\rho_{AX_A X_{A'} R} \in \mathcal{V}(AX_A X_{A'} R)$ be classically coherent on $X_A X_{A'}$ with respect to the basis $\{|x\rangle\}$. Then there exists a classical $(\varepsilon + \varepsilon' + \sqrt{8\varepsilon'} + |X_{A'}|^{-1/2})$-error state splitting protocol for $\rho_{AX_A X_{A'} R}$ with*

$$c \leq I^{\varepsilon'}_{\max}(X_{A'} : R)_\rho + 4 \cdot \log \frac{1}{\varepsilon} + 4 + \log \log |X_{A'}| \tag{36}$$

$$c + s \leq H^{\varepsilon'}_0(X_{A'})_\rho + 2 + \log \log |X_{A'}| \, , \tag{37}$$

where c denotes the classical communication cost, and s the shared randomness cost.

Proof. The idea for the protocol is as follows. Let $\rho_{AX_A X_{A'} R} = |\rho\rangle\langle\rho|_{AX_A X_{A'} R}$ with

$$|\rho\rangle_{AX_A X_{A'} R} = \sum_x \sqrt{p_x} \cdot |xx\rangle_{X_A X_{A'}} \otimes |\rho^x\rangle_{AR} \, . \tag{38}$$

First, in our proof, we disregard all the x with $p_x \leq |X_{A'}|^{-2}$. This introduces an error $|X_{A'}|^{-1/2}$, but the error at the end of the protocol is still upper bounded by $|X_{A'}|^{-1/2}$ due to the monotonicity of the purified distance. As the next step, we let Alice perform a measurement $W_{X_{A'} \rightarrow X_{A'} Y_A}$ with roughly $2 \cdot \log |X_{A'}|$ measurement outcomes in the basis $\{|x\rangle\}_{x \in X_{A'}}$. That is, the state after the measurement is of the form

$$\omega_{AX_A X_{A'} R Y_A} = \sum_y q_y \cdot \rho^y_{AX_A X_{A'} R} \otimes |y\rangle\langle y|_{Y_A} \, , \tag{39}$$

where the index y indicates which measurement outcome occurs, q_y denotes its probability, and $\rho^y_{AX_A X_{A'} R}$ is the corresponding post-measurement state. Then conditioned on the index y, we use the classical state splitting protocol for classically coherent states from Lemma 4 for each state $\rho^y_{AX_A X_{A'} R}$, and denote the corresponding classical communication cost and shared randomness cost by c_y and s_y, respectively. The total amount of classical communication we need for this is no larger than $\max_y c_y$, plus the amount needed to send the register Y_A (which is of order $\log \log |X_{A'}|$). The sum cost is no larger than $\max_y c_y + s_y$ (along with the amount for sending Y_A). This completes the description of the classical state splitting protocol for $\rho_{AX_A X_{A'} R}$. All that remains to do is to bring the expression for the classical communication cost and the sum cost into the right form. In the following, we describe the proof in detail.

Let $Q = \lceil 2 \cdot \log |X_{A'}| - 1 \rceil$, $Y = \{0, 1, \ldots, Q, (Q+1)\}$ and let $\{T^y_{X_{A'}}\}_{y \in Y}$ be a collection of projectors on $X_{A'}$ defined as

$$T^{Q+1}_{X_{A'}} = \sum_{\substack{x \\ 0 \le p_x \le 2^{-2 \log |X_{A'}|}}} |x\rangle\langle x|_{X_{A'}} , \qquad T^Q_{X_{A'}} = \sum_{\substack{x \\ 2^{-2 \log |X_{A'}|} \le p_x \le 2^{-Q}}} |x\rangle\langle x|_{X_{A'}} , \tag{40}$$

and for $y = 0, 1, \ldots, (Q-1)$ as

$$T^y_{X_{A'}} = \sum_{\substack{x \\ 2^{-(y+1)} \le p_x \le 2^{-y}}} |x\rangle\langle x|_{X_{A'}} . \tag{41}$$

These define a measurement

$$W_{X_{A'} \to X_{A'} Y_A}(\cdot) = \sum_{y \in Y} T^y_{X_{A'}}(\cdot) T^y_{X_{A'}} \otimes |y\rangle\langle y|_{Y_A} , \tag{42}$$

where the vectors $|y\rangle_{Y_A}$ form an orthonormal basis, and Y_A is at Alice's side. Furthermore let

$$q_y = \mathrm{tr}\left[T^y_{X_{A'}} \rho_{X_{A'}} \right], \tag{43}$$

$$\rho^y_{AX_A X_{A'} R} = q_y^{-1} \cdot T^y_{X_{A'}} \rho_{AX_A X_{A'} R} T^y_{X_{A'}}, \tag{44}$$

and define the sub-normalized state

$$\bar{\rho}_{AX_A X_{A'} R} = \sum_{y=0}^{Q} q_y \cdot \rho^y_{AX_A X_{A'} R} . \tag{45}$$

We have

$$P(\bar{\rho}_{AX_A X_{A'} R}, \rho_{AX_A X_{A'} R}) = \sqrt{1 - F^2(\bar{\rho}_{AX_A X_{A'} R}, \rho_{AX_A X_{A'} R})} \tag{46}$$

$$\le \sqrt{1 - \sum_{y=0}^{Q} q_y} = \sqrt{q_{Q+1}} \le \sqrt{|X_{A'}| \cdot 2^{-2 \log |X_{A'}|}} = |X_{A'}|^{-1/2} . \tag{47}$$

We proceed by defining the operations that we need for the classical state splitting protocol for $\bar{\rho}_{AX_A X_{A'} R}$. We want to use the ε-error classical state splitting protocol from Corollary 4 for each $\rho^y_{AX_A X_{A'} R}$. For $y = 0, 1, \ldots, Q$ this protocol has a classical communication cost

$$c_y \le H_0(X_{A'})_{\rho^y} - H_{\min}(X_{A'}|R)_{\rho^y} + 4 \cdot \log \frac{1}{\varepsilon} + 1 , \tag{48}$$

and sum cost

$$c_y + s_y \le H_0(X_{A'})_{\rho^y} , \tag{49}$$

where s_y denotes the shared randomness cost.

FIG. 4. Our final one-shot protocol for state splitting that achieves the smooth max-information rate of Theorem 5. The converse theorem in Theorem 6 guarantees that this one-shot protocol is essentially optimal in terms of its classical communication cost.

For X_{A_1} on Alice's side, X_{B_1} on Bob's side, and $X_{A_1^y}$, $X_{B_1^y}$ 2^{s_y}-dimensional subspaces of X_{A_1}, X_{B_1} respectively, the classical state splitting protocol from Corollary 4 has basically the following form: apply some isometry $V_{AX_{A'}X_AX_{A_1^y} \to AX_{(A_2')^y}X_A}$ on Alice's side, send $X_{(A_2')^y}$ from Alice to Bob (relabel it to $X_{B_2^y}$), and then apply some isometry $U_{X_{B_1^y}X_{B_2^y} \to B}$ on Bob's side ($U_{X_{B_1^y}X_{B_2^y} \to B}$ is the inverse permutation discussed in the proof of Corollary 4). As the next ingredient, we define the operations that supply the shared randomness of size s_y. For $y = 0, 1, \ldots, Q$, let $S_{X_{A_1^y}}$ and $S_{X_{B_1^y}}$ be the local operations at Alice's and Bob's side respectively, that put shared randomness of size s_y on $X_{A_1^y}X_{B_1^y}$.

We are now ready to put the steps together and give the protocol for classical state splitting of $\bar{\rho}_{AX_AX_{A'}R}$ (depicted in Figure 4). Alice applies the measurement $W_{X_{A'} \to X_{A'}Y_A}$ from (42) followed by

$$S_{A_1Y_A} = \sum_{y=1}^{Q} S_{X_{A_1^y}} \otimes |y\rangle\langle y|_{Y_A} , \tag{50}$$

and the isometry

$$V_{AX_AX_{A'}X_{A_1}Y_A \to AX_{A_2'}X_AY_A} = \sum_{y=0}^{Q} V_{AX_AX_{A'}X_{A_1^y} \to AX_{(A_2')^y}X_A} \otimes |y\rangle\langle y|_{Y_A} . \tag{51}$$

Afterwards she sends $X_{A_2'}$ and Y_A, that is

$$c \leq \max_{y}[H_0(X_{A'})_{\rho^y} - H_{\min}(X_{A'}|R)_{\rho^y}] + 4 \cdot \log\frac{1}{\varepsilon} + 1 + \log\lceil 2 \cdot \log|X_{A'}|\rceil \tag{52}$$

bits to Bob (and we now rename $X_{A_2'}$ to X_{B_2} and Y_A to Y_B). Then Bob applies

$$S_{B_1 Y_B} = \sum_{y=1}^{Q} S_{B_1^y} \otimes |y\rangle\langle y|_{Y_B} \ , \tag{53}$$

followed by the isometry

$$U_{X_{B_1} X_{B_2} Y_B \to X_B Y_B} = \sum_{y=0}^{Q} U_{X_{B_1^y} X_{B_2^y} \to X_B} \otimes |y\rangle\langle y|_{Y_B} \ . \tag{54}$$

We obtain a sub-normalized state

$$\sigma_{A X_A X_B R Y_B} = \sum_{y=0}^{Q} q_y \cdot \tilde{\rho}^y_{A X_A X_B R} \otimes |y\rangle\langle y|_{Y_B} \ , \tag{55}$$

with $\tilde{\rho}^y_{A X_A X_B R} \approx_\varepsilon \mathcal{I}_{X_{A'} \to X_B}(\rho^y_{A X_A X_{A'} R})$ for $y = 0, 1, \ldots, Q$. By the (quasi) convexity of the purified distance in its arguments (Lemma 25), and the monotonicity of the purified distance under partial trace, we have

$$\sigma_{A X_A X_B R} \approx_\varepsilon \mathcal{I}_{X_{A'} \to X_B}(\bar{\rho}_{A X_A X_{A'} R}) \ . \tag{56}$$

Hence, we have shown the existence of an ε-error classical state splitting protocol for $\bar{\rho}_{A X_A X_{A'} R}$ with classical communication cost as in (52). But by the monotonicity of the purified distance, and the triangle inequality for the purified distance, this implies the existence of an $(\varepsilon + |X_{A'}|^{-1/2})$-error classical state splitting protocol for $\rho_{A X_A X_{A'} R}$, with the same classical communication cost as in (52).

We now proceed by simplifying (52). We have $H_0(X_{A'})_{\rho^y} \le H_{\min}(X_{A'})_{\rho^y} + 1$ for $y = 0, 1, \ldots, Q$ as can be seen as follows:

$$2^{-(y+1)} \le \lambda_{\min}(q_y \cdot \rho^y_{X_{A'}}) \le \mathrm{rank}^{-1}\left(q_y \cdot \rho^y_{X_{A'}}\right) \le \left\| q_y \cdot \rho^y_{X_{A'}} \right\|_\infty \le 2^{-y} \ , \tag{57}$$

where $\lambda_{\min}(\rho^y_{X_{A'}})$ denotes the smallest non-zero eigenvalue of $\rho^y_{X_{A'}}$. Thus,

$$\mathrm{rank}\left(q_y \cdot \rho^y_{X_{A'}}\right) \le 2^{y+1} = 2^y \cdot 2 \le \left\| q_y \cdot \rho^y_{X_{A'}} \right\|_\infty^{-1} \cdot 2 \ , \tag{58}$$

and this is equivalent to the claim. Hence, we get an $(\varepsilon + |X_{A'}|^{-1/2})$-error classical state splitting protocol for $\rho_{A X_A X_{A'} R}$ with classical communication cost

$$c \le \max_y [H_{\min}(X_{A'})_{\rho^y} - H_{\min}(X_{A'}|R)_{\rho^y}] + 4 \cdot \log \frac{1}{\varepsilon} + 2 + \log\lceil 2 \cdot \log |X_{A'}| \rceil \tag{59}$$

$$\le \max_y [H_{\min}(X_{A'})_{\rho^y} - H_{\min}(X_{A'}|R)_{\rho^y}] + 4 \cdot \log \frac{1}{\varepsilon} + 4 + \log\log |X_{A'}| \ . \tag{60}$$

Using a lower bound for the max-information in terms of min-entropies (Lemma 13), and the behaviour of the max-information under projective measurements (Lemma 14) this simplifies to

$$c \leq \max_y I_{\max}(X_{A'} : R)_{\rho^y} + 4 \cdot \log \frac{1}{\varepsilon} + 4 + \log \log |X_{A'}| \tag{61}$$

$$\leq I_{\max}(X_{A'} : R)_{\rho} + 4 \cdot \log \frac{1}{\varepsilon} + 4 + \log \log |X_{A'}| \ . \tag{62}$$

Furthermore, it easily seen from (49) that

$$c + s \leq H_0(X_{A'})_{\rho} + 2 + \log \log |X_{A'}| \ . \tag{63}$$

As the last step, we reduce the classical communication and shared randomness cost by smoothing the max-information and the zero-Rényi entropy in (62) and (63), respectively. For that, we do not apply the protocol as described above to the state $\rho_{AX_AX_{A'}R}$, but pretend that we have another classically coherent (sub-normalised) state $\bar{\rho}_{AX_AX_{A'}R}$ that is $(\sqrt{8\varepsilon'} + \varepsilon')$-close to $\rho_{AX_AX_{A'}R}$, and then apply the protocol for $\bar{\rho}_{AX_AX_{A'}R}$. By the monotonicity of the purified distance, the additional error term from this is upper bounded by $\sqrt{8\varepsilon'} + \varepsilon'$, and by the triangle inequality for the purified distance this results in a total accuracy of $\varepsilon + \varepsilon' + \sqrt{8\varepsilon'} + |X_{A'}|^{-1/2}$. We now proceed by defining $\bar{\rho}_{AX_AX_{A'}R}$. Let $\tilde{\rho}_{X_{A'}R} \in \mathcal{B}^{\varepsilon'}(\rho_{X_{A'}R})$ such that

$$I_{\max}^{\varepsilon'}(X_{A'} : R)_{\rho} = I_{\max}(X_{A'} : R)_{\tilde{\rho}} \ . \tag{64}$$

Furthermore, since the zero-Rényi entropy can be smoothed by applying a projection (Lemma 22), there exists $\Pi_{X_{A'}} \in \mathcal{P}(X_{A'})$ with $\Pi_{X_{A'}} \leq \mathbb{1}_{X_{A'}}$ such that

$$H_0^{2\varepsilon'}(X_{A'})_{\tilde{\rho}} \geq H_0(X_{A'})_{\bar{\rho}} \ , \tag{65}$$

with $\bar{\rho}_{X_{A'}} = \Pi_{X_{A'}} \tilde{\rho}_{X_{A'}} \Pi_{X_{A'}} \in \mathcal{B}^{\sqrt{8\varepsilon'}}(\tilde{\rho}_{X_{A'}})$ classical with respect to the basis $\{|x\rangle\}$. By the properties of the purified distance [52, Chapter 3], there exists a purification $\bar{\rho}_{AX_AX_{A'}R} \in \mathcal{B}^{\sqrt{8\varepsilon'}+\varepsilon'}(\rho_{AX_AX_{A'}R})$ that is classically coherent on $X_AX_{A'}$ with respect to the basis $\{|x\rangle\}$. Applying the protocol for this state $\bar{\rho}_{AX_AX_{A'}R}$, the classical communication cost (62) becomes by the monotonicity of the max-information (Lemma 16) and (64),

$$c \leq I_{\max}^{\varepsilon'}(X_{A'} : R)_{\rho} + 4 \cdot \log \frac{1}{\varepsilon} + 4 + \log \log |X_{A'}| \ , \tag{66}$$

and by (65) the sum cost (63) becomes

$$c + s \leq H_0^{\varepsilon'}(X_{A'})_{\rho} + 2 + \log \log |X_{A'}| \ . \tag{67}$$

$\square$

For completeness we also state a converse for the classical communication cost of classical state splitting of classically coherent states.

Theorem 6. *Let $\varepsilon \geq 0$, $\varepsilon' > 0$, and $\rho_{AX_A X_{A'} R} \in \mathcal{V}_{\leq}(\mathcal{H}_{AX_A X_{A'} R})$ be classically coherent on $X_A X_{A'}$ with respect to the basis $\{|x\rangle\}_{x \in X_A X_{A'}}$. Then the classical communication cost for any ε-error classical state splitting protocol for $\rho_{AX_A X_{A'} R}$ is lower bounded by*[3]

$$c \geq I_{\max}^{\varepsilon + \varepsilon'}(X_{A'} : R)_\rho - \log\left(\frac{8}{(\varepsilon')^2} + 2\right) . \tag{68}$$

Proof. We have a look at the correlations between Bob and the reference by analyzing the max-information that Bob has about the reference (recall that this will be a max-information of the form $I_{\max}(R : B)$ where R is the reference system and B here is a general label for whatever Bob's system is). At the beginning of any protocol, there is no register at Bob's side correlated with the reference and therefore the max-information that Bob has about the reference is zero. Since back communication is not allowed, we can assume that the protocol for state splitting has the following form: applying local operations at Alice's side, sending bits from Alice to Bob and then applying local operations at Bob's side. Local operations at Alice's side have no influence on the max-information that Bob has about the reference. By sending c bits from Alice to Bob, the max-information that Bob has about the reference can increase, but at most by c (Corollary 18). By applying local operations at Bob's side, the max-information that Bob has about the reference can only decrease (Lemma 12). So the max-information that Bob has about the reference is upper bounded by c. Therefore, any state $\omega_{X_B R}$ at the end of a state splitting protocol must satisfy $I_{\max}(R : X_B)_\omega \leq c$. But we also need $\omega_{X_B R} \approx_\varepsilon \rho_{X_B R} \equiv \mathcal{I}_{X_{A'} \to X_B}(\rho_{X_{A'} R})$ by the definition of ε-error state splitting (Definition 3). Using the definition of the smooth max-information, and that the smooth max-information is approximately symmetric in its arguments (Lemma 1), we obtain the bound in the statement of the theorem. $\qquad \square$

V. UNIVERSAL MEASUREMENT COMPRESSION

In this section, we establish our main result: feedback and non-feedback universal measurement compression. Theorem 7 characterizes the trade-off between shared randomness and classical communication required to simulate many instances of a measurement on an arbitrary input state in such a way that both the sender and receiver obtain the outcomes of the measurement (feedback

---

[3] We do not mention the cost of the shared randomness resource, since the statement holds independently of it.

simulation), and Theorem 10 characterizes the trade-off for the non-feedback case when only the receiver is required to get the outcomes of the measurement.

Definition 5 (One-shot Measurement Compression). *Consider a bipartite system with parties Alice and Bob. Let $\delta \geq 0$, and $\mathcal{M} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_X)$ be a quantum-classical channel, with quantum input A and classical output X. A quantum protocol $\mathcal{P}$ is a one-shot feedback measurement compression for $\mathcal{M}$ with error δ if it consists of using s bits of shared randomness, applying local operations at Alice's side, sending c classical bits from Alice to Bob, applying local operations at Bob's side, and*

$$\|\mathcal{P} - \overline{\Delta} \circ \mathcal{M}\|_\Diamond \leq \delta , \tag{69}$$

where $\overline{\Delta} : \mathcal{L}(\mathcal{H}_X) \to \mathcal{L}(\mathcal{H}_{X_A}) \otimes \mathcal{L}(\mathcal{H}_{X_B})$ is a classical copying map,

$$\overline{\Delta}(\sigma) \equiv \sum_x \langle x|\sigma|x\rangle \ |x\rangle\langle x|_{X_A} \otimes |x\rangle\langle x|_{X_B}, \tag{70}$$

ensuring that both Alice and Bob obtain the measurement outcome. The quantity c is called the classical communication cost, and s is the shared randomness cost. For the case of a non-feedback measurement compression, we only require the following condition to hold

$$\|\mathcal{P} - \mathcal{M}\|_\Diamond \leq \delta , \tag{71}$$

because Alice does not need to recover the output of the simulation in this case.[4]

Definition 6 (Universal Measurement Compression). *Let $\mathcal{M} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_X)$ be a quantum-classical channel. An asymptotic measurement compression for $\mathcal{M}$ is a sequence of one-shot measurement compressions $\mathcal{P}^n$ for $\mathcal{M}^{\otimes n}$ with error δ_n, such that $\lim_{n\to\infty} \delta_n = 0$. The classical communication rate is $\limsup_{n\to\infty} \frac{\log c_n}{n}$ and the shared randomness rate is $\limsup_{n\to\infty} \frac{\log s_n}{n}$ (where c_n and s_n denote the corresponding costs for the one-shot measurement compressions).*

A. Feedback Simulation

Theorem 7. *Let $\mathcal{M} : \mathcal{L}(A) \to \mathcal{L}(X)$ be a quantum to classical channel. Then there exist asymptotic feedback measurement compressions for $\mathcal{M}$ if and only if the classical communication rate C and shared randomness rate S lie in the following rate region:[5]*

$$C \geq \max_\rho I(X : R)_{(\mathcal{M} \otimes \mathcal{I})(\rho)} \tag{72}$$

$$C + S \geq \max_\rho H(X)_{\mathcal{M}(\rho)} , \tag{73}$$

---

[4] If we state the task of measurement compression as being that a verifier who is given the reference system and classical output should not be able to distinguish the true channel from the simulation, then we should also demand that the common randomness and classical communication be private from the verifier.

[5] Note that the two maxima in (72) and (73) can be achieved for different states.

where $\rho_{AR} \in \mathcal{V}(AR)$ is a purification of the input state $\rho_A \in \mathcal{S}(A)$. Or equivalently, for a given shared randomness rate S, the optimal rate of classical communication is equal to

$$C(S) = \max\left\{\max_\rho I(X:R)_{(\mathcal{M}\otimes\mathcal{I})(\rho)}, \; \max_\rho H(X)_{\mathcal{M}(\rho)} - S\right\} . \tag{74}$$

In particular, when sufficient shared randomness is available, the rate of classical communication is given by

$$C(\infty) = \max_\rho I(X:R)_{(\mathcal{M}\otimes\mathcal{I})(\rho)}. \tag{75}$$

Proof. We first show that the right-hand side of (72) is a lower bound on the classical communication rate, and that (73) is a lower bound on the sum rate (Propositition 8). Then we show that these lower bounds can be achieved (Proposition 9). The general rate trade-off in (72)-(73) and (74) immediately follows, since the shared randomness can always be created by classical communication. $\quad\square$

Proposition 8 (Converse). *Let $\mathcal{M} : \mathcal{L}(A) \to \mathcal{L}(X)$ be a quantum to classical channel. Then we have for any asymptotic measurement compression for $\mathcal{M}$ that*

$$C \geq \max_\rho I(X:R)_{(\mathcal{M}\otimes\mathcal{I})(\rho)} \tag{76}$$

$$C + S \geq \max_\rho H(X)_{\mathcal{M}(\rho)} , \tag{77}$$

where $\rho_{AR} \in \mathcal{V}(AR)$ is a purification of the input state $\rho_A \in \mathcal{S}(A)$.

Proof. This proposition follows from the converse for the case of a fixed IID source [60, Theorem 8], since the asymptotic measurement compressions must in particular work for any fixed IID input state $\rho_A^{\otimes n}$ (for $n \to \infty$). To see this explicitly worked out with the feedback assumption, see Section 2.4 of [57]. $\quad\square$

Proposition 9 (Achievability). *Let $\mathcal{M} : \mathcal{L}(A) \to \mathcal{L}(X)$ be a quantum to classical channel. Then there exist asymptotic feedback measurement compressions for $\mathcal{M}$ with*

$$C \leq \max_\rho I(X:R)_{(\mathcal{M}\otimes\mathcal{I})(\rho)} \tag{78}$$

$$C + S \leq \max_\rho H(X)_{\mathcal{M}(\rho)} , \tag{79}$$

where $\rho_{AR} \in \mathcal{V}(AR)$ is a purification of the input state $\rho_A \in \mathcal{S}(A)$.

Proof. We show the existence of a sequence of one-shot feedback measurement compressions $\mathcal{P}^n$ for $\mathcal{M}^{\otimes n}$ with asymptotically vanishing error ε_n, a classical communication rate $\frac{c_n}{n}$ as in (78), and a shared randomness rate $\frac{s_n}{n}$ such that the sum rate becomes as in (79). Without loss of generality, we choose $\mathcal{P}^n$ to be permutation covariant.[6] The post-selection technique for quantum channels (Proposition 28) then applies and upper bounds the error by

$$\delta_n = \|\mathcal{M}_{A \to X_B}^{\otimes n} - \mathcal{P}_{A^n \to X_B^n}^n\|_\diamond \leq (n+1)^{|A|^2-1} \cdot \|((\mathcal{M}_{A \to X_B}^{\otimes n} - \mathcal{P}_{A^n \to X_B^n}^n) \otimes \mathcal{I}_R^{\otimes n} \otimes \mathcal{I}_{R'}(\zeta_{ARR'}^n)\|_1 \,, \tag{80}$$

where $\zeta_{ARR'}^n$ is a purification of the de Finetti state $\zeta_{AR}^n = \int \psi_{AR}^{\otimes n} \, d(\psi_{AR})$ with $\psi_{AR} \in \mathcal{V}(AR)$, $A \cong R$ and $d(\cdot)$ the measure on the normalized pure states on AR induced by the Haar measure on the unitary group acting on AR, normalized to $\int d(\cdot) = 1$. Hence, it is sufficient to consider simulating the measurement on a purification of the de Finetti state:

$$\omega_{X_B RR'}^n = \left(\mathcal{M}_{A \to X_B}^{\otimes n} \otimes \mathcal{I}_R^{\otimes n} \otimes \mathcal{I}_{R'}\right)(\zeta_{ARR'}^n) \,, \tag{81}$$

up to an error $o\left((n+1)^{1-|A|^2}\right)$ in trace distance, for an asymptotic classical communication cost smaller than (78). For this, we consider a local Stinespring dilation $U_{A \to EX_A X_{A'}}$ of the measurement $\mathcal{M}_{A \to X_{A'}}$ at Alice's side, followed by classical state splitting of the resulting classically coherent state (Theorem 5). Let $U_{A^n \to E^n X_A^n X_{A'}^n} = U_{A \to EX_A X_{A'}}^{\otimes n}$ and

$$\omega_{E^n X_A^n X_{A'}^n, R^n R'} = U_{A^n \to E^n X_A^n X_{A'}^n}(\zeta_{ARR'}^n) \,. \tag{82}$$

As mentioned above, this map can be made permutation invariant. For fixed $\varepsilon_n > 0$, Theorem 5 then assures that the map outputs a state which is

$$4 \cdot \varepsilon_n + 4\sqrt{2\varepsilon_n} + 2 \cdot |X_{A'}|^{-n/2} \tag{83}$$

close to (81) in trace distance,[7] for a classical communication cost

$$c_n \leq I_{\max}^{\varepsilon_n}(X_{A'} : RR')_\omega + 4 \cdot \log \frac{1}{\varepsilon_n} + 4 + \log\log|X_{A'}| + \log n \,, \tag{84}$$

and a sum cost

$$c_n + s_n \leq H_0^{\varepsilon_n}(X_{A'})_\omega + 2 + \log\log|X_{A'}| + \log n \,, \tag{85}$$

---

[6] By the following argument, every protocol can be made permutation covariant. To start with, Alice applies a random permutation π on the input system chosen according to some shared randomness. This is then followed by the original protocol (which might not yet be permutation covariant), and Bob who undoes the permutation by applying π^{-1} on the output system. The shared randomness cost of this procedure can be kept sub-linear in n by using randomness recycling as discussed in [2, Section IV. D].

[7] The trace distance is upper bounded by two times the purified distance (Lemma 24).

where the last two terms on the right in each of the above expressions come from the fact that $\log \log |X_{A'}|^n = \log \log |X_{A'}| + \log n$. We now analyse the asymptotic behaviour of (84) and (85). By a dimension upper bound for the smooth max-information (Lemma 17), and the fact that we can assume $|R'| \leq (n+1)^{|A|^2-1}$ (Proposition 28), we get

$$c_n \leq I_{\max}^{\varepsilon_n}(X_{A'} : R)_\omega + 2 \cdot \log\left[(n+1)^{|A|^2-1}\right] + 4 \cdot \log \frac{1}{\varepsilon_n} + 4 + \log \log |X_{A'}| + \log n . \tag{86}$$

By a corollary of Carathéodory's theorem (Lemma 29), we write

$$\zeta_{AR}^n = \sum_{i \in I} p_i (\sigma_{AR}^i)^{\otimes n} , \tag{87}$$

where $\sigma_{AR}^i \in \mathcal{V}(AR)$, $I = \{1, 2, \ldots, (n+1)^{2|A||R|-2}\}$, and $\{p_i\}_{i \in I}$ a probability distribution. Using a quasi-convexity property of the smooth max-information (Lemma 19), and for

$$\chi = 2 \cdot \log\left[(n+1)^{|A|^2-1}\right] + 4 \cdot \log \frac{1}{\varepsilon_n} + 4 + \log \log |X_{A'}| + \log n , \tag{88}$$

we obtain

$$c_n \leq I_{\max}^{\varepsilon_n}(X_{A'} : R)_{(\mathcal{M}^{\otimes n} \otimes \mathcal{I})(\sum_i p_i(\sigma^i)^{\otimes n})} + \chi \tag{89}$$

$$\leq \max_i I_{\max}^{\varepsilon_n}(X_{A'} : R)_{[(\mathcal{M} \otimes \mathcal{I})(\sigma^i)]^{\otimes n}} + \log\left[(n+1)^{2|A||R|-2}\right] + \chi \tag{90}$$

$$\leq \max_\rho I_{\max}^{\varepsilon_n}(X_{A'} : R)_{[(\mathcal{M} \otimes \mathcal{I})(\rho)]^{\otimes n}} + \log\left[(n+1)^{2|A||R|-2}\right] + \chi , \tag{91}$$

where the last maximum ranges over all $\rho_{AR} \in \mathcal{V}(AR)$. From the asymptotic equipartition property for the smooth max-information (Lemma 23) we obtain

$$c_n \leq n \cdot \max_\rho I(X_{A'} : R)_{(\mathcal{M} \otimes \mathcal{I})(\rho)} + \sqrt{n} \cdot \xi(\varepsilon_n) - 2 \cdot \log \frac{\varepsilon_n^2}{24} + \log\left[(n+1)^{2|A||R|-2}\right] + \chi , \tag{92}$$

where $\xi(\varepsilon_n) = 8\sqrt{13 - 4 \cdot \log \varepsilon_n} \cdot (2 + \frac{1}{2} \cdot \log |A|)$. By choosing

$$\varepsilon_n = (n+1)^{4(1-|A|^2)} , \tag{93}$$

we get an asymptotic classical communication cost of

$$c = \limsup_{n \to \infty} \frac{c_n}{n} \leq \max_\rho I(X_{A'} : R)_{(\mathcal{M} \otimes \mathcal{I})(\rho)} , \tag{94}$$

for a vanishing asymptotic error (80), (83), (93):

$$\limsup_{n \to \infty} \delta_n \leq \limsup_{n \to \infty}\left[\left(4 \cdot (n+1)^{4(1-|A|^2)} + 4\sqrt{2} \cdot (n+1)^{2(1-|A|^2)} + 2 \cdot |X_{A'}|^{-n/2}\right)(n+1)^{|A|^2-1}\right]$$

$$= 0 . \tag{95}$$

Furthermore, we estimate the asymptotic behaviour of the sum cost (85) by using (87) and a quasi-convexity property of the smooth zero-Rényi entropy (Lemma 20). For $\chi' = 2 + \log\log|X_{A'}| + \log n$ we get

$$c_n + s_n \leq \max_i H_0^{\varepsilon_n}(X_{A'})_{\mathcal{M}(\sigma^i)^{\otimes n}} + \log\left[(n+1)^{2|A||R|-2}\right] + \chi' \tag{96}$$

$$\leq \max_\rho H_0^{\varepsilon_n}(X_{A'})_{\mathcal{M}(\rho)^{\otimes n}} + \log\left[(n+1)^{2|A||R|-2}\right] + \chi' \ , \tag{97}$$

where $\rho_A \in \mathcal{S}(A)$. By the equivalence of the smooth zero-Rényi entropy and the smooth max-entropy (Lemma 21), and the asymptotic equipartition property for the smooth max-entropy (Lemma 23), we arrive at

$$c_n + s_n \leq \max_\rho H_{\max}^{\varepsilon_n/2}(X_{A'})_{\mathcal{M}(\rho)^{\otimes n}} + 2 \cdot \log\frac{8}{\varepsilon_n^2} + \log\left[(n+1)^{2|A||R|-2}\right] + \chi' \tag{98}$$

$$\leq n \cdot \max_\rho H(X_{A'})_{\mathcal{M}(\rho)} + \sqrt{n} \cdot 4\sqrt{1 - 2 \cdot \log\frac{\varepsilon_n}{2}} \cdot (2 + \frac{1}{2} \cdot \log|X_{A'}|)$$

$$+ 2 \cdot \log\frac{8}{\varepsilon_n^2} + \log\left[(n+1)^{2|A||R|-2}\right] + \chi' \ , \tag{99}$$

where $\rho_A \in \mathcal{S}(A)$. By employing (93), we get for the asymptotic limit

$$c + s = \limsup_{n\to\infty} \frac{1}{n}(c_n + s_n) \leq \max_\rho H(X_{A'})_{\mathcal{M}(\rho)} \ , \tag{100}$$

where $\rho_A \in \mathcal{S}(A)$. $\qquad\square$

B. Non-Feedback Simulation

Theorem 10. *Let $\mathcal{M} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_X)$ be a quantum-to-classical channel. Then there exist asymptotic non-feedback measurement compressions for $\mathcal{M}$ if and only if the classical communication rate C and shared randomness rate S lie in the rate region given by the union of the following regions,*

$$C \geq \max_\rho I(W : R)_\beta \tag{101}$$

$$C + S \geq \max_\rho I(W : XR)_\beta \ , \tag{102}$$

where the state β_{WXR} has the form

$$\beta_{WXR} = \sum_{w,x} q_{x|w} \cdot |w\rangle\langle w|_W \otimes |x\rangle\langle x|_X \otimes \mathrm{tr}_A\left[(\mathcal{N}_w \otimes \mathcal{I})(\rho_{AR})\right] \ , \tag{103}$$

$\rho_{AR} \in \mathcal{V}(\mathcal{H}_{AR})$ *is a purification of the input state* $\rho_A \in \mathcal{S}(\mathcal{H}_A)$, *and the union is with respect to all decompositions of the measurement* $\mathcal{M}$ *in terms of internal measurements* $\mathcal{N} = \{\mathcal{N}_w\}$ *and conditional post-processing distributions* $q_{x|w}$. *That is, for all states* σ, *it should hold that*

$$\sum_x \mathcal{M}_x(\sigma)|x\rangle\langle x| = \sum_{x,w} q_{x|w} \mathcal{N}_w(\sigma)|x\rangle\langle x|. \tag{104}$$

Or equivalently, for a given shared randomness rate S, *the optimal rate of classical communication is equal to*

$$C(S) = \min_{\mathcal{N} : \sum_w q_{x|w}\mathcal{N}_w = \mathcal{M}_x} \max\left\{ \max_\rho I(W:R)_\beta,\ \max_\rho I(W:XR)_\beta - S \right\}. \tag{105}$$

By the data processing inequality for the mutual information, it holds that $I(W:R)_\beta \geq I(X:R)_{\mathcal{M}(\rho)}$, and hence, the classical communication cost can only increase compared to a feedback simulation (Theorem 7). However, if the savings in common randomness consumption are larger than the increase in classical communication cost, then there is an advantage to performing a non-feedback simulation. It follows from the considerations in [40, 57] that the rate trade-offs (74) and (105) become identical if and only if the elements of the measurement to simulate are all rank-one operators.

Proof. We see from the converse for the case of a fixed IID source [57, Theorem 9], that the right-hand side of (101) is a lower bound on the classical communication rate, and that (102) is a lower bound on the sum rate. This is because the asymptotic non-feedback measurement compression must work in particular for any fixed IID input state $\rho_A^{\otimes n}$ (as $n \to \infty$).

As the next step, we show that these lower bounds can be achieved. The general rate trade-off in (101)-(102) and (105) then immediately follows, since shared randomness can always be created by classical communication.

The idea for the achievability part is as follows. Given a particular decomposition of the measurement $\mathcal{M} = \{\mathcal{M}_x\}$ as $\left\{\sum_w q_{x|w} \cdot \mathcal{N}_w\right\}$ as stated above, Alice and Bob just use a feedback measurement compression protocol (as in the proof of Theorem 7) to simulate the measurement $\mathcal{N} = \{\mathcal{N}_w\}$. This is followed by a local simulation of the classical map $q_{x|w}$ at no cost at Bob's side. Finally, Alice and Bob can use randomness recycling to extract $H_{\min}(W|RX)_\beta$ bits of shared randomness back [2]. In the one-shot case, this leads to a classical communication cost of $I_{\max}(W:R)_\beta$, and a sum cost $I_{\max}(W:RX)_\beta$. For technical reasons, we smooth the states using typical projectors (see Appendix E for background on typical projectors) and arrive at the rates given in the statement of the theorem.

Let $\{q_{x|w}, \mathcal{N}_\omega\}$ be a fixed decomposition of $\mathcal{M}$. As in the feedback case (Theorem 7) we employ the post-selection technique (Proposition 28) to upper bound the error for one-shot non-feedback compressions $\mathcal{P}^n$ for $\mathcal{M}^{\otimes n}$ by

$$\delta_n = \|\mathcal{M}^{\otimes n}_{A \to X_B} - \mathcal{P}^n_{A \to X_B}\|_\diamond \tag{106}$$

$$\leq (n+1)^{|A|^2-1} \cdot \|((\mathcal{M}^{\otimes n}_{A \to X_B} - \mathcal{P}^n_{A \to X_B}) \otimes \mathcal{I}^{\otimes n}_R \otimes \mathcal{I}_{R'}(\zeta^n_{ARR'})\|_1 \,, \tag{107}$$

where $\zeta^n_{ARR'}$ is a purification of the de Finetti state $\zeta^n_{AR} = \int \psi^{\otimes n}_{AR} \, d(\psi_{AR})$ with $\psi_{AR} \in \mathcal{V}(\mathcal{H}_{AR})$, $A \cong R$ and $d(\cdot)$ the measure on the normalized pure states on $\mathcal{H}_{AR}$ induced by the Haar measure on the unitary group acting on $\mathcal{H}_{AR}$, normalized to $\int d(\cdot) = 1$. Hence, it is sufficient to consider simulating the measurement $\mathcal{M}^{\otimes n}$ on a purification of the de Finetti state

$$\omega^n_{X_B RR'} = \left(\mathcal{M}^{\otimes n}_{A \to X_B} \otimes \mathcal{I}^{\otimes n}_R \otimes \mathcal{I}_{R'}\right)(\zeta^n_{ARR'}) \,, \tag{108}$$

up to an error $o\left((n+1)^{1-|A|^2}\right)$ in trace distance, for an asymptotic simulation cost smaller than in (101) and (102). For this, the idea is to consider a local Stinespring dilation $V_{A \to EW_A W_{A'}}$ of the measurement $\mathcal{N}_{A \to W_A}$ at Alice's side, followed by classical state splitting of the resulting classically coherent state (along Theorem 5). Let $V^n_{A \to EW_A W_{A'}} = V^{\otimes n}_{A \to EW_A W_{A'}}$ and

$$\omega^n_{EW_A W_{A'} RR'} = V^n_{A \to EW_A W_{A'}}(\zeta^n_{ARR'}) \,. \tag{109}$$

However, Alice and Bob will not execute the protocol with respect to the state $\omega^n_{EW_A W_{A'} RR'}$ directly, but they will do so with respect to another pure, sub-normalized state $\bar{\gamma}^n_{EW_A W_{A'} RR'}$ that is also classically coherent on $W_A W_{A'}$ with respect to the basis $\{|w\rangle\}_{w \in W_A}$, and such that

$$\|\bar{\gamma}^n_{EW_A W_{A'} RR'} - \omega^n_{EW_A W_{A'} RR'}\|_1 \leq \varepsilon_n \,, \tag{110}$$

for some $\varepsilon_n > 0$. By a corollary of Carathéodory's theorem (Lemma 29), we write

$$\zeta^n_{AR} = \sum_{i \in I} p_i \cdot (\sigma^i_{AR})^{\otimes n} \,, \tag{111}$$

where $\sigma^i_{AR} \in \mathcal{V}(\mathcal{H}_{AR})$, $I = \{1, 2, \ldots, (n+1)^{2|A||R|-2}\}$, and $\{p_i\}_{i \in I}$ a probability distribution. From this, we define

$$\gamma^{i,n}_{EW_A W_{A'} R} = [(V_{A \to EW_A W_{A'}} \otimes \mathcal{I}_R)(\sigma^i_{AR})]^{\otimes n}, \tag{112}$$

as well as its reduction as a classical-quantum state $\gamma^{i,n}_{W_A R}$ on the systems $W^n_{A'} R^n$:

$$\gamma^{i,n}_{W_A R} = \sum_{w^n} p_{W^n|i}(w^n|i) \, |w^n\rangle\langle w^n|_{W^n_{A'}} \otimes \gamma^{i,w^n}_{R^n}, \tag{113}$$

for some distribution $p_{W^n|i}(w^n|i)$. On this state, we act with typical projectors to flatten its spectrum as we need, defining the projected state $\bar{\gamma}^{i,n}_{W_A R}$ as follows:

$$\bar{\gamma}^{i,n}_{W_A R} = \sum_{w^n} p_{W^n|i}(w^n|i)\, \Pi^{W^n|i}_{\delta} |w^n\rangle\langle w^n|_{W^n_{A'}}\, \Pi^{W^n|i}_{\delta} \otimes \Pi^n_{\gamma^i,\delta}\, \Pi^n_{\gamma^{i,w^n},\delta}\, \gamma^{i,w^n}_{R^n}\, \Pi^n_{\gamma^{i,w^n},\delta}\, \Pi^n_{\gamma^i,\delta}, \tag{114}$$

where $\Pi^{W^n|i}_{\delta}$ is a typical projector corresponding to the distribution $p_{W^n|i}(w^n|i)$, $\Pi^n_{\gamma^{i,w^n},\delta}$ is a conditionally typical projector corresponding to the conditional state γ^{i,w^n} on the system R^n, and $\Pi^n_{\gamma^i,\delta}$ is a typical projector corresponding to the state $\gamma^{i,n}_R$ (see Appendix E for details of typical projectors). It follows from the properties of typical projectors that the projected state $\bar{\gamma}^{i,n}_{W_A R}$ becomes arbitrarily close in trace distance to the original state $\gamma^{i,n}_{W_A R}$:

$$\|\gamma^{i,n}_{W_A R} - \bar{\gamma}^{i,n}_{W_A R}\|_1 \leq \frac{\varepsilon_n^2}{4}, \tag{115}$$

for some $\varepsilon_n > 0$ and sufficiently large n. The equivalence of the trace distance and the purified distance (Lemma 24) together with Uhlmann's theorem then imply the existence of some subnormalized pure state $\bar{\gamma}^{i,n}_{EW_A W_{A'} R}$ such that

$$P(\gamma^{i,n}_{EW_A W_{A'} R}, \bar{\gamma}^{i,n}_{EW_A W_{A'} R}) \leq \frac{\varepsilon_n}{2}. \tag{116}$$

Hence, we get by (111) and (109) that

$$\bar{\gamma}^n_{EW_A W_{A'} R} = \sum_{i \in I} p_i \cdot \bar{\gamma}^{i,n}_{EW_A W_{A'} R} \tag{117}$$

is ε_n-close to $\omega^n_{EW_A W_{A'} R}$ in purified distance. By features of the purified distance [52, Chapter 3], and the equivalence of the trace distance and the purified distance (Lemma 24), we then get that there exists an extension $\bar{\gamma}^n_{EW_A W_{A'} RR'}$ of $\bar{\gamma}^n_{EW_A W_{A'} R}$ with the desired properties such that (110) holds.

Alice and Bob will now act with a classical state splitting protocol for $W_A W_{A'}$ with respect to the classically coherent state $\bar{\gamma}^n_{EW_A W_{A'} RR'}$. However, we do not directly use our result about classical state splitting (Theorem 5), but instead employ a non-smooth version that is implicit in the proof of Theorem 5. It follows from (61) and (62) that for an $(\varepsilon_n + |W_{A'}|^{-n})$-error (in purified distance) classical state splitting protocol for $W_A W_{A'}$, a classical communication cost

$$c_n \leq \max_y I_{\max}(W_{A'} : RR')_{\bar{\gamma}^{n,y}} + 4 \cdot \log \frac{1}{\varepsilon_n} + 4 + \log\log|W_{A'}| + \log n \tag{118}$$

is achievable, and it follows from (49) and (63) that the sum cost becomes

$$c_n + s_n \leq \max_y H_0(W_{A'})_{\bar{\gamma}^{n,y}} + 2 + \log\log|W_{A'}| + \log n, \tag{119}$$

where the measurement outcomes y are with respect to the pre-processing measurement defined in (42). This provides Bob with the measurement outcomes of $\mathcal{N}$ for the fixed de Finetti type input state $\zeta_{ARR'}$, and a total error of $(3 \cdot \varepsilon_n + 2 \cdot |W_{A'}|^{-n})$ in trace distance. A local simulation of the classical map $q_{x^n|w^n}$ at no cost at Bob's side then provides Bob with the measurement outcomes of $\mathcal{M}$ as desired (again for the fixed de Finetti type input state $\zeta_{ARR'}$ and the same error). However, the sum cost of this non-feedback measurement simulation can be reduced by invoking an additional randomness recycling step as in Ref. [2]. We do this by having Alice and Bob apply, conditioned on y, a strong classical min-entropy extractor on W against the (quantum) side information XRR' (Proposition 27), and this lowers the sum cost to

$$c_n + s_n \leq \max_y \left(H_0(W_{A'})_{\bar{\gamma}^{n,y}} - H_{\min}(W_{A'}|RR'X_{A'})_{\bar{\gamma}^{n,y}} \right) + 4 \cdot \log \frac{1}{\varepsilon_n} + 2 + \log \log |W_{A'}| \ , \quad (120)$$

for an additional error ε_n in trace distance, leading to a total error of

$$(4 \cdot \varepsilon_n + 2 \cdot |W_{A'}|^{-n}) \tag{121}$$

in trace distance. The min-entropy extractor is performed with respect to the following typical projected state, in order to increase the amount of randomness that can be extracted:

$$\bar{\gamma}^{i,n}_{XW_AR} = \sum_{w^n,x^n} q(x^n|w^n)p_{W^n|i}(w^n|i)\, \Pi_\delta^{X^n|W^n,i}\, |x^n\rangle\langle x^n|_{X^n}\, \Pi_\delta^{X^n|W^n,i} \otimes$$
$$\Pi_\delta^{W^n|i}\, |w^n\rangle\langle w^n|_{W_{A'}^n}\, \Pi_\delta^{W^n|i} \otimes \Pi_{\gamma^i,\delta}^n\, \Pi_{\gamma^i,w^n,\delta}^n\, \gamma^{i,w^n}_{R^n}\, \Pi_{\gamma^i,w^n,\delta}^n\, \Pi_{\gamma^i,\delta}^n \ . \quad (122)$$

In the rest of the proof, we bring the classical communication cost (118) and the sum cost (119) into the right form, and show that the asymptotic error for the measurement simulation (106) becomes zero. By the behavior of the max-information under projective measurements (Corollary 14), a dimension upper bound for the max-information (Lemma 17), the fact that we can assume $|R'|^{|A|^2-1}$ (Proposition 28), and a quasi-convexity property of the max-information (Lemma 19), we get

$$c_n \leq \max_{i \in I} I_{\max}(W_{A'} : R)_{\bar{\gamma}^{i,n}} + \chi \ , \tag{123}$$

where

$$\chi = 2 \cdot \log \left((n+1)^{|A|^2-1}\right) + \log \left((n+1)^{2|A||R|-2}\right) + 4 \cdot \log \frac{1}{\varepsilon_n} + 4 + \log \log |W_{A'}| + \log n \ . \quad (124)$$

By an upper bound on the max-information (Lemma 13), and a lower bound on the conditional min-entropy (Lemma 11), this can be estimated to be

$$c_n \leq \max_{i \in I} \left(H_R(W_{A'})_{\bar{\gamma}^{i,n}} - H_{\min}(W_{A'}R)_{\bar{\gamma}^{i,n}} + H_0(R)_{\bar{\gamma}^{i,n}} \right) + \chi \ . \tag{125}$$

By (114), as well as the properties of typical projectors (see Appendix E), we get

$$c_n \leq n \cdot \max_{i \in I} \left(H(W_{A'})_{\gamma^i} - H(W_{A'}R)_{\gamma^i} + H(R)_{\gamma^i} \right) + 5nc\delta + \chi \tag{126}$$

$$\leq n \cdot \max_{\rho} \left(H(W_{A'})_{\mathcal{N}(\rho)} - H(W_{A'}R)_{(\mathcal{N} \otimes \mathcal{I})(\rho)} + H(R)_{\rho} \right) + 5nc\delta + \chi \ , \tag{127}$$

where $\rho_{AR} \in \mathcal{V}(\mathcal{H}_{AR})$, c is a constant, and $\delta > 0$ is the typicality tolerance.

By choosing

$$\varepsilon_n = (n+1)^{4(1-|A|^2)} \ , \tag{128}$$

we finally get an asymptotic classical communication cost of

$$c = \limsup_{n \to \infty} \frac{c_n}{n} \leq \max_{\rho} I(W_{A'} : R)_{\beta} \ , \tag{129}$$

where $\rho_{AR} \in \mathcal{V}(\mathcal{H}_{AR})$, $\beta_{W_{A'}R}$ is as in (103), and a vanishing asymptotic error (106), (121),

$$\limsup_{n \to \infty} \delta_n \leq \limsup_{n \to \infty} \left((n+1)^{|A|^2-1} \cdot \left(4 \cdot (n+1)^{4(1-|A|^2)} + 2 \cdot |X_{A'}|^{-n} \right) \right) = 0 \ . \tag{130}$$

For the sum cost (119) we get by the definition of the measurement in (42) with outcomes y, and a line of argument as in (58) that

$$c_n + s_n \leq \max_{y} \left(H_{\min}(W_{A'})_{\bar{\gamma}^{n,y}} - H_{\min}(W_{A'}|RR'X_{A'})_{\bar{\gamma}^{n,y}} \right)$$
$$+ 4 \cdot \log \frac{1}{\varepsilon_n} + 2 + \log \log |W_{A'}| + \log n \tag{131}$$

$$\leq \max_{y} I_{\max}(W_{A'} : RX_{A'})_{\bar{\gamma}^{n,y}}$$
$$+ 2 \cdot \log \left((n+1)^{|A|^2-1} \right) + 4 \cdot \log \frac{1}{\varepsilon_n} + 2 + \log \log |W_{A'}| + \log n \ , \tag{132}$$

where we used a lower bound on the max-information (Lemma 13), as well as a dimension upper bound for the max-information (Lemma 17), and the fact that $|R'| \leq (n+1)^{|A|^2-1}$ (Proposition 28). Using similar arguments (see Appendix E) as in the estimation of the classical communication cost, we arrive at

$$c + s = \limsup_{n \to \infty} \frac{c_n + s_n}{n} \leq \max_{\rho} I(W_{A'} : RX_{A'})_{\beta} \ , \tag{133}$$

where $\rho_{AR} \in \mathcal{V}(\mathcal{H}_{AR})$, and $\beta_{W_{A'}RX_{A'}}$ is as in (103). By minimizing over all decompositions of the measurement $\mathcal{M}$ as in (104), the claim follows. $\qquad \square$

VI.   EXTENSIONS AND APPLICATIONS

a.   Structured State Splitting Scheme   The state splitting protocol presented in Theorem 5 has the drawback that the permutations U_y used by Bob must be chosen at random and little is known about the structure of the unitaries V_y. We can remedy this by basing the state splitting protocol used in Theorem 5 on a modified state merging protocol instead of that in Lemma 2. The new protocol has the advantage that Alice's classical operation P (recall that the roles are reversed) is a *linear* function rather than an arbitrary permutation, though still randomly-chosen, and Bob's unitary operation V is based on the decoder of an information reconciliation protocol. We now give a sketch of this modified state merging protocol.

The protocol is based on the observation from [8, 45] that state merging is a by-product of an entanglement distillation protocol in which Alice measures the stabilizers of a Calderbank-Shor-Steane (CSS) code such that, given the resulting (classical) syndrome results, Bob could determine both the amplitude (logical X value) and phase (logical Z value)[8] of Alice's remaining encoded system by using his systems. Indeed, for state merging of classically coherent states such as $\rho_{X_A X_B B R}$ in Lemma 2, the situation is considerably simpler since Bob can already determine the amplitude of Alice's system X_A by measuring X_B. For simplicity, let us regard X_A as a collection of $k = \log|X_A|$ qubits.

Thus, from the analysis of [8, 45], all that remains is for Alice to measure a sufficient number of phase stabilizers from an error-correcting code to enable Bob to determine the phase of her encoded systems by using the syndromes and his systems X_B and B, with probability of error at most ϵ. Use of a linear code ensures that Alice does not damage Bob's amplitude information in the course of trying to increase his phase information. Since the task at hand is equivalent to information reconciliation, the number of phase stabilizers needed for this purpose is no more than $H_{\max}(\widetilde{X}_A|X_B B)_\rho + 2\log\frac{1}{\epsilon} + 4$ [46], where $\widetilde{X}_A$ denotes the phase observable conjugate to the amplitude observable X_A.

To measure the phase stabilizers, Alice can apply a suitable unitary operation to all of her systems and then simply measure the phases of a certain subset of the outputs which correspond to the stabilizers [41]. But for stablizer codes, this unitary just implements a linear transformation in the phase basis of the k qubits, which can equally well be regarded as a linear transformation in the amplitude basis $\{|x\rangle\}_{x\in X_A}$. Therefore, just as in the original protocol, Alice applies a "classical" transformation of her system and sends one part of the output to Bob.

---

[8] Associating X with amplitude instead of phase contravenes the usual convention in the QECC literature, but better fits the notation of the current paper.

For his part, Bob can complete the state merging protocol by coherently implementing the decoder from the information reconciliation protocol, a construction of which based on the pretty good measurement is given in [46].

Finally, the number of entangled systems generated in the state merging protocol is equal to the number of systems left at Alice's side, or $|X_A| - H_{\max}(\widetilde{X}_A|X_B B)_\rho - 2\log\frac{1}{\epsilon} - 4$. Lemma 30 shows that this is in fact greater than $H_{\min}(X_A|R)_\rho - 2\log\frac{1}{\epsilon} - 4$. Thus, the stabilizer-based state merging protocol achieves the same costs as the state merging protocol of Lemma 2 (up to terms of order $\log\frac{1}{\epsilon}$).

b. Fixed IID Source The case of a fixed IID source also follows easily from our analysis. We can simply apply the one-shot protocol from Theorem 5 to the case of a fixed IID source and then invoke the asymptotic equipartition property for the smooth max-information and the smooth max-entropy. In this way, we provide an alternative proof of this special case that avoids the use of typical projectors and the operator Chernoff bound [60].

c. Instrument Compression In Winter's original paper on measurement compression, additional arguments were required to establish that a POVM (positive operator-valued measure) compression protocol can function as an instrument compression protocol, where for an instrument compression protocol, Alice and Bob receive the classical outcomes of the measurement while Alice obtains the post-measurement states (see Section V of [60]). We note that our protocol here already functions as an instrument compression protocol due to our use of the classical state splitting protocol as a coding primitive.

d. Universal Measurement Compression with Quantum Side Information We briefly mention that there is no point in considering a protocol for universal measurement compression with quantum side information (similar to the observation in Section 6.3 of [18]). In such a scenario, the receiver would obtain some quantum side information correlated with the state on which the measurement should be simulated (see [57] for the case of measurement compression with quantum side information for a fixed IID source). Though, since a universal protocol should simulate the measurement with respect to an *arbitrary* input state, a special case of this input is one in which the quantum side information and input state are in a product state. Thus, the universal protocol given here is suitable for this case. This occurs simply because our simulation is with respect to the diamond norm, and the diamond norm is known to be robust under tensoring with other systems upon which the channel of interest does not act.

Another way to see this is that one could imagine devising a protocol for which quantum side information is taken into account. Based on the results in Ref. [57], we would expect the rate

of classical communication for such a protocol to be equal to the following information quantity: $\max_\rho I(X : R \,|\, B)_{(\mathcal{I} \otimes \mathcal{M} \otimes \mathcal{I})(\rho)}$, where $\rho_{AB} \in \mathcal{S}(AB)$ is an input state with quantum side information in the system B, and $\rho_{RAB} \in \mathcal{V}(RAB)$ is a purification of ρ_{AB}. Though, as shown in Theorem 16 of [18], the above information quantity is actually equal to the information quantity in (5), so that there is no improvement in the communication rate from the availability of quantum side information.

VII. CONCLUSION

We have justified the information-theoretic measure in (5) as quantifying the information gain of a quantum measurement, by providing an operational interpretation in terms of a protocol for universal measurement compression. The main tools used to prove this result are the post-selection technique for quantum channels and a novel classical state splitting protocol based on permutation-based extractors.

There are a number of open questions to consider going forward from here. Given that there are applications of "information gain" or "entropy reduction" in thermodynamics [35] and quantum feedback control [21], it would be interesting to explore whether the quantity in (5) has some application in these domains. Also, Buscemi *et al.* showed that the static measure of information gain in (3) plays a role in quantifying the trade-off between information extraction and disturbance [9], and it would be interesting to determine if there is a role in this setting for the information quantity in (5).

ACKNOWLEDGMENTS

Appendix A: Entropies

Lemma 11. *[47, Lemma 3.1.10] Let $\rho_{AB} \in \mathcal{S}_{\leq}(\mathcal{H}_{AB})$. Then we have that*

$$H_{\min}(A|B)_\rho \geq H_{\min}(AB)_\rho - H_0(B)_\rho \ . \tag{A1}$$

The max-mutual information is monotone under local operations.

Lemma 12. *[7, Lemma B.14] Let $\rho_{AB} \in \mathcal{S}_{\leq}(AB)$, and let $\mathcal{E}$ be a quantum channel of the form $\mathcal{E} = \mathcal{E}_A \otimes \mathcal{E}_B$. Then we have that*

$$I_{\max}(A:B)_\rho \geq I_{\max}(A:B)_{\mathcal{E}(\rho)} \ . \tag{A2}$$

The max-information can be upper and lower bounded in terms of entropies.

Lemma 13. *[7, Lemma B.10] Let $\rho_{AB} \in \mathcal{S}_{\leq}(AB)$. Then we have that*

$$H_R(A)_\rho - H_{\min}(A|B)_\rho \geq I_{\max}(A:B)_\rho \geq H_{\min}(A)_\rho - H_{\min}(A|B)_\rho \ , \tag{A3}$$

$H_R(\rho)$ is defined as the negative logarithm of the smallest eigenvalue of ρ on its support [7].

The following lemma is about the behavior of the max-information under projective measurements.

Lemma 14. *[7, Corollary B.16] Let $\rho_{AB} \in \mathcal{S}_{\leq}(AB)$, and let $P = \left\{ P_A^i \right\}_{i \in I}$ be a collection of projectors that describe a projective measurement on A. For $\mathrm{tr}\left[P_A^i \rho_A \right] \neq 0$, let $p_i = \mathrm{tr}\left[P_A^i \rho_A \right]$, and $\rho_{AB}^i = p_i^{-1} \cdot P_A^i \rho_{AB} P_A^i$. Then we have that*

$$I_{\max}(A:B)_\rho \geq \max_i I_{\max}(A:B)_{\rho^i} \ , \tag{A4}$$

where the maximum ranges over all i for which ρ_{AB}^i is defined.

Lemma 15. *Let $\varepsilon \geq 0$, and let $\rho_{XR} \in \mathcal{S}(XR)$ be classical on X with respect to the basis $\{|x\rangle\}_{x \in X}$. Then there exists $\bar{\rho}_{XR} \in \mathcal{B}^\varepsilon(\rho_{XR})$ classical on X with respect to the basis $\{|x\rangle\}_{x \in X}$ such that*

$$I_{\max}^\varepsilon(X:R)_\rho = I_{\max}(X:R)_{\bar{\rho}} \ . \tag{A5}$$

Proof. This is standard and can be proven exactly as in [52, Proposition 5.8]. □

We need the following monotonicity of the max-information.

Lemma 16. *Let $\rho_{AR} \in \mathcal{S}(AR)$, and $\Pi_A \in \mathcal{P}(A)$ with $\Pi_A \leq \mathbb{1}_A$. Then we have that*

$$I_{\max}(A:R)_\rho \geq I_{\max}(A:R)_{\Pi\rho\Pi} \ . \tag{A6}$$

Proof. Let $\sigma_R \in \mathcal{S}(R)$, and let $\lambda \in \mathbb{R}$ be such that $I_{\max}(A : R)_\rho = D_{\max}(\rho_{AR} \| \rho_A \otimes \sigma_R) = \log \lambda$. Then we have that $\lambda \cdot \rho_A \otimes \sigma_R \geq \rho_{AR}$, and with this

$$\lambda \cdot \rho_A \otimes \sigma_R \geq \lambda \cdot \Pi_A \rho_A \Pi_A \otimes \sigma_R \geq \Pi_A \rho_{AR} \Pi_A . \tag{A7}$$

Hence, we have $\log \lambda \geq D_{\max}(\Pi_A \rho_{AR} \Pi_A \| \Pi_A \rho_A \Pi_A \otimes \sigma_R) \geq I_{\max}(A : R)_{\Pi \rho \Pi}$. $\qquad \square$

The following is a bound on the increase of the smooth max-information when an additional subsystem is added.

Lemma 17. *[7, Lemma B.9] Let $\varepsilon \geq 0$, and let $\rho_{ABR} \in \mathcal{S}(ABR)$. Then we have that*

$$I_{\max}^\varepsilon(A : BR)_\rho \leq I_{\max}^\varepsilon(A : B)_\rho + 2 \cdot \log |R| . \tag{A8}$$

The following is a strengthening of the bound in Lemma 17 when the additional system is classical.

Lemma 18. *Let $\rho_{ABX} \in \mathcal{S}(ABX)$ be classical on $\mathcal{H}_X$ with respect to the basis $\{|x\rangle\}_{x \in X}$. Then we have that*

$$I_{\max}(A : BX)_\rho \leq I_{\max}(A : B)_\rho + \log |X| . \tag{A9}$$

Proof. Let $\sigma_B \in \mathcal{S}(B)$ be such that

$$I_{\max}(A : B)_\rho = D_{\max}(\rho_{AB} \| \rho_A \otimes \sigma_B) = \log \mu , \tag{A10}$$

that is, $\mu \in \mathbb{R}$ is minimal such that $\mu \cdot \rho_A \otimes \sigma_B \geq \rho_{AB}$. This implies $\mu \cdot \rho_A \otimes \sigma_B \otimes \frac{\mathbb{1}_X}{|X|} \geq \frac{1}{|X|} \cdot \rho_{AB} \otimes \mathbb{1}_X$. But we have by [47, Lemma 3.1.9] that $\rho_{AB} \otimes \mathbb{1}_X \geq \rho_{ABC}$, and hence $\mu \cdot \rho_A \otimes \sigma_B \otimes \frac{\mathbb{1}_X}{|X|} \geq \frac{1}{|X|} \cdot \rho_{ABX}$. Now, let $\lambda \in \mathbb{R}$ be minimal such that $\lambda \cdot \rho_A \otimes \sigma_B \otimes \frac{\mathbb{1}_X}{|X|} \geq \rho_{ABX}$. Thus, it follows that $\lambda \leq \mu \cdot |X|$, and from this we get

$$I_{\max}(A : BX)_\rho \leq D_{\max}(\rho_{ABX} \| \rho_A \otimes \sigma_B \otimes \frac{\mathbb{1}_X}{|X|}) = \log \lambda \tag{A11}$$

$$\leq D_{\max}(\rho_{AB} \| \rho_A \otimes \sigma_B) + \log |X| = I_{\max}(A : B)_\rho + \log |X| . \tag{A12}$$

$$\square$$

The smooth max-information is quasi-convex in its argument in the following sense.

Lemma 19. *[7, Lemma B.18] Let $\varepsilon \geq 0$, and let $\rho_{AB} = \sum_{i \in I} p_i \rho_{AB}^i \in \mathcal{S}_\leq(AB)$ with $\rho_{AB}^i \in \mathcal{S}_\leq(AB)$ for $i \in I$. Then we have that*

$$I_{\max}^\varepsilon(A : B)_\rho \leq \max_{i \in I} I_{\max}^\varepsilon(A : B)_{\rho^i} + \log |I| . \tag{A13}$$

The following is a quasi-convexity property of the zero-Rényi entropy.

Lemma 20. [4, Lemma 26] Let $\varepsilon \geq 0$, and let $\rho_A = \sum_{j=1}^N p_j \rho_A^j \in \mathcal{S}(A)$ with $p_j > 0$ for $j = 1, \ldots, N$. Then we have that

$$H_0^\varepsilon(A)_\rho \leq \max_j H_0^\varepsilon(A)_{\rho^j} + \log N \ . \tag{A14}$$

The smooth max-entropy and smooth zero-Rényi entropy are equivalent in the following sense.

Lemma 21. Let $\varepsilon > 0$, $\varepsilon' \geq 0$, and $\rho_A \in \mathcal{S}(A)$. Then we have that

$$H_0^{\varepsilon'}(A)_\rho \geq H_{\max}^{\varepsilon'}(A)_\rho > H_0^{\varepsilon'+\sqrt{2\varepsilon}}(A)_\rho - 2 \cdot \log \frac{1}{\varepsilon} \ . \tag{A15}$$

Proof. Since the (unconditional) max-entropy is the Rényi entropy of order $1/2$, the first inequality just follows from the ordering of the Rényi entropies [42, 48].

The idea for the proof of the second inequality is from the supplementary material [6, Lemma 13]. Let $\sigma_A \in \mathcal{B}^{\varepsilon'}(\rho_A)$ such that $H_{\max}^{\varepsilon'}(A)_\rho = H_{\max}(A)_\sigma$, and let $\sigma_A = \sum_i t_i |i\rangle\langle i|_A$ be a spectral decomposition of σ_A where the eigenvalues t_i are ordered non-increasingly. Define the projector $\Pi_A^k = \sum_{i \geq k} |i\rangle\langle i|_A$, let j be the smallest index such that $\mathrm{tr}\left[\Pi_A^j \sigma_A\right] \leq \varepsilon$, and define $\Pi_A = \mathbb{1}_A - \Pi_A^j$ as well as $\bar{\sigma}_A = \Pi_A \sigma_A \Pi_A$. By [6, Lemma 13] we have

$$H_{\max}(A)_\sigma > -\log \sup\{\lambda : \bar{\sigma}_A \geq \lambda \cdot \bar{\sigma}_A^0\} - 2 \cdot \log \frac{1}{\varepsilon} \geq \log \mathrm{tr}\left[\bar{\sigma}_A^0\right] - 2 \cdot \log \frac{1}{\varepsilon} \tag{A16}$$

$$= H_0(A)_{\bar{\sigma}} - 2 \cdot \log \frac{1}{\varepsilon} \ , \tag{A17}$$

and furthermore

$$P(\bar{\sigma}_A, \rho_A) \leq P(\sigma_A, \rho_A) + P(\bar{\sigma}_A, \sigma_A) \leq \varepsilon' + P(\Pi_A \sigma_A \Pi_A, \sigma_A) \leq \varepsilon' + \sqrt{1 - \left(\mathrm{tr}\left[\Pi_A^2 \sigma_A\right]\right)^2} \tag{A18}$$

$$\leq \varepsilon' + \sqrt{1 - (1 - \varepsilon)^2} \leq \varepsilon' + \sqrt{2\varepsilon} \ , \tag{A19}$$

where we used the triangle inequality for the purified distance, and a gentle measurement lemma for the purified distance (Lemma 26). Thus, we have

$$H_{\max}^{\varepsilon'}(A)_\rho = H_{\max}(A)_\sigma > H_0(A)_{\bar{\sigma}} - 2 \cdot \log \frac{1}{\varepsilon} \geq H_0^{\varepsilon'+\sqrt{2\varepsilon}}(A)_\rho - 2 \cdot \log \frac{1}{\varepsilon} \ . \tag{A20}$$

$$\square$$

The zero-Rényi entropy can be smoothed by applying a projection.

Lemma 22. *Let $\varepsilon \geq 0$, and let $\rho_A \in \mathcal{S}(A)$. Then there exists $\Pi_A \in \mathcal{P}(A)$ with $\Pi_A \leq \mathbb{1}_A$, diagonal in any eigenbasis of ρ_A,*

$$H_0^\varepsilon(A)_\rho \geq H_0(A)_{\Pi\rho\Pi} \ , \tag{A21}$$

and $\Pi_A \rho_A \Pi_A \in \mathcal{B}^{\sqrt{4\varepsilon}}(\rho_A)$.

Proof. The idea for the proof is from the supplementary material [6, Lemma 14]. Let $\sigma_A \in \mathcal{B}^\varepsilon(\rho_A)$ such that $H_0^\varepsilon(A)_\rho = H_0(A)_\sigma$. It follows from the supplementary material [6, Lemma 8], that σ_A can be taken to be diagonal in any eigenbasis of ρ_A. Define

$$\bar{\sigma}_A = \sigma_A - \{\sigma_A - \rho_A\}_+ = \rho_A - \{\rho_A - \sigma_A\}_+ \ , \tag{A22}$$

where $\{\cdot\}$ denotes the positive part of an operator. This implies $\bar{\sigma}_A \leq \sigma_A$, and we then have $H_0^\varepsilon(A)_\rho \geq H_0(A)_{\bar\sigma}$. Since $\bar{\sigma}_A$ and ρ_A also have the same eigenbasis, it follows that there exists $\Pi_A \in \mathcal{P}(A)$ with $\Pi_A \leq \mathbb{1}_A$ such that $\bar{\sigma}_A = \Pi_A \rho_A \Pi_A$. Furthermore, we get by the equivalence of the trace distance and the purified distance (Lemma 24) that

$$P(\rho_A, \bar{\sigma}_A) \leq \sqrt{\|\rho_A - \bar{\sigma}_A\|_1 + |\mathrm{tr}[\rho_A] - \mathrm{tr}[\bar{\sigma}_A]|} = \sqrt{2 \cdot \mathrm{tr}\left[\{\rho_A - \sigma_A\}_+\right]} \leq \sqrt{2 \cdot \|\rho_A - \sigma_A\|_1} \quad \text{(A23)}$$

$$\leq \sqrt{4 \cdot P(\rho_A, \sigma_A)} \leq \sqrt{4\varepsilon} \ . \tag{A24}$$

$\square$

The fully quantum asymptotic equipartition property for the smooth max-information and the smooth max-entropy is as follows.

Lemma 23. *[7, Lemma B.21][53, Theorem 9] Let $\varepsilon > 0$, $n \geq 2 \cdot (1 - \varepsilon^2)$, and $\rho_{AB} \in \mathcal{S}(AB)$. Then we have that*

$$\frac{1}{n} I_{\max}^\varepsilon(A : B)_{\rho^{\otimes n}} \leq I(A : B)_\rho + \frac{\xi(\varepsilon)}{\sqrt{n}} - \frac{2}{n} \cdot \log \frac{\varepsilon^2}{24} \tag{A25}$$

$$\frac{1}{n} H_{\max}^\varepsilon(A)_{\rho^{\otimes n}} \leq H(A)_\rho + \frac{\eta(\varepsilon)}{\sqrt{n}} \ , \tag{A26}$$

where $\xi(\varepsilon) = 8\sqrt{13 - 4 \cdot \log \varepsilon} \cdot (2 + \frac{1}{2} \cdot \log |A|)$, and $\eta(\varepsilon) = 4\sqrt{1 - 2 \cdot \log \varepsilon} \cdot (2 + \frac{1}{2} \cdot \log |A|)$.

Appendix B: Misc Lemmas

The following gives lower and upper bounds to the purified distance in terms of the trace distance.

Lemma 24. *[54, Lemma 6] Let $\rho, \sigma \in \mathcal{S}_{\leq}(A)$. Then we have that*

$$\frac{1}{2} \cdot \|\rho_A - \sigma_A\|_1 \leq P(\rho_A, \sigma_A) \leq \sqrt{\|\rho_A - \sigma_A\|_1 + |\mathrm{tr}[\rho_A] - \mathrm{tr}[\sigma_A]|} \ . \tag{B1}$$

The purified distance is convex in its arguments in the following sense.

Lemma 25. *[7, Lemma A.3] Let ρ_A^i, $\sigma_A^i \in \mathcal{S}_{\leq}(A)$ be with $\rho_A^i \approx_\varepsilon \sigma_A^i$ for $i \in I$, and $\{p_i\}_{i \in I}$ a probability distribution. Then we have that*

$$\sum_{i \in I} p_i \rho_A^i \approx_\varepsilon \sum_{i \in I} p_i \sigma_A^i \ . \tag{B2}$$

The following is a gentle measurement lemma for the purified distance.

Lemma 26. *[6, Lemma 7] Let $\rho_A \in \mathcal{S}(A)$, and $\Pi_A \in \mathcal{P}(A)$ with $\Pi_A \leq \mathbb{1}_A$. Then we have that*

$$P(\rho_A, \Pi_A \rho_A \Pi_A) \leq \sqrt{1 - \left(\mathrm{tr}\left[\Pi_A^2 \rho_A\right]\right)^2} \ . \tag{B3}$$

Appendix C: Extractors Based on Permutations

The following proposition concerns permutation-based extractors (operations that extract uniform randomness independent of an adversary's information), and it is critical in establishing our protocol for state merging of classically coherent states.

Proposition 27. *[51, Section 5.2] Let $\rho_{XR} \in \mathcal{S}(XR)$ be classical on X with respect to $\{|x\rangle\}_{x \in X}$, and $X = X_1 X_2$. Then we have that*

$$\frac{1}{|X|!} \cdot \sum_{P_X \in \mathbb{P}(X)} \left\| \mathrm{tr}_{X_2}\left[(P_X \otimes \mathbb{1}_R) \rho_{XR} (P_X^\dagger \otimes \mathbb{1}_R) \right] - \frac{\mathbb{1}_{X_1}}{|X_1|} \otimes \rho_R \right\|_1 \leq \sqrt{|X_1| \cdot 2^{-H_{\min}(X|R)_\rho}} \ , \tag{C1}$$

where $\mathbb{P}(X)$ denotes the group of permutations matrices on $\mathcal{H}_X$ with respect to $\{|x\rangle\}_{x \in X}$, defined as $P(\pi)|x\rangle = |\pi(x)\rangle$ for $\pi \in S_{|X|}$, the symmetric group on $\{1, 2, \ldots, |X|\}$.

Appendix D: The Post-Selection Technique

The following proposition lies at the heart of the post-selection technique for quantum channels.

Proposition 28. *[10] Let $\varepsilon > 0$, and let $\mathcal{E}_A^n$ and $\mathcal{F}_A^n$ be quantum channels from $\mathcal{L}(A^{\otimes n})$ to $\mathcal{L}(B)$. If there exists a quantum channel K_π for any permutation π such that $(\mathcal{E}_A^n - \mathcal{F}_A^n) \circ \pi = K_\pi \circ (\mathcal{E}_A^n - \mathcal{F}_A^n)$, then $\mathcal{E}_A^n$ and $\mathcal{F}_A^n$ are ε-close whenever*

$$\|((\mathcal{E}_A^n - \mathcal{F}_A^n) \otimes \mathcal{I}_{RR'})(\zeta_{ARR'}^n)\|_1 \leq \varepsilon (n+1)^{-(|A|^2 - 1)} \ , \tag{D1}$$

where $\zeta_{ARR'}^n$ is a purification of the de Finetti state $\zeta_{AR}^n = \int \sigma_{AR}^{\otimes n} d(\sigma_{AR})$ with $\sigma_{AR} \in \mathcal{V}(AR)$, $A \cong R$ and $d(\cdot)$ the measure on the normalized pure states on AR induced by the Haar measure on the unitary group acting on AR, normalized to $\int d(\cdot) = 1$. Furthermore, we can assume without loss of generality that $|R'| \leq (n+1)^{|A|^2-1}$.

A straightforward application of Carathéodory's theorem gives the following.

Lemma 29. *[7, Corollary D.6] Let $\zeta_{AR}^n = \int \sigma_{AR}^{\otimes n} d(\sigma_{AR})$ as in Proposition 28. Then we have that $\zeta_{AR}^n = \sum_i p_i (\omega_{AR}^i)^{\otimes n}$ with $\omega_{AR}^i \in \mathcal{V}(AR)$, $i \in \{1, 2, \ldots, (n+1)^{2|A||R|-2}\}$, and $\{p_i\}$ a probability distribution.*

Appendix E: Typical Projectors

A sequence x^n is typical with respect to some probability distribution $p_X(x)$ if its empirical distribution has maximum deviation δ from $p_X(x)$. The typical set $T_\delta^{X^n}$ is the set of all such sequences:

$$T_\delta^{X^n} \equiv \left\{ x^n : \left| \frac{1}{n} N(x|x^n) - p_X(x) \right| \leq \delta \quad \forall x \in \mathcal{X} \right\}, \tag{E1}$$

where $N(x|x^n)$ counts the number of occurrences of the letter x in the sequence x^n. The above notion of typicality is the "strong" notion (as opposed to the weaker "entropic" version of typicality sometimes employed [12]). The typical set enjoys three useful properties: its probability approaches unity in the large n limit, it has exponentially smaller cardinality than the set of all sequences, and every sequence in the typical set has approximately uniform probability. That is, suppose that X^n is a random variable distributed according to $p_{X^n}(x^n) \equiv p_X(x_1) \cdots p_X(x_n)$, ϵ is positive number that becomes arbitrarily small as n becomes large, and c is some positive constant. Then the following three properties hold [12]

$$\Pr\{X^n \in T_\delta^{X^n}\} \geq 1 - \epsilon, \tag{E2}$$

$$\left| T_\delta^{X^n} \right| \leq 2^{n[H(X)+c\delta]}, \tag{E3}$$

$$\forall x^n \in T_\delta^{X^n} : \quad 2^{-n[H(X)+c\delta]} \leq p_{X^n}(x^n) \leq 2^{-n[H(X)-c\delta]}. \tag{E4}$$

These properties translate straightforwardly to the quantum setting by applying the spectral theorem to a density operator ρ. That is, suppose that

$$\rho \equiv \sum_x p_X(x) |x\rangle\langle x|, \tag{E5}$$

for some orthonormal basis $\{|x\rangle\}_x$. Then there is a typical subspace defined as follows:

$$T_{\rho,\delta}^n \equiv \operatorname{span}\left\{|x^n\rangle : \left|\frac{1}{n}N(x|x^n) - p_X(x)\right| \leq \delta \quad \forall x \in \mathcal{X}\right\}, \tag{E6}$$

and let $\Pi_{\rho,\delta}^n$ denote the projector onto it. Then properties analogous to (E2-E4) hold for the typical subspace. The probability that a tensor power state $\rho^{\otimes n}$ is in the typical subspace approaches unity as n becomes large, the rank of the typical projector is exponentially smaller than the rank of the full n-fold tensor-product Hilbert space of $\rho^{\otimes n}$, and the state $\rho^{\otimes n}$ "looks" approximately maximally mixed on the typical subspace:

$$\operatorname{Tr}\left\{\Pi_{\rho,\delta}^n \, \rho^{\otimes n}\right\} \geq 1 - \epsilon, \tag{E7}$$

$$\operatorname{Tr}\left\{\Pi_{\rho,\delta}^n\right\} \leq 2^{n[H(B)+c\delta]}, \tag{E8}$$

$$2^{-n[H(B)+c\delta]} \, \Pi_{\rho,\delta}^n \leq \Pi_{\rho,\delta}^n \, \rho^{\otimes n} \, \Pi_{\rho,\delta}^n \leq 2^{-n[H(B)-c\delta]} \, \Pi_{\rho,\delta}^n, \tag{E9}$$

where $H(B)$ is the entropy of ρ.

Suppose now that we have an ensemble of the form $\{p_X(x), \rho_x\}$, and suppose that we generate a typical sequence x^n according to a "pruned" distribution (defined as a normalized version of $p_{X^n}(x^n)$ with support on its typical set and zero otherwise), leading to a tensor product state $\rho_{x^n} \equiv \rho_{x_1} \otimes \cdots \otimes \rho_{x_n}$. Then there is a conditionally typical subspace with a conditionally typical projector defined as follows:

$$\Pi_{\rho_{x^n},\delta}^n \equiv \bigotimes_{x \in \mathcal{X}} \Pi_{\rho_x,\delta}^{I_x}, \tag{E10}$$

where $I_x \equiv \{i : x_i = x\}$ is an indicator set that selects the indices i in the sequence x^n for which the ith symbol x_i is equal to $x \in \mathcal{X}$ and $\Pi_{\rho_x,\delta}^{I_x}$ is the typical projector for the state ρ_x. The conditionally typical subspace has the three following properties:

$$\operatorname{Tr}\left\{\Pi_{\rho_{x^n},\delta}^n \, \rho_{x^n}\right\} \geq 1 - \epsilon, \tag{E11}$$

$$\operatorname{Tr}\left\{\Pi_{\rho_{x^n},\delta}^n\right\} \leq 2^{n[H(B|X)+c\delta]}, \tag{E12}$$

$$2^{-n[H(B|X)+c\delta]} \, \Pi_{\rho_{x^n},\delta}^n \leq \Pi_{\rho_{x^n},\delta}^n \, \rho_{x^n} \, \Pi_{\rho_{x^n},\delta}^n \leq 2^{-n[H(B|X)-c\delta]} \, \Pi_{\rho_{x^n},\delta}^n, \tag{E13}$$

where $H(B|X) = \sum_x p_X(x)H(\rho_x)$ is the conditional quantum entropy.

Let ρ be the expected density operator of the ensemble $\{p_X(x), \rho_x\}$ so that $\rho = \sum_x p_X(x)\rho_x$. The following properties are proved in Refs. [19, 56, 58]:

$$\forall x^n \in T_\delta^{X^n} : \operatorname{Tr}\{\rho_{x^n} \, \Pi_\rho\} \geq 1 - \epsilon,$$

$$\sum_{x^n} p'_{X'^n}(x)\rho_{x^n} \leq [1-\epsilon]^{-1}\rho^{\otimes n}. \tag{E14}$$

In order to justify some of the estimates made in Section V B, we use the above estimates on eigenvalues and support sizes. For the classical communication cost, we consider

$$H_R(W^n)_{\bar{\gamma}^i} - H_{\min}(W^n R^n)_{\bar{\gamma}^i} + H_0(R^n)_{\bar{\gamma}^i}. \tag{E15}$$

The smallest nonzero eigenvalue of the reduced state on W^n is larger than $2^{-n[H(W)+c\delta]}$ due to the typical projection on W^n. Thus, we have that

$$H_R(W^n) \leq n[H(W) + c\delta]. \tag{E16}$$

The largest eigenvalue of $\bar{\gamma}_{WR}^{i,n}$ is bounded by

$$2^{-n\left[H(W)_{\bar{\gamma}^i} - c\delta\right]} 2^{-n\left[H(R|W)_{\bar{\gamma}^i} - c\delta\right]}, \tag{E17}$$

due to the typical projection on W^n and the conditionally typical projection on R^n. So we have that

$$H_{\min}(W^n R^n)_{\bar{\gamma}^i} \geq n\left[H(WR)_{\bar{\gamma}^i} + 2c\delta\right]. \tag{E18}$$

The size of the support of R^n is bounded from above by

$$2^{n\left[H(R)_{\bar{\gamma}^i} + \delta\right]}, \tag{E19}$$

due to the outermost projection on R^n. Thus, we have that

$$H_0(R^n)_{\bar{\gamma}^i} \leq n\left[H(R)_{\bar{\gamma}^i} + 2c\delta\right]. \tag{E20}$$

The above development then gives the following bound:

$$H_R(W^n)_{\bar{\gamma}^i} - H_{\min}(W^n R^n)_{\bar{\gamma}^i} + H_0(R^n)_{\bar{\gamma}^i} \leq n\left[I(W;R)_{\bar{\gamma}^i} + 5c\delta\right]. \tag{E21}$$

We have similar arguments for bounding the shared randomness cost:

$$H_R(W^n)_{\bar{\gamma}^i} - H_{\min}(W^n X^n R^n)_{\bar{\gamma}^i} + H_0(R^n X^n)_{\bar{\gamma}^i}. \tag{E22}$$

By the same argument as above, we have that

$$H_R(W^n)_{\bar{\gamma}^i} \leq n\left[H(W)_{\bar{\gamma}^i} + c\delta\right]. \tag{E23}$$

The largest eigenvalue of $\bar{\gamma}_{WXR}^{i,n}$ is bounded by

$$2^{-n\left[H(W)_{\bar{\gamma}^i} - c\delta\right]} 2^{-n\left[H(X|W)_{\bar{\gamma}^i} - c\delta\right]} 2^{-n\left[H(R|W)_{\bar{\gamma}^i} - c\delta\right]} \tag{E24}$$

$$= 2^{-n\left[H(WX)_{\bar{\gamma}^i} - 2c\delta\right]} 2^{-n\left[H(R|WX)_{\bar{\gamma}^i} - c\delta\right]} \tag{E25}$$

$$= 2^{-n\left[H(WXR)_{\bar{\gamma}^i} - 3c\delta\right]}, \tag{E26}$$

where we have used the fact that $H(R|W)_{\bar{\gamma}i} = H(R|WX)_{\bar{\gamma}i}$ because the state on R is independent of X. Thus, we have that

$$H_{\min}(W^n X^n R^n)_{\bar{\gamma}i} \geq n\Big[H(WXR)_{\bar{\gamma}i} - 3c\delta\Big]. \tag{E27}$$

Finally, the support of $R^n X^n$ is bounded again by $2^{n\big[H(RX)_{\bar{\gamma}i}+2c\delta\big]}$, due to the typical projections, so that we have

$$H_0(R^n X^n)_{\bar{\gamma}i} \leq n\Big[H(RX)_{\bar{\gamma}i} + 2c\delta\Big]. \tag{E28}$$

The above development then gives the following bound:

$$H_R(W^n)_{\bar{\gamma}i} - H_{\min}(W^n X^n R^n)_{\bar{\gamma}i} + H_0(R^n X^n)_{\bar{\gamma}i} \leq n\Big[I(W;XR)_{\bar{\gamma}i} + 6c\delta\Big]. \tag{E29}$$

Appendix F: Uncertainty Relation

Lemma 30. *For every* $|\psi\rangle\langle\psi|_{ABR} \in \mathcal{V}(ABR)$ *and observable (measurement)* Z_A, *we have that*

$$H_{\min}(A|B)_\psi + H_{\max}(Z_A|R)_\psi \leq \log|A|, \tag{F1}$$

Proof. Define $\lambda = H_{\min}(A|B)_\psi$ and let $\sigma_B \in \mathcal{S}(B)$ be such that

$$\psi_{AB} \leq 2^{-\lambda}\mathbb{1}_A \otimes \sigma_B. \tag{F2}$$

The measurement procedure can be described by an isometry $U_{A\to ZA}$ whose action is specified by $U_{A\to ZA}|z\rangle_A = |z\rangle_Z|z\rangle_A$, where $\{|z\rangle\}$ are the basis states associated with the (projective) measurement. Applied to ψ_{AB} this yields

$$\xi_{ZAB} = U_{A\to ZA}\psi_{AB}U^\dagger_{A\to ZA} \tag{F3}$$

$$\leq 2^{-\lambda} U_{A\to ZA}(\mathbb{1}_A \otimes \sigma_B)U^\dagger_{A\to ZA} \tag{F4}$$

$$= 2^{-\lambda}\sum_z |z\rangle\langle z|_Z \otimes |z\rangle\langle z|_A \otimes \sigma_B \tag{F5}$$

$$\leq 2^{-\lambda}\mathbb{1}_{ZA} \otimes \sigma_B \tag{F6}$$

$$= 2^{-(\lambda-\log|A|)}\mathbb{1}_Z \otimes \pi_A \otimes \sigma_B, \tag{F7}$$

where $\pi_A = \mathbb{1}_A/|A|$. Thus, $\mu = \lambda - \log|A|$ and $\pi_A \otimes \sigma_B$ are feasible for $H_{\min}(Z|AB)_\xi$, meaning

$$H_{\min}(Z|AB)_\xi \geq \lambda - \log|A| . \tag{F8}$$

Therefore the first claim follows, since $H_{\min}(Z|AB)_\xi = -H_{\max}(Z|R)_\xi = -H_{\max}(Z_A|R)_\psi$. $\qquad\square$

[1] *Albert Einstein: Philosopher-Scientist*, chapter Discussions with Einstein on Epistemological Problems in Atomic Physics, page 199. Cambridge University Press, 1949.

[2] Charles H. Bennett, Igor Devetak, Aram W. Harrow, Peter W. Shor, and Andreas Winter. The quantum reverse Shannon theorem. 2009. arXiv:0912.5537v2.

[3] Charles H. Bennett, Peter W. Shor, John A. Smolin, and Ashish V. Thapliyal. Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem. *IEEE Transactions on Information Theory*, 48:2637, 2002.

[4] Mario Berta, Fernando Brandao, Matthias Christandl, and Stephanie Wehner. Entanglement cost of quantum channels. arXiv:1108.5357v2, 2011.

[5] Mario Berta, Matthias Christandl, Fernando Brandao, and Stephanie Wehner. Entanglement cost of quantum channels. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, page 900, 2012.

[6] Mario Berta, Matthias Christandl, Roger Colbeck, Joseph M. Renes, and Renato Renner. The uncertainty principle in the presence of quantum memory. *Nature Physics*, 6:659, 2010.

[7] Mario Berta, Matthias Christandl, and Renato Renner. The quantum reverse Shannon theorem based on one-shot information theory. *Communications in Mathematical Physics*, 306:579, 2011.

[8] J.-C. Boileau and J. M. Renes. Optimal state merging without decoupling. In *Proceedings of the Fourth Workshop on Theory of Quantum Computation, Communication, and Cryptography*, volume 5906 of *Lecture Notes in Computer Science*, page 76, 2009.

[9] Francesco Buscemi, Masahito Hayashi, and Michal Horodecki. Global information balance in quantum measurements. *Physical Review Letters*, 100:210504, 2008.

[10] Matthias Christandl, Robert König, and Renato Renner. Post-selection technique for quantum channels with applications to quantum cryptography. *Physical Review Letters*, 102:020504, 2009.

[11] N. Ciganovic. Smooth max-mutual information as a generalization of von neumann mutual information for the one-shot setting. Master's thesis, ETH Zurich, 2012.

[12] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.

[13] Paul Cuff. Communication requirements for generating correlated random variables. In *Information Theory Proceedings (ISIT), 2008 IEEE International Symposium on*, page 1393, 2008.

[14] Paul Cuff. Distributed channel synthesis. 2012. arXiv:1208.4415v1.

[15] Nilanjana Datta. Min- and max- relative entropies and a new entanglement monotone. *IEEE Transactions on Information Theory*, 55:2816, 2009.

[16] Nilanjana Datta, Min-Hsiu Hsieh, and Mark M. Wilde. Quantum rate distortion, reverse Shannon theorems, and source-channel separation. *IEEE Transactions on Information Theory*, 59:615, January 2013.

[17] Nilanjana Datta, Min-Hsiu Hsieh, Mark M. Wilde, and Andreas Winter. Quantum-to-classical rate

distortion coding. 2012. arXiv:1210.6962v1.

[18] Nilanjana Datta, Mark M. Wilde, Min-Hsiu Hsieh, and Andreas Winter. Quantum rate distortion coding with auxiliary resources. 2012. arXiv:1212.5316v1.

[19] Igor Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Transactions on Information Theory*, 51:44–55, January 2005. arXiv:quant-ph/0304127.

[20] Paul Dirac. *The Principles of Quantum Mechanics (International Series of Monographs on Physics)*. Oxford University Press, 1982.

[21] Andrew C. Doherty, Kurt Jacobs, and Gerard Jungman. Information, disturbance, and hamiltonian quantum feedback control. *Physical Review A*, 63:062306, 2001.

[22] Frédéric Dupuis, Oleg Szehr, and Marco Tomamichel. A decoupling approach to classical data transmission over quantum channels. 2012. arXiv:1207.0067v1.

[23] Christopher A. Fuchs and Kurt Jacobs. Information-tradeoff relations for finite-strength quantum measurements. *Phys. Rev. A*, 63:062305, 2001.

[24] Hilbrand J. Groenewold. A problem of information gain by quantal measurements. *International Journal of Theoretical Physics*, 4:327, 1971.

[25] Aram Harrow. Coherent communication of classical messages. *Physical Review Letters*, 92:097902, 2004.

[26] Alexander S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problems of Information Transmission*, 9:177, 1973.

[27] Alexander S. Holevo. The capacity of the quantum communication channel with general signal states. *IEEE Transactions on Information Theory*, 44:269, 1998.

[28] Alexander S. Holevo. On entanglement assisted classical capacity. *Journal of Mathematical Physics*, 43:4326, 2002.

[29] Alexander S. Holevo. Information capacity of quantum observable. *Problems of Information Transmission*, 48:1, 2012.

[30] Michal Horodecki, Jonathan Oppenheim, and Andreas Winter. Partial quantum information. *Nature*, 436:673, 2005.

[31] Michal Horodecki, Jonathan Oppenheim, and Andreas Winter. Quantum state merging and negative information. *Communications in Mathematical Physics*, 269:107, 2006.

[32] Min-Hsiu Hsieh, Igor Devetak, and Andreas Winter. Entanglement-assisted capacity of quantum multiple-access channels. *IEEE Transactions on Information Theory*, 54:3078, 2008.

[33] Kurt Jacobs. On the properties of information gathering in quantum and classical measurements. 2003. arXiv:quant-ph/0304200v1.

[34] Kurt Jacobs. A bound on the mutual information, and properties of entropy reduction, for quantum channels with inefficient measurements. *Journal of Mathematical Physics*, 47:012102, 2006.

[35] Kurt Jacobs. Second law of thermodynamics and quantum feedback control: Maxwell's demon with weak measurements. *Physical Review A*, 80:012322, 2009.

[36] Richard Jozsa. Fidelity for mixed quantum states. *Journal of Modern Optics*, 41:2315, 1994.

[37] Alexei Y. Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52:1191, 1997.

[38] Göran Lindblad. An entropy inequality for quantum measurements. *Communications in Mathematical Physics*, 28:245, 1972.

[39] Shunlong Luo. Information conservation and entropy change in quantum measurements. *Physical Review A*, 82:052103, 2010.

[40] Hans Martens and WillemM. Muynck. Nonideal quantum measurements. *Foundations of Physics*, 20:255–281, 1990.

[41] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information.* Cambridge University Press, 2000.

[42] Masanori Ohya and Denes Petz. *Quantum Entropy and Its Use.* Springer, 1993.

[43] Masanao Ozawa. On information gain by quantum measurements of continuous observables. *Journal of Mathematical Physics*, 27:759, 1986.

[44] Vern I. Paulsen. *Completely bounded maps and operator algebras.* Cambridge University Press, 2002.

[45] J. M. Renes. The physics of quantum information: Complementarity, uncertainty, and entanglement, 2011. arXiv:1212.2379.

[46] J. M. Renes and R. Renner. One-shot classical data compression with quantum side information and the distillation of common randomness or secret keys. 2010. arXiv:1008.0452v2.

[47] Renato Renner. Security of quantum key distribution. *International Journal of Quantum Information*, 6:1, 2008.

[48] Alfréd Rényi. On measures of information and entropy. *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, page 547, 1960.

[49] Benjamin Schumacher and Michael D. Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56:131, 1997.

[50] Maksim E. Shirokov. Entropy reduction of quantum measurements. *Journal of Mathematical Physics*, 52:052202, 2011.

[51] Oleg Szehr. Decoupling theorems. Master's thesis, ETH Zurich, 2011. arXiv:1207.3927v1.

[52] Marco Tomamichel. *A Framework for Non-Asymptotic Quantum Information Theory.* PhD thesis, ETH Zurich, 2012. arXiv:1203.2142.

[53] Marco Tomamichel, Roger Colbeck, and Renato Renner. A fully quantum asymptotic equipartition property. *IEEE Transactions on Information Theory*, 55:5840, 2009.

[54] Marco Tomamichel, Roger Colbeck, and Renato Renner. Duality between smooth min- and max-entropies. *IEEE Transactions on Information Theory*, 56:4674, 2010.

[55] Armin Uhlmann. The transition probability in the state space of a *-algebra. *Report on Mathematical Physics*, 9:273, 1976.

[56] Mark M. Wilde. *From Classical to Quantum Shannon Theory.* June 2011. arXiv:1106.1445.

[57] Mark M. Wilde, Patrick Hayden, Francesco Buscemi, and Min-Hsiu Hsieh. The information-theoretic costs of simulating quantum measurements. *Journal of Physics A: Mathematical and Theoretical*, 45:453001, 2012.

[58] Andreas Winter. Coding theorem and strong converse for quantum channels. *IEEE Transactions on Information Theory*, 45(7):2481–2485, 1999.

[59] Andreas Winter. Compression of sources of probability distributions and density operators. 2002. arXiv:quant-ph/0208131v1.

[60] Andreas Winter. "Extrinsic" and "Intrinsic" data in quantum measurements: Asymptotic convex decomposition of positive operator valued measures. *Communications in Mathematical Physics*, 244:157, 2004.

[61] Wojciech H. Zurek. Decoherence and the transition from quantum to classical. *Physics Today*, 44:36, 1991.

[62] Wojciech H. Zurek. Quantum Darwinism. *Nature Physics*, 5:181, 2009.