IOPscience iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

The information-theoretic costs of simulating quantum measurements

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2012 J. Phys. A: Math. Theor. 45 453001

(http://iopscience.iop.org/1751-8121/45/45/453001)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 173.178.178.167
The article was downloaded on 19/10/2012 at 01:36

Please note that terms and conditions apply.

IOP Publishing Journal of Physics A: Mathematical and Theoretical

J. Phys. A: Math. Theor. **45** (2012) 453001 (67pp) doi:10.1088/1751-8113/45/45/453001

TOPICAL REVIEW

The information-theoretic costs of simulating quantum measurements

Mark M Wilde[1], Patrick Hayden[1], Francesco Buscemi[2] and Min-Hsiu Hsieh[3]

[1] School of Computer Science, McGill University, Montreal, Quebec, H3A 2A7, Canada
[2] Institute for Advanced Research, Nagoya University, Chikusa-ku, Nagoya 464-8601, Japan
[3] Centre for Quantum Computation and Intelligent Systems (QCIS), Faculty of Engineering and Information Technology, University of Technology, Sydney, NSW 2007, Australia

E-mail: mwilde@gmail.com, patrick@cs.mcgill.ca, buscemif@gmail.com and minhsiuh@gmail.com

Received 26 June 2012, in final form 18 September 2012
Published 18 October 2012
Online at stacks.iop.org/JPhysA/45/453001

Abstract

Winter's measurement compression theorem stands as one of the most penetrating insights of quantum information theory. In addition to making an original and profound statement about measurement in quantum theory, it also underlies several other general protocols used for entanglement distillation and local purity distillation. The theorem provides for an asymptotic decomposition of any quantum measurement into *noise* and *information*. This decomposition leads to an optimal protocol for having a sender simulate many independent instances of a quantum measurement and send the measurement outcomes to a receiver, using as little communication as possible. The protocol assumes that the parties have access to some amount of common randomness, which is a strictly weaker resource than classical communication. In this review, we provide a second look at Winter's measurement compression theorem, detailing the information processing task, giving examples for understanding it, reviewing Winter's achievability proof, and detailing a new approach to its single-letter converse theorem. We prove an extension of the theorem to the case in which the sender is not required to receive the outcomes of the simulated measurement. The total cost of common randomness and classical communication can be lower for such a 'non-feedback' simulation, and we prove a single-letter converse theorem demonstrating optimality. We then review the Devetak–Winter theorem on classical data compression with quantum side information, providing new proofs of its achievability and converse parts. From there, we outline a new protocol that we call 'measurement compression with quantum side information,' announced previously by two of us in our work on triple trade-offs in quantum Shannon theory. This protocol has several applications, including its part in the 'classically-assisted state redistribution' protocol, which is the most general protocol on the static side

 © 2012 IOP Publishing Ltd Printed in the UK & the USA

of the quantum information theory tree, and its role in reducing the classical communication cost in a task known as local purity distillation. We also outline a connection between measurement compression with quantum side information and recent work on entropic uncertainty relations in the presence of quantum memory. Finally, we prove a single-letter theorem characterizing measurement compression with quantum side information when the sender is not required to obtain the measurement outcome.

PACS numbers: 03.67.−a, 03.67.Hk, 03.65.Ta

(Some figures may appear in colour only in the online journal)

Contents

1. Introduction

Measurement plays an important role in quantum theory. It is the interface between the macroscopic world of everyday experience and the quantum world, which is characterized by

noncommutativity and superposition. The translation is imperfect, however, with superposition and noncommutativity leading necessarily to uncertainty in the outcomes of measurements. In any given measurement, there will be noise inherent to the measurement procedure, uncertainty due to the state being measured and, most importantly, *information*. The objective of this review is to explain how to separate out these components, precisely identifying and quantifying them in the data produced by a quantum measurement. To do so, it will be crucial to adopt an information-theoretic point of view, not just to provide the necessary techniques to solve the problem, but even to figure out how to properly formulate the question.

If we are only concerned with capturing the statistics of the outcomes of a quantum measurement, the most general mathematical description is to use the positive operator-valued measure (POVM) formalism [21, 34, 40]. In the POVM formalism, a quantum measurement is specified as a set $\Lambda \equiv \{\Lambda_x\}$ of operators indexed by some classical label x corresponding to the classical outcomes of the measurement. These operators should be positive and form a resolution of the identity on the Hilbert space of the system that is being measured:

$$\forall x : \Lambda_x \geqslant 0, \quad \sum_x \Lambda_x = I.$$

Given a quantum state described by a density operator ρ (a positive, unit trace operator) and a POVM Λ, a measurement of ρ specified by Λ induces a random variable X, and the probability $p_X(x)$ for the classical outcome x to occur is given by the Born rule:

$$p_X(x) = \mathrm{Tr}\{\Lambda_x \rho\}. \tag{1}$$

Positivity of the operators Λ_x and ρ guarantees positivity of the distribution $p_X(x)$, and that the set Λ forms a resolution of the identity and the density operator ρ has unit trace guarantees normalization of the distribution $p_X(x)$.

The above definition of a POVM makes it clear that the set of all POVMs is a convex set, i.e. given a POVM $\Lambda \equiv \{\Lambda_x\}$ and another $\Gamma \equiv \{\Gamma_x\}$, with $0 < \lambda < 1$, the convex combination $\lambda\Lambda + (1-\lambda)\,\Gamma \equiv \{\lambda\Lambda_x + (1-\lambda)\,\Gamma_x\}$ is also a POVM. The physical interpretation of this convexity is that it might be possible to decompose any particular measuring apparatus into noise and information. If an apparatus does not admit a decomposition of this form, then it is an extremal POVM, lying on the boundary of the convex set. If it does, however, as in the above example apparatus, one could first flip a biased coin with distribution $(\lambda, 1-\lambda)$ to determine whether to perform Λ or Γ and then perform the corresponding measurement. The coin flip is a source of noise because it is independent of the physical measurement outcome, and the distribution for the outcome corresponds to the information. Decomposing an apparatus in this way is a useful idea with many applications.

To develop a robust quantitative theory, however, it is surprisingly effective to consider the above ideas from an information-theoretic standpoint, in the sense of Shannon [57]. In this context, that approach will have three main features: a tolerance for small imperfections, a focus on asymptotics, and an emphasis on communication. To begin with, let us focus on the first two. From an operational point of view, there is little justification for requiring an exact convex decomposition of a given measurement. As long as any imperfections are very small, approximation by a convex decomposition leads to experimentally indistinguishable consequences. Moreover, measurement statistics are most meaningful in a setting in which the measurement is repeated many times on identical state preparations. As such, it is sensible, and remarkably powerful, to ask about approximate convex decomposition of repeated measurements, with the permissible imperfection required to vanish in the limit of infinite repetitions.

The relevance of communication is less immediate. In the example described above, the measurement $\lambda\Lambda + (1-\lambda)\Gamma$ could be implemented by first flipping a coin and then either

measuring Λ or Γ. This opens up the possibility of significantly compressing the measurement outcomes because there will generically be less uncertainty about the outcome of either Λ or Γ alone than the convex combination $\lambda\Lambda + (1-\lambda)\Gamma$. To formalize this notion, one could imagine that two parties, traditionally named Alice and Bob, are trying to collectively implement a measurement. They share some common random bits that can be used to perform the $(\lambda, 1-\lambda)$ coin flip without communicating, and Alice holds the quantum system on which $\lambda\Lambda + (1-\lambda)\Gamma$ is to be measured. Based on the result of the coin flip, Alice would apply either Λ or Γ and compress the outcome as much as possible, minimizing the number of bits she needs to send to Bob in order to allow him to reconstruct the outcome of the measurement. Optimizing the number of bits required over all possible measurement simulation strategies, of which we have only described one, then provides a robust operational measure of the amount of information generated by the quantum measurement.

In a seminal paper, Winter successfully performed this information-theoretic analysis of measurement, and in so doing, was able to make a profound and original statement about the nature of information in quantum measurement [65]. The content of his 'measurement compression theorem' is the specification of an optimal two-dimensional rate region, characterizing the resources needed for an asymptotically faithful simulation of a quantum measurement Λ on a state ρ in terms of common randomness and classical communication. The sender (Alice) and receiver (Bob) both obtain the outcome of the measurement, and as such, this is known as a 'feedback simulation' (terminology introduced in a different though related context [5]). His measurement compression protocol achieves one important optimal rate pair in this region: if, to first order, at least $nH(X|R)$ bits of common randomness are available, then it is possible to simulate the measurement $\Lambda^{\otimes n}$ on the state $\rho^{\otimes n}$ with only about $nI(X;R)$ bits of classical communication. We allow n, the number of repetitions of the measurement Λ, to go to infinity, in which limit the simulation becomes asymptotically faithful. The entropies $H(X|R)$ and $I(X;R)$ are defined as

$$H(X|R) \equiv H(XR) - H(R),$$
$$I(X;R) \equiv H(X) - H(X|R),$$

with the von Neumann entropy of a state σ defined as $H(\sigma) \equiv -\mathrm{Tr}\{\sigma \log_2 \sigma\}$. The above entropies are taken with respect to the state

$$\sum_x |x\rangle\langle x|^X \otimes \mathrm{Tr}_A\left\{\left(I^R \otimes \Lambda_x^A\right)\phi_\rho^{RA}\right\}, \tag{2}$$

where ϕ_ρ^{RA} is any purification of the state ρ, meaning that ϕ_ρ^{RA} is a rank-1 density operator satisfying $\mathrm{Tr}_R\left\{\phi_\rho^{RA}\right\} = \rho$.[4] One can think of (2) as the post-measurement state, including both the classical outcome of the measurement and the subsequent state of the reference system R. The other important rate pair corresponds to Shannon's protocol. If no common randomness is available and both the sender and receiver are to obtain the measurement outcome, then the lowest achievable rate of classical communication is $H(X)$, the Shannon entropy of the distribution of measurement outcomes in (1). Time-sharing between these two protocols, converting classical communication to common randomness, and wasting common randomness then give all other optimal rate pairs. (See figure 3 for an example plot of the region.)

[4] Here and throughout this review, we use superscripts such as A, B, R, and E to denote quantum systems with corresponding Hilbert spaces $\mathcal{H}_A$, $\mathcal{H}_B$, $\mathcal{H}_R$, and $\mathcal{H}_E$. Such a labeling is useful in quantum information theory because we often deal with states that are defined over many systems. We also use the shorthand $\phi \equiv |\phi\rangle\langle\phi|$ to denote a pure-state density operator. So, for example, the state ϕ_ρ^{RA} is shared between systems A and R, implying that ϕ_ρ^{RA} is an operator acting on the tensor-product Hilbert space $\mathcal{H}_R \otimes \mathcal{H}_A$. We also freely identify Roman capital letters W, X, Y, and Z with both random variables and quantum systems containing only classical data (as in (2)). There should be no confusion here because these entities are in direct correspondence.

Winter's measurement compression protocol has an important place in the constellation of quantum Shannon theoretic protocols. It evolved from earlier work in [46, 66], and it is the predecessor to the quantum reverse Shannon theorem, which was conjectured in [6] and proved later in [5, 10][5]. The quantum reverse Shannon theorem quantifies the noiseless resources required to simulate a noisy quantum channel. Since Winter's measurement compression theorem applies to a quantum measurement and a quantum measurement is a special type of quantum channel with quantum input and classical output, it is clear that the measurement compression protocol gives a special type of quantum reverse Shannon theorem. The quantum reverse Shannon theorem may seem on first encounter to correspond to a pointless task. After all, in the words of [6], why would we want to dilute fresh water into salt water? First appearances notwithstanding, it has at least two nearly immediate and significant information-theoretic applications: in proving strong converses [6, 64, 5, 10, 8] and in lossy data compression, otherwise known as rate distortion theory [64, 44, 43, 19]. The connection to strong converses follows from a *reductio ad absurdum* argument: if one were able to simulate a channel at a rate larger than its capacity, then it would be possible to bootstrap a channel code and a simulation code to achieve more communication than a noiseless channel would allow for. With the aid of an appropriate reverse Shannon theorem, one can then argue that coding at a rate beyond the capacity should make the error probability converge to one exponentially fast in the number of channel uses. The connection to rate distortion theory [7] follows from the observation that a reverse Shannon theorem achieves a task strictly stronger than the usual average distortion criterion considered in rate distortion theory. There, one requires that an information source be represented by the receiver up to some average distortion $D \geqslant 0$. If one were to simulate a channel on the information source that does not distort it by more than D on average, then clearly such a protocol would already satisfy the demands of rate distortion.

There are two other useful applications of Winter's measurement compression theorem. The first is in local purity distillation [36, 37, 22, 41], where the task is for two spatially separated parties to distill local pure states from an arbitrary bipartite mixed state ρ^{AB} by using only local unitary operations and classical communication. The measurement compression theorem is helpful in determining the classical communication cost of such protocols, as considered in [41]. Another application of measurement compression is in realizing the first step of the so-called grandmother protocol of quantum information theory [26], where the objective is to distill entanglement from a noisy bipartite state ρ^{AB} with the help of noiseless classical and quantum communication. It is possible to improve upon both of these protocols by exploiting one of the new measurement compression theorems that we outline in this review.

Once one takes the first step of splitting the implementation of a measurement between Alice and Bob, it becomes natural to consider different notions of simulation. What if only Bob needs to get the outcome of the measurement, not Alice? What if Bob holds a quantum system entangled with the system being measured? These and related variations provide a very precise and diverse set of tools for analyzing the dichotomy between noise and information in quantum measurements. Beyond providing a detailed review of Winter's theorem, the main contribution of this review will be to develop these variations and generalizations of his original theorem. More specifically, our contributions are as follows.

- We provide a full review of Winter's measurement compression theorem, detailing the basic information processing task, the statement of the theorem, Winter's achievability proof, and a simple converse theorem that demonstrates an optimal characterization of the rate region. We also review Winter's extension of the theorem to quantum instruments.

[5] We should clarify here that, while [5] appeared on the arXiv in 2009, that paper contains ideas developed and publicized by the authors over a nine year period starting with the publication of [6] in 2001. Reference [10] features a different proof from that in [5], but it exploits many of the important ingredients developed in [5].

- We extend Winter's measurement compression theorem to the setting in which the sender is not required to receive the outcome of the measurement simulation. Such a task is known as a 'non-feedback' simulation, in analogy with a similar setting in the quantum reverse Shannon theorem [5]. A benefit of a 'non-feedback' simulation is that the total cost of common randomness and classical communication can be lower than that of a 'feedback' simulation, leading to interesting, non-trivial trade-off curves for the rates of these resources. Also, we prove a single-letter converse theorem for this case, demonstrating that our protocol is optimal.
- We then review Devetak and Winter's theorem regarding classical data compression with quantum side information (CDC–QSI) [27]. The setting of the problem is that an information source distributes a random classical sequence to one party and a quantum state correlated with the sequence to another party. The objective is for the first party to transmit the classical sequence to the second party using as few noiseless classical bit channels as possible. As such, it is one particular quantum generalization of the classic Slepian–Wolf problem [58]. In the Slepian–Wolf protocol, the first party hashes the sequence received from the source, transmits the hash, and the second party uses his side information to search among all the sequences for any that are consistent with the hash and are a 'reasonable cause' for his side information. We provide a novel achievability proof for CDC–QSI that is a direct quantization of this strategy, replacing the latter search with binary-outcome quantum measurements. We also provide a simple converse proof that is along the lines of the standard converses in [15, 30].
- The above reviews of measurement compression and CDC–QSI then prepare us for another novel contribution: measurement compression in the presence of quantum side information (MC–QSI). The setting for this new protocol is that a sender and receiver share many copies of some bipartite state ρ^{AB}, and the sender would like to simulate the action of many independent and identical measurements on the A system according to some POVM Λ. The protocol is a 'feedback simulation,' such that the sender also obtains the outcomes of the measurement (though we still refer to it as MC–QSI for short). The MC–QSI protocol combines ideas from the measurement compression theorem and CDC–QSI in order to reduce the classical communication rate and common randomness needed to simulate the measurement. The idea is that Alice performs the measurement compression protocol as she would before, but she hashes the output of the simulated measurement and sends this along to Bob. Bob then searches among all the post-measurement states that are consistent with the hash and his share of the common randomness, similar to the way that he would in the CDC–QSI protocol. The result is a reduction in the classical communication and common randomness rate to $I(X;R|B)$ and $H(X|RB)$, respectively, where the entropies are with respect to the following state:

$$\sum_x |x\rangle\langle x|^X \otimes \mathrm{Tr}_A\left\{\left(I^{RB} \otimes \Lambda_x^A\right)\phi_\rho^{RBA}\right\},$$

  and ϕ_ρ^{RBA} is a purification of the state ρ^{AB}. These rates are what we would intuitively expect of such a protocol—they are the same as in Winter's original theorem, except the entropic quantities are conditioned on Bob's quantum side information in the system B.
- After developing MC–QSI, we briefly discuss three of its applications. The first is an application that two of us announced in [38] : MC–QSI along with state redistribution [29, 68] acts as a replacement for the 'grandmother' protocol discussed above. The resulting protocol uses less classical and quantum communication and can in fact generate the grandmother by combining it with entanglement distribution. As such, MC–QSI and state redistribution form the backbone of the best known 'static' protocols in quantum

Shannon theory (though, one should be aware that these results are only optimal up to a regularization, so it could very well be that further improvements are possible). The second application is an observation that the above protocol leads to a quantum reverse Shannon theorem for a quantum instrument, that is, a way to simulate the action of a quantum instrument on a quantum state by employing common randomness, classical communication, entanglement, and quantum communication. The third application is an improvement of the local purity distillation protocol from [41], so that we can lower the classical communication cost from $I(Y;BE)$ to $I(Y;E|B)$, as one should expect when taking quantum side information into account.

- We then discuss a way that we can relate recent work on entropic uncertainty relations with quantum side information [52, 9, 59, 14, 31] to provide a lower bound on the classical resources required in two different complementary MC–QSI protocols.
- Finally, we analyze the MC–QSI problem in the case where the sender is not required to receive the outcomes of the measurement simulation. For this non-feedback MC–QSI problem, we once again develop optimal protocols and find a single-letter characterization of the achievable rate region. While the necessary protocols are simply the natural combinations of those used in MC–QSI with those used for non-feedback MC, the optimality proof is different and remarkably subtle.

All the simulation theorems appearing in this review are 'single-letter,' meaning that we can calculate the optimal rate regions as simple entropic functions of one copy of the state or resource. This type of result occurs more often in quantum information theory when the resources considered are of a hybrid classical-quantum nature, as is our case here. The single-letter results here mean that we can claim to have a complete information-theoretic understanding of the tasks of MC, non-feedback MC, CDC–QSI, MC–QSI, and non-feedback MC–QSI.

2. Measurement compression

This section provides a detailed review of the main results in Winter's original paper on measurement compression [65], and it also serves to establish notation used in the rest of the paper. Consider a quantum state ρ and a POVM $\Lambda \equiv \{\Lambda_x\}_{x\in\mathcal{X}}$, such that $\Lambda_x \geqslant 0$ and $\sum_x \Lambda_x = I$. Measuring the POVM Λ on the state ρ induces a random variable X with the following distribution $p_X(x)$:

$$p_X(x) \equiv \mathrm{Tr}\{\Lambda_x \rho\}.$$

Suppose that a quantum information source outputs many copies of the state ρ and the POVM is performed many times, producing the IID distribution $p_{X^n}(x^n)$ (where $x^n = x_1 x_2 \ldots x_n$):

$$\begin{aligned} p_{X^n}(x^n) &\equiv \mathrm{Tr}\{\Lambda_{x^n} \rho^{\otimes n}\} \\ &= \mathrm{Tr}\{(\Lambda_{x_1} \otimes \Lambda_{x_2} \otimes \cdots \otimes \Lambda_{x_n})(\rho \otimes \rho \otimes \cdots \otimes \rho)\} \\ &= \prod_{i=1}^{n} \mathrm{Tr}\{\Lambda_{x_i} \rho\}. \end{aligned}$$

In order to communicate the result of the measurement to a receiver using a noiseless classical channel, one could compress the data sequence x^n using Shannon compression [15] and communicate the sequence x^n faithfully by transmitting only $nH(X)$ bits. Such a strategy is optimal if no other resource is shared between the sender and receiver. But supposing that the sender and receiver have access to some shared randomness (a fairly innocuous resource), would it be possible for the sender to simulate the outcome of the measurement using some of

the shared randomness and then communicate fewer classical bits to the receiver in order for him to reconstruct the sequence x^n?

The goal of Winter's POVM compression protocol [65] is to do exactly that: accurately simulate the distribution induced by the POVM, by exploiting shared randomness. The starting point for Winter's protocol is the observation that any POVM Λ can be decomposed as a convex combination of some other POVMs $\{\Gamma^{(m)}\} = \{\{\Gamma_x^{(m)}\}\}$, such that

$$\Lambda_x = \sum_m p_M(m)\Gamma_x^{(m)}. \tag{3}$$

This is due to the fact that the set of all POVMs is a convex set[6]. The set of POVMs $\{\Gamma^{(m)}\}$ then provides a simulation of the original POVM Λ by the following procedure.

(1) Generate the variable M according to the distribution $p_M(m)$.
(2) Measure the state ρ with the POVM $\Gamma^{(M)}$.

The resulting distribution for the random variable X, when marginalizing over the random variable M, is then as follows:

$$\sum_m p_M(m)\mathrm{Tr}\{\Gamma_x^{(m)}\rho\} = \mathrm{Tr}\left\{\sum_m p_M(m)\Gamma_x^{(m)}\rho\right\} = \mathrm{Tr}\{\Lambda_x\rho\} = p_X(x).$$

Thus, the random variable M is a source of noise for simulating the POVM Λ, and the output X represents information. Separating these two components is a useful idea, and it is what allows us to simulate a POVM by a protocol similar to the above.

2.1. Information processing task for measurement compression

We can now define the information processing task for a measurement compression protocol. Given the original POVM $\Lambda \equiv \{\Lambda_x\}$, suppose that it acts on an n-fold tensor product state $\rho^{\otimes n}$. The POVM then has the form

$$\Lambda^{\otimes n} \equiv \{\Lambda_{x^n}\}_{x^n \in \mathcal{X}^n},$$

where

$$\Lambda_{x^n} \equiv \Lambda_{x_1} \otimes \Lambda_{x_2} \otimes \cdots \otimes \Lambda_{x_n}.$$

The ideal measurement compression protocol would be for the sender Alice to simply perform this measurement on each copy of her state and transmit the classical output to the receiver Bob. Figure 1 depicts this ideal protocol.

Our goal is to find an approximate convex decomposition of the tensor-product POVM of the sort in (3), but in this case it should have the form:

$$\Lambda_{x_1} \otimes \Lambda_{x_2} \otimes \cdots \otimes \Lambda_{x_n} \approx \widetilde{\Lambda}_{x^n},$$

where

$$\widetilde{\Lambda}_{x^n} \equiv \sum_m p_M(m)\Gamma_{x^n}^{(m)},$$

so that each POVM element $\Gamma_{x^n}^{(m)}$ is a collective measurement on the n-fold tensor product Hilbert space. One might expect that such a collective measurement would have some compression capabilities built into it, in the sense that it could reduce the number of bits needed to represent the sequence x^n. Figure 2 depicts the most general protocol for measurement compression when both sender and receiver are to obtain the outcome of the simulated measurement (known as a 'feedback' simulation).

[6] See [18] for an explicit algorithm that decomposes any non-extremal POVM in this way.

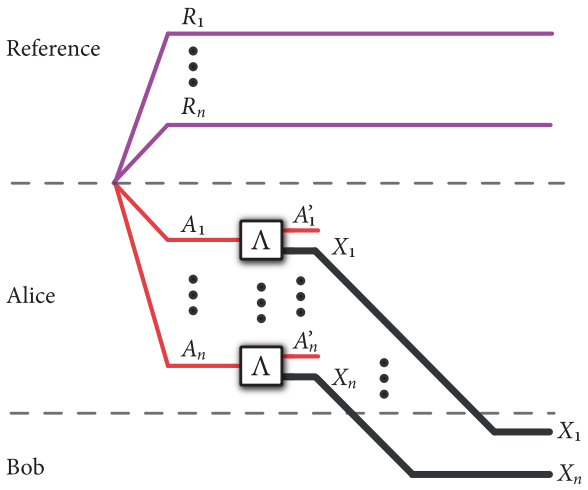

Figure 1. *Ideal measurement compression.* In an ideal protocol for measurement compression, Alice performs the POVM $\Lambda \equiv \{\Lambda_x\}$ on n copies of the state ρ, which for the ith copy leads to a quantum system A'_i and a classical output X_i. The goal of the protocol is to transmit the classical output X^n to a receiver. Doing so perfectly would require $n \log |\mathcal{X}|$ bits of communication, where $\mathcal{X}$ is the alphabet for the random variable X. Winter's measurement compression protocol gives a way of doing so by allowing for a small error but demanding that this error vanish in the asymptotic limit of many copies of the state ρ. The idea is to simulate the measurement in such a way that a third party would not be able to distinguish between the true measurement and the simulated one. An assumption of this protocol is that the sender obtains the outcome of the simulated measurement in addition to the receiver.

We now make precise the above notion of the approximation of a POVM acting on a source state. Suppose that there is some convex decomposition of the tensor-product source $\rho^{\otimes n}$ as

$$\rho^{\otimes n} = \sum_k p_K(k)\sigma_k, \tag{4}$$

where the states σ_k are generally entangled states living on the n-fold tensor product Hilbert space. Thus, one could view the preparation of the source as a selection of a random variable K according to $p_K(k)$, followed by a preparation of the state σ_K. There is then a joint distribution $p_{K,X^n}(k, x^n)$ for the selection of the source and the true measurement result:

$$p_{K,X^n}(k, x^n) \equiv p_K(k)\mathrm{Tr}\{\Lambda_{x^n}\sigma_k\}, \tag{5}$$

and a joint distribution $p_{K,\widetilde{X}^n}(k, x^n)$ for the selection of the source and the approximation measurement's result:

$$\begin{aligned} p_{K,\widetilde{X}^n}(k, x^n) &\equiv p_K(k) \sum_m p_M(m)\mathrm{Tr}\left\{\Gamma^{(m)}_{x^n}\sigma_k\right\} \\ &= p_K(k)\mathrm{Tr}\{\widetilde{\Lambda}_{x^n}\sigma_k\}. \end{aligned}$$

Definition 1 (Faithful simulation)**.** *A sequence of protocols provides a faithful simulation of the POVM Λ on the source ρ, if for all decompositions of the source of the form in (4), the above joint distributions are ϵ-close in variational distance for all $\epsilon > 0$ and sufficiently large n:*

$$\sum_{k,x^n} |p_{K,X^n}(k, x^n) - p_{K,\widetilde{X}^n}(k, x^n)| \leqslant \epsilon. \tag{6}$$

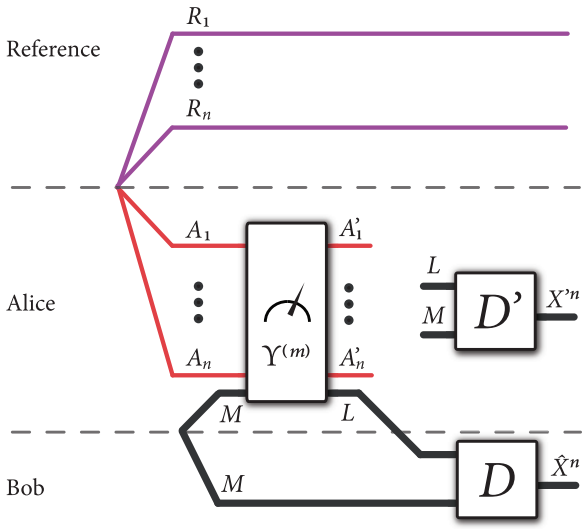

Figure 2. *Measurement compression protocol*. The most general protocol for 'feedback' measurement compression that exploits common randomness and classical communication. Alice selects a POVM $\Upsilon^{(m)} \equiv \{\Upsilon_l^{(m)}\}$ according to the common randomness M. She then performs this POVM on many copies of the state ρ, and receives an outcome l from it, modeled by the random variable L. She transmits the variable L over $\log_2 |\mathcal{L}|$ noiseless classical bit channels. Bob receives this variable, and by combining it with his share of the common randomness, he can reconstruct the output $\hat{X}^n$ of the simulated measurement. In a feedback simulation, the sender also reconstructs a variable X'^n, which is the output of the simulated measurement. The goal of a feedback measurement compression protocol is for the classical outputs of the simulated measurement to be statistically indistinguishable from the output of the ideal measurement (this is from the perspective of someone holding both the reference systems and the classical outputs).

The following lemma states a condition for faithful simulation that implies the above one, and it is the one that we will strive to meet when constructing a protocol for measurement compression.

Lemma 2. *If for all* $\epsilon > 0$ *and sufficiently large n, it holds that*

$$\sum_{x^n} \|\sqrt{\omega}(\Lambda_{x^n} - \widetilde{\Lambda}_{x^n})\sqrt{\omega}\|_1 \leqslant \epsilon, \tag{7}$$

where $\omega \equiv \rho^{\otimes n}$,[7] *then the measurement simulation is faithful, in the sense that the above inequality implies the following one for all decompositions of the source of the form in (4):*

$$\sum_{k,x^n} |p_{K,X^n}(k, x^n) - p_{K,\widetilde{X}^n}(k, x^n)| \leqslant \epsilon.$$

Proof. We rewrite the joint distribution $p_{K,X^n}(k, x^n)$ in (5) as follows:

$$\begin{aligned} p_{K,X^n}(k, x^n) &= p_K(k)\mathrm{Tr}\{\Lambda_{x^n}\sigma_k\} \\ &= \mathrm{Tr}\{(\sqrt{\omega}\Lambda_{x^n}\sqrt{\omega})(\omega^{-1/2}\, p_K(k)\sigma_k\, \omega^{-1/2})\} \\ &= \mathrm{Tr}\{\sqrt{\omega}\Lambda_{x^n}\sqrt{\omega}\, S_k\}, \end{aligned}$$

[7] The trace norm $\|A\|_1$ of an operator A is equal to $\|A\|_1 = \mathrm{Tr}\{\sqrt{A^\dagger A}\}$. The trace distance $\|\rho - \sigma\|_1$ is commonly used as a measure of distinguishability between the states ρ and σ because it is equal to $2(1 - 2p_e)$ where p_e is the probability of error in distinguishing these states if they are chosen uniformly at random.

where we define S_k as

$$S_k \equiv \omega^{-1/2}\, p_K(k)\sigma_k\, \omega^{-1/2}.$$

Observe that the operators S_k are positive and sum to the identity on the support of ω. Thus, they form a POVM $\{S_k\}$. Similarly, we can rewrite the joint distribution $p_{K,\widetilde{X}^n}(k, x^n)$ as

$$p_{K,\widetilde{X}^n}(k, x^n) = \mathrm{Tr}\{\sqrt{\omega}\widetilde{\Lambda}_{x^n}\sqrt{\omega}\, S_k\}.$$

So we can rewrite and upper bound the simulation approximation condition in (6) as

$$\begin{aligned}\sum_{k,x^n} |p_{K,X^n}(k, x^n) - p_{K,\widetilde{X}^n}(k, x^n)| &= \sum_{k,x^n} |\mathrm{Tr}\{\sqrt{\omega}(\Lambda_{x^n} - \widetilde{\Lambda}_{x^n})\sqrt{\omega}S_k\}| \\ &\leqslant \sum_{x^n} \|\sqrt{\omega}(\Lambda_{x^n} - \widetilde{\Lambda}_{x^n})\sqrt{\omega}\|_1,\end{aligned}$$

where the inequality follows from the following chain of inequalities that hold for all Hermitian operators τ:

$$\begin{aligned}\sum_k |\mathrm{Tr}\{\tau S_k\}| &= \sum_k |\mathrm{Tr}\{(\tau_+ - \tau_-)S_k\}| \\ &\leqslant \sum_k |\mathrm{Tr}\{\tau_+ S_k\}| + |\mathrm{Tr}\{\tau_- S_k\}| \\ &= \sum_k \mathrm{Tr}\{\tau_+ S_k\} + \mathrm{Tr}\{\tau_- S_k\} \\ &= \mathrm{Tr}\{\tau_+\} + \mathrm{Tr}\{\tau_-\} \\ &= \|\tau\|_1.\end{aligned}$$

In the above, we exploit the decomposition $\tau = \tau_+ - \tau_-$, where τ_+ is the positive part of τ and τ_- is the negative part, and the fact that the operators S_k form a POVM. □

We now introduce the quantum-to-classical measurement maps $\mathcal{M}_{\Lambda^{\otimes n}}$ and $\mathcal{M}_{\widetilde{\Lambda}^n}$, defined as

$$\mathcal{M}_{\Lambda^{\otimes n}}(\sigma) \equiv \sum_{x^n} \mathrm{Tr}\{\Lambda_{x^n}\sigma\}|x^n\rangle\langle x^n|, \tag{8}$$

$$\mathcal{M}_{\widetilde{\Lambda}^n}(\sigma) \equiv \sum_{x^n} \mathrm{Tr}\{\widetilde{\Lambda}_{x^n}\sigma\}|x^n\rangle\langle x^n|, \tag{9}$$

where $|x^n\rangle\langle x^n| \equiv |x_1\rangle\langle x_1| \otimes |x_2\rangle\langle x_2| \otimes \cdots \otimes |x_n\rangle\langle x_n|$ and $\{|x\rangle\}$ is some orthonormal basis. By introducing a purification $|\phi_\rho\rangle$ of the source ρ, we can then formulate another notion of faithful simulation, as given in the following definition:

Definition 3 (Faithful simulation for purification)**.** *A sequence of protocols provides a faithful simulation of the POVM Λ on the source ρ, if for a purification $|\phi_\rho\rangle$ of the source, the states on the reference and source systems after applying the measurement maps in (8)–(9) are ϵ-close in trace distance for all $\epsilon > 0$ and sufficiently large n:*

$$\left\|(id \otimes \mathcal{M}_{\Lambda^{\otimes n}})\left(\phi_\rho^{\otimes n}\right) - (id \otimes \mathcal{M}_{\widetilde{\Lambda}^n})\left(\phi_\rho^{\otimes n}\right)\right\|_1 \leqslant \epsilon. \tag{10}$$

In the above, it is implicit that the measurement maps act on the n source systems and the identity map acts on the n reference systems.

One might think that the above definition of faithful simulation is stronger than the condition in (7), but the following lemma demonstrates that they are equivalent.

Lemma 4 (Faithful simulation equivalence)**.** *The notions of faithful simulation from lemma 2 and definition 3 are equivalent, in the sense that*

$$\sum_{x^n} \|\sqrt{\omega}(\Lambda_{x^n} - \widetilde{\Lambda}_{x^n})\sqrt{\omega}\|_1 = \left\|(\mathrm{id}^{\otimes n} \otimes \mathcal{M}_{\Lambda^{\otimes n}})\left(\phi_\rho^{\otimes n}\right) - (\mathrm{id}^{\otimes n} \otimes \mathcal{M}_{\widetilde{\Lambda}^n})\left(\phi_\rho^{\otimes n}\right)\right\|_1, \tag{11}$$

for all states $\omega = \rho^{\otimes n}$, purifications of $\rho^{\otimes n}$, POVMs $\Lambda^{\otimes n}$ and $\widetilde{\Lambda}^n$, and the resulting measurement maps $\mathcal{M}_{\Lambda^{\otimes n}}$ and $\mathcal{M}_{\widetilde{\Lambda}^n}$.

Proof. We can prove this result by considering the single-copy case. Consider a state ρ, a purification ϕ_ρ, and measurements $\{\Lambda_x\}$ and $\{\widetilde{\Lambda}_x\}$. We choose the purification ϕ_ρ to be as follows:

$$\sqrt{d}(\sqrt{\rho}^R \otimes I^A)|\Phi\rangle^{RA},$$

where $|\Phi\rangle^{RA}$ is the maximally entangled state:

$$|\Phi\rangle^{RA} \equiv \frac{1}{\sqrt{d}} \sum_x |x\rangle^R |x\rangle^A,$$

and $\{|x\rangle\}$ is an orthonormal basis that diagonalizes ρ (this basis is not related to the one used in (8)–(9). Then the unnormalized state after the measurement on A is equal to

$$\left(I^R \otimes \sqrt{\Lambda_x}^A\right)|\phi_\rho\rangle\langle\phi_\rho|^{RA}\left(I^R \otimes \sqrt{\Lambda_x}^A\right) = d\left(\sqrt{\rho}^R \otimes \sqrt{\Lambda_x}^A\right)|\Phi\rangle\langle\Phi|^{RA}\left(\sqrt{\rho}^R \otimes \sqrt{\Lambda_x}^A\right). \tag{12}$$

Given the following 'transpose trick' identity that holds for a maximally entangled state (and where the transpose is with respect to the basis chosen for $|\Phi\rangle$)

$$(I \otimes M)|\Phi\rangle = (M^T \otimes I)|\Phi\rangle,$$
$$\langle\Phi|(I \otimes M) = \langle\Phi|(M^* \otimes I),$$

we then have that (12) is equal to

$$d\left(\left(\sqrt{\rho}\sqrt{\Lambda_x}^T\right)^R \otimes I^A\right)|\Phi\rangle\langle\Phi|^{RA}\left(\sqrt{\Lambda_x}^T\sqrt{\rho}^R \otimes I^A\right),$$

where the rightmost equivalence $\sqrt{\Lambda_x}^* = \sqrt{\Lambda_x}^T$ follows because Λ_x is Hermitian. Tracing over the A system then leaves the following unnormalized state on the reference system

$$\sqrt{\rho}\Lambda_x^T\sqrt{\rho}, \tag{13}$$

an observation first made in [39].

Now consider that a measurement map $\mathrm{id} \otimes \mathcal{M}_\Lambda$ has the following action on the purification $|\phi_\rho\rangle$:

$$\begin{aligned}(\mathrm{id} \otimes \mathcal{M}_\Lambda)(|\phi_\rho\rangle\langle\phi_\rho|) &= \sum_x \mathrm{Tr}_A\left\{\left(\mathrm{id}^R \otimes \Lambda_x^A\right)(|\phi_\rho\rangle\langle\phi_\rho|^{RA})\right\} \otimes |x\rangle\langle x|^X \\ &= \sum_x \left(\sqrt{\rho}\Lambda_x^T\sqrt{\rho}\right)^R \otimes |x\rangle\langle x|^X,\end{aligned}$$

where the last line follows from the conclusion in (13). Thus, we have that

$$\begin{aligned}&\|(\mathrm{id} \otimes \mathcal{M}_\Lambda)(\phi_\rho) - (\mathrm{id} \otimes \mathcal{M}_{\widetilde{\Lambda}})(\phi_\rho)\|_1 \\ &= \left\|\sum_x \left(\sqrt{\rho}\Lambda_x^T\sqrt{\rho}\right)^R \otimes |x\rangle\langle x|^X - \sum_x \left(\sqrt{\rho}\widetilde{\Lambda}_x^T\sqrt{\rho}\right)^R \otimes |x\rangle\langle x|^X\right\|_1 \\ &= \left\|\sum_x \sqrt{\rho}\left(\Lambda_x^T - \widetilde{\Lambda}_x^T\right)\sqrt{\rho} \otimes |x\rangle\langle x|^X\right\|_1 \\ &= \sum_x \left\|\sqrt{\rho}\left(\Lambda_x^T - \widetilde{\Lambda}_x^T\right)\sqrt{\rho}\right\|_1 \\ &= \sum_x \|\sqrt{\rho}(\Lambda_x - \widetilde{\Lambda}_x)\sqrt{\rho}\|_1,\end{aligned}$$

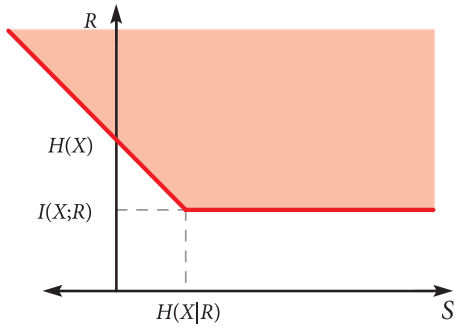

Figure 3. *Optimal rate region for measurement compression with feedback*. The figure plots the optimal rate region from theorem 5. The measurement compression protocol demonstrates that the rate pair $(S = H(X|R), R = I(X;R))$ is achievable. Wasting common randomness achieves all of the rate pairs to the right of this corner point. Time-sharing between measurement compression and Shannon compression $(S = 0, R = H(X))$ achieves all of the rate pairs between them. Finally, employing Shannon compression and converting the extra classical communication to common randomness achieves all of the optimal rate pairs along the line extending northwest from Shannon compression. The converse theorem in section 2.4 proves that this rate region is optimal.

where the third equality follows because the trace norm of a block-diagonal operator is just the sum of the trace norms of the blocks. The fourth equality follows because

$$\sqrt{\rho}\left(\Lambda_x^T - \widetilde{\Lambda}_x^T\right)\sqrt{\rho} = [\sqrt{\rho}(\Lambda_x - \widetilde{\Lambda}_x)\sqrt{\rho}]^T,$$

the trace norm depends only on the singular values of a matrix, and these are invariant under transposition. □

2.2. Measurement compression theorem

We can now state Winter's main result:

Theorem 5 (Measurement compression theorem)**.** *Let ρ be a source state and Λ a POVM to simulate on this state. A protocol for a faithful feedback simulation of the POVM with classical communication rate R and common randomness rate S exists if and only if the following set of inequalities hold*

$$\begin{aligned} R &\geqslant I(X;R), \\ R + S &\geqslant H(X), \end{aligned}$$

where the entropies are with respect to the state

$$\sum_x |x\rangle\langle x|^X \otimes \mathrm{Tr}_A\left\{\left(I^R \otimes \Lambda_x^A\right)\phi^{RA}\right\}, \tag{14}$$

and ϕ^{RA} is any purification of the state ρ.

Note that $I(X;R)$ and $H(X)$ are independent of the choice of purification ϕ^{RA}. Moreover, the entropies are invariant with respect to transposition in the basis that diagonalizes ρ so that we could instead evaluate entropies with respect to the following classical-quantum state:

$$\sum_x |x\rangle\langle x|^X \otimes \mathrm{Tr}_A\left\{\left(I^R \otimes \left(\Lambda_x^T\right)^A\right)\phi^{RA}\right\}.$$

Figure 3 provides a plot of the optimal rate region given in the above theorem for this case of a feedback simulation in which the sender also obtains the outcome of the measurement simulation.

After giving a simple example of an application of the above theorem, we prove it in two parts. First, we prove that there exists a measurement compression protocol achieving the rates in the above theorem, specifically the corner point ($S = H(X|R), R = I(X;R)$). Next, we prove the converse part of the theorem: that one cannot do better than the rates given in the above theorem.

2.2.1. Examples. We review two simple examples to illustrate some applications of theorem 5. Our first example is for the case that the initial state on which Alice performs the measurement is some pure state ϕ^A. In this case, the state in (14) becomes

$$\sum_x |x\rangle\langle x|^X \otimes \mathrm{Tr}_A\left\{\left(I^R \otimes \Lambda_x^A\right)(\psi^R \otimes \phi^A)\right\} = \sum_x \mathrm{Tr}\{\Lambda_x \phi\}|x\rangle\langle x|^X \otimes \psi^R.$$

Thus, the reference has no correlations with the outcome of the measurement, so that $I(X;R) = 0$ and $H(X|R) = H(X)$, where $H(X)$ is the Shannon entropy of the distribution $p(x) = \mathrm{Tr}\{\Lambda_x \phi\}$. No classical communication is required in this case—common randomness suffices for this simulation. Indeed, the protocol just has Alice and Bob operate as in randomness dilution, whereby they dilute their uniform, shared randomness to match the distribution $p(x)$. The idea here is that there are no correlations with some reference system, or similarly, there is only a trivial decomposition of the form in (5), so that K is a degenerate random variable.

Our next example is a natural one discussed in the conclusion of [27]. Consider the POVM

$$\left\{\tfrac{1}{2}|0\rangle\langle 0|, \tfrac{1}{2}|1\rangle\langle 1|, \tfrac{1}{2}|+\rangle\langle +|, \tfrac{1}{2}|-\rangle\langle -|\right\} \tag{15}$$

acting on the maximally mixed state $\pi^A = I^A/2$. We would like to determine the resources required to simulate the action of this measurement on the maximally mixed state. Consider that the Bell state

$$|\Phi\rangle^{RA} \equiv \tfrac{1}{\sqrt{2}}(|00\rangle^{RA} + |11\rangle^{RA}) = \tfrac{1}{\sqrt{2}}(|++\rangle^{RA} + |--\rangle^{RA})$$

is a purification of the maximally mixed state π^A. The post-measurement classical-quantum state in (14) for this case is as follows:

$$\tfrac{1}{4}(|0\rangle\langle 0|^X \otimes |0\rangle\langle 0|^R + |1\rangle\langle 1|^X \otimes |1\rangle\langle 1|^R + |2\rangle\langle 2|^X \otimes |+\rangle\langle +|^R + |3\rangle\langle 3|^X \otimes |-\rangle\langle -|^R).$$

A simple calculation reveals that the mutual information $I(X;R)$ of the above state is equal to one bit. Also, the conditional entropy $H(X|R)$ of the above state is equal to one bit. Thus, one bit of classical communication and one bit of common randomness are required to simulate this measurement.

For this case, the simulation is straightforward since the original POVM decomposes as a random choice of a Z or X Pauli measurement:

$$\left\{\tfrac{1}{2}\{|0\rangle\langle 0|, |1\rangle\langle 1|\}, \tfrac{1}{2}\{|+\rangle\langle +|, |-\rangle\langle -|\}\right\}.$$

Thus, Alice and Bob can use one bit of common randomness to select which measurement to perform. Alice then performs the Z or X measurement and sends the outcome to Bob using one classical bit channel. Bob can then determine which of the four outcomes in (15) has occurred by combining the two bits.

2.3. Achievability proof for measurement compression

The resource inequality [25, 26] characterizing measurement compression is as follows:

$$I(X;R)[c \to c] + H(X|R)[cc] \geqslant \langle \Lambda(\rho)\rangle,$$

where the entropic quantities are with respect to a state of the following form:

$$\sum_x |x\rangle\langle x|^X \otimes \mathrm{Tr}_A\left\{\left(I^R \otimes \Lambda_x^A\right)\phi^{RA}\right\} = \sum_x p_X(x)|x\rangle\langle x|^X \otimes \theta_x^R,$$
$$\theta_x^R \equiv \mathrm{Tr}_A\left\{\left(I^R \otimes \Lambda_x^A\right)\phi^{RA}\right\}/p_X(x),$$
$$p_X(x) \equiv \mathrm{Tr}\left\{\left(I^R \otimes \Lambda_x^A\right)\phi^{RA}\right\},$$

and ϕ^{RA} is some purification of the state ρ. The operators θ_x^R take on the following special form

$$\theta_x^R = \sqrt{\rho}\Lambda_x^T\sqrt{\rho},$$

if the spectral decomposition of ρ is $\rho = \sum_x \lambda_x|x\rangle\langle x|$ and the purification ϕ^{RA} is taken as $|\phi\rangle^{RA} = \sum_x \sqrt{\lambda_x}|x\rangle^R|x\rangle^A$. The meaning of the above resource inequality is that $nI(X;R)$ bits of classical communication $[c \to c]$ and $nH(X|R)$ bits of common randomness $[cc]$ are required in order to simulate the action of the POVM $\Lambda^{\otimes n}$ on the tensor product state $\rho^{\otimes n}$, and the simulation becomes exact in the asymptotic limit as $n \to \infty$.

The main idea of the proof is to 'steer' the state of the reference to be close to the ensemble produced by the ideal measurement. In order to do so, we construct a measurement at random, chosen from an ensemble of operators built from the ideal measurement Λ and the state ρ. By employing the Ahlswede–Winter operator Chernoff bound [2], we can then guarantee that there exists a particular POVM satisfying the faithful simulation condition in (3), as long as the amount of classical communication and common randomness is sufficiently large.

The achievability part of the theorem begins by considering the following ensemble derived from the state ρ and the POVM $\Lambda = \{\Lambda_x\}$:

$$p_X(x) \equiv \mathrm{Tr}\{\Lambda_x\rho\},$$
$$\hat{\rho}_x \equiv \frac{1}{p_X(x)}\sqrt{\rho}\Lambda_x\sqrt{\rho}.$$

(Recall our statement that the entropies are invariant with respect to transposition in the basis that diagonalizes ρ.) Observe that the expected density operator of this ensemble is just the state ρ:

$$\sum_x p_X(x)\hat{\rho}_x = \sum_x \sqrt{\rho}\Lambda_x\sqrt{\rho} = \rho.$$

We will prove that there exist POVMs $\Gamma^{(m)} = \left\{\Gamma_{x^n}^{(m)}\right\}_{x^n \in \mathcal{L}}$ with $m \in \mathcal{M}$,

$$|\mathcal{L}| = 2^{n[I(X;R)+3\delta]}, \tag{16}$$

$$|\mathcal{M}| = 2^{n[H(X|R)+\delta]}, \tag{17}$$

for some $\delta > 0$, such that the mixed POVM $\widetilde{\Lambda}$ with elements $\widetilde{\Lambda}_{x^n} \equiv \frac{1}{|\mathcal{M}|}\sum_m \Gamma_{x^n}^{(m)}$ provides a faithful simulation of Λ on ρ according to the criterion in (7).

In order to prove achievability, we require the Ahlswede–Winter Operator Chernoff bound, which we recall below:

Lemma 6 (Operator Chernoff bound)**.** *Let $\xi_1, \ldots, \xi_M$ be M independent and identically distributed random variables with values in the algebra $\mathcal{B}(\mathcal{H})$ of linear operators acting on some finite dimensional Hilbert space $\mathcal{H}$. Each ξ_m has all of its eigenvalues between zero and one, so that the following operator inequality holds*

$$\forall m \in [M] : 0 \leqslant \xi_m \leqslant I. \tag{18}$$

Let $\overline{\xi}$ denote the sample average of the M random variables:

$$\overline{\xi} = \frac{1}{M}\sum_{m=1}^{M} \xi_m. \tag{19}$$

Suppose that the expectation $\mathbb{E}_\xi\{\xi_m\} \equiv \mu$ of each operator ξ_m exceeds the identity operator scaled by a number $a > 0$:

$$\mu \geqslant aI. \tag{20}$$

Then for every η where $0 < \eta < 1/2$ and $a\,(1+\eta) \leqslant 1$, we can bound the probability that the sample average $\overline{\xi}$ lies inside the operator interval $[(1 \pm \eta)\mu]$:

$$\Pr_{\xi}\{(1-\eta)\mu \leqslant \overline{\xi} \leqslant (1+\eta)\mu\} \geqslant 1 - 2\dim\mathcal{H}\exp\left(-\frac{M\eta^2 a}{4\ln 2}\right). \tag{21}$$

Thus it is highly likely that the sample average operator $\overline{\xi}$ becomes close to the true expected operator μ as M becomes large.

We first define some operators that we will use to generate the POVM elements $\Gamma^{(m)}_{x^n}$. For all $x^n \in T^{X^n}_\delta$ (where $T^{X^n}_\delta$ is the strongly typical set—see appendix A) consider the following positive operators with trace less than one:

$$\xi'_{x^n} \equiv \Pi^n_{\rho,\delta}\ \Pi_{\hat{\rho}_{x^n},\delta}\ \hat{\rho}_{x^n}\ \Pi_{\hat{\rho}_{x^n},\delta}\ \Pi^n_{\rho,\delta}. \tag{22}$$

These operators ξ'_{x^n} have a trace almost equal to one because

$$\begin{aligned}
\mathrm{Tr}\{\xi'_{x^n}\} &= \mathrm{Tr}\{\Pi^n_{\rho,\delta}\ \Pi_{\hat{\rho}_{x^n},\delta}\ \hat{\rho}_{x^n}\ \Pi_{\hat{\rho}_{x^n},\delta}\ \Pi^n_{\rho,\delta}\} \\
&= \mathrm{Tr}\{\Pi^n_{\rho,\delta}\ \Pi_{\hat{\rho}_{x^n},\delta}\ \hat{\rho}_{x^n}\ \Pi_{\hat{\rho}_{x^n},\delta}\} \\
&\geqslant \mathrm{Tr}\{\Pi^n_{\rho,\delta}\ \hat{\rho}_{x^n}\} - \|\hat{\rho}_{x^n} - \Pi_{\hat{\rho}_{x^n},\delta}\ \hat{\rho}_{x^n}\ \Pi_{\hat{\rho}_{x^n},\delta}\|_1 \\
&\geqslant 1 - \epsilon - 2\sqrt{\epsilon}.
\end{aligned} \tag{23}$$

The first inequality follows from the trace inequality in lemma 17, and the second inequality follows by appealing to the properties of quantum typicality reviewed in appendix A. Also, we set S to be the probability of the typical set $T^{X^n}_\delta$, and recall that this probability is near to one:

$$S \equiv \Pr\{X^n \in T^{X^n}_\delta\} = \sum_{x^n \in T^{X^n}_\delta} p_{X^n}(x^n) \geqslant 1 - \epsilon.$$

We define ξ' to be the expectation of the operators ξ'_{x^n}, when each one is chosen according to a pruned distribution $p_{X^m}(x^n)$:

$$\xi' \equiv \mathbb{E}_{X^m}\{\xi'_{X^m}\} = \sum_{x^n} p_{X^m}(x^n)\xi'_{x^n},$$

where we define $p_{X^m}(x^n)$ as

$$p_{X^m}(x^n) \equiv \begin{cases} p_{X^n}(x^n)/S & x^n \in T^{X^n}_\delta \\ 0 & \text{else} \end{cases}. \tag{24}$$

It follows that $\mathrm{Tr}\{\xi'\} \geqslant 1 - \epsilon - 2\sqrt{\epsilon}$ because

$$\begin{aligned}
\mathrm{Tr}\{\xi'\} &= \sum_{x^n} p_{X^m}(x^n)\mathrm{Tr}\{\xi'_{x^n}\} \\
&\geqslant 1 - \epsilon - 2\sqrt{\epsilon}.
\end{aligned} \tag{25}$$

The inequality follows from the one in (23). From properties of quantum typicality, we know that

$$\Pi^n_{\rho,\delta}\ \rho^{\otimes n}\ \Pi^n_{\rho,\delta} \geqslant 2^{-n[H(\rho)+\delta]}\ \Pi^n_{\rho,\delta}. \tag{26}$$

We now define Π to be the projector onto the subspace spanned by the eigenvectors of ξ' with eigenvalue larger than $\epsilon\alpha$, where $\alpha \equiv 2^{-n[H(\rho)+\delta]} = 2^{-n[H(R)+\delta]}$. Defining the operator Ω as

$$\Omega \equiv \Pi \xi' \Pi, \tag{27}$$

it follows that $\mathrm{Tr}\{\Omega\} \geqslant 1 - 2\epsilon - 2\sqrt{\epsilon}$ because

$$\mathrm{rank}(\Omega) \leqslant \mathrm{Tr}\{\Pi\} \leqslant \mathrm{Tr}\left\{\Pi^n_{\rho,\delta}\right\} \leqslant 2^{n[H(\rho)+\delta]} = \alpha^{-1},$$

so that eigenvalues smaller than $\epsilon\alpha$ contribute at most ϵ to $\mathrm{Tr}\{\Omega\}$, giving

$$\mathrm{Tr}\{\Omega\} \geqslant (1-\epsilon)\mathrm{Tr}\{\xi'\} \geqslant (1-\epsilon)(1-\epsilon-2\sqrt{\epsilon}) \geqslant 1 - 2\epsilon - 2\sqrt{\epsilon}. \tag{28}$$

We now exploit a random selection of the operators in (22) in order to build up a POVM that has desirable properties that we can use to prove the achievability part of this theorem. Let ξ_{x^n} denote the following operators:

$$\xi_{x^n} \equiv \Pi\ \xi'_{x^n}\ \Pi,$$

so that we confine them to be in the subspace onto which Π projects. Define $|\mathcal{L}||\mathcal{M}|$ random variables $X^n(l, m)$ that are chosen independently according to the pruned distribution $p_{X'^n}(x^n)$. We can group these variables into $|\mathcal{M}|$ sets $\mathcal{C}_m \equiv \{X^n(l, m)\}_{l\in\mathcal{L}}$, according to the value of the common randomness m. Under the pruned distribution, the expectation of the random operator $\xi_{X^n(l,m)}$ is equal to Ω:

$$\mathbb{E}_{X^n(l,m)}\{\xi_{X^n(l,m)}\} = \sum_{x^n} p_{X'^n}(x^n)\xi_{x^n} = \Omega.$$

Let E_m denote the event that the sample average of the operators in the mth set $\mathcal{C}_m$ falls close to its mean Ω (in the operator interval sense):

$$\Omega(1-\epsilon) \leqslant \frac{1}{|\mathcal{L}|}\sum_l \xi_{x^n(l,m)} \leqslant \Omega(1+\epsilon). \tag{29}$$

The above event E_m is equivalent to the following rescaled event:

$$\beta\Omega(1-\epsilon) \leqslant \frac{\beta}{|\mathcal{L}|}\sum_l \xi_{x^n(l,m)} \leqslant \beta\Omega(1+\epsilon),$$

where $\beta \equiv 2^{n[H(R|X)-\delta]}$. It is then clear that the expectation of the operators $\beta\xi_{x^n(l,)}$ satisfies the following operator inequality needed in the operator Chernoff bound:

$$\mathbb{E}_{X'^n}\{\beta\xi_{X^n(l,m)}\} = \beta\Omega \geqslant \alpha\beta\epsilon\Pi.$$

Also, each individual rescaled operator $\beta\xi_{x^n(l,m)}$ admits a tight operator upper bound with the identity operator because

$$\begin{aligned}
\beta\xi_{x^n(l,m)} &= 2^{n[H(R|X)-\delta]}\ \Pi\ \Pi^n_{\rho,\delta}\ \Pi_{\hat{\rho}_{x^n},\delta}\ \hat{\rho}_{x^n}\ \Pi_{\hat{\rho}_{x^n},\delta}\ \Pi^n_{\rho,\delta}\ \Pi \\
&\leqslant 2^{n[H(R|X)-\delta]}\ 2^{-n[H(R|X)-\delta]}\ \Pi\ \Pi^n_{\rho,\delta}\ \Pi_{\hat{\rho}_{x^n},\delta}\ \Pi^n_{\rho,\delta}\ \Pi \\
&= \Pi\ \Pi^n_{\rho,\delta}\ \Pi_{\hat{\rho}_{x^n},\delta}\ \Pi^n_{\rho,\delta}\ \Pi \\
&\leqslant I,
\end{aligned}$$

where we applied the operator inequality $\Pi_{\hat{\rho}_{x^n},\delta}\ \hat{\rho}_{x^n}\ \Pi_{\hat{\rho}_{x^n},\delta} \leqslant 2^{-n[H(R|X)-\delta]}\ \Pi_{\hat{\rho}_{x^n},\delta}$ for the first inequality. Applying the operator Chernoff bound then gives us an upper bound on the probability that event E_m does not occur

$$\begin{aligned}
\Pr\{\neg E_m\} &\leqslant 2\mathrm{rank}(\Pi)\exp\left(-\frac{|\mathcal{L}|\epsilon^2(\epsilon\alpha\beta)}{4\ln 2}\right) \\
&\leqslant 2\cdot 2^{n[H(R)+\delta]}\exp\left(-\frac{2^{n[I(X;R)+3\delta]}\epsilon^3 2^{-n[H(R)+\delta]}2^{n[H(R|X)-\delta]}}{4\ln 2}\right)
\end{aligned}$$

$$\leqslant 2 \cdot 2^{n[H(R)+\delta]} \exp\left(-\frac{2^{n\delta}\epsilon^3}{4\ln 2}\right)$$
$$= 2\exp\left(-\frac{2^{n\delta}\epsilon^3}{4\ln 2} + n[H(R)+\delta]\ln 2\right).$$

Thus, by choosing $|\mathcal{L}|$ as we did in (16), it is possible to make the probability of the complement of E_m doubly-exponentially small in n. Also, the above application of the operator Chernoff bound makes it clear why we rescale according to β—doing so allows for the rescaled operators $\beta\xi_{x^n(l,m)}$ to admit a tight operator upper bound with the identity (so that the demands of the operator Chernoff bound are met), while allowing for $|\mathcal{L}|$ to be as small as $2^{n[I(X;R)+3\delta]}$ and $\Pr\{E_m^c\}$ to be arbitrarily small.

We now define a counting function $c_{x^n}(\mathcal{L},\mathcal{M})$ on the sets $\mathcal{L}$ and $\mathcal{M}$, which counts the fraction of occurrences of a sequence $x^n \in T_\delta^{X^n}$ in the set $\{x^n(l,m)\}_{l\in\mathcal{L},m\in\mathcal{M}}$:

$$c_{x^n}(\mathcal{L},\mathcal{M}) \equiv \frac{1}{|\mathcal{L}||\mathcal{M}|}|\{l,m : x^n(l,m) = x^n\}|.$$

(This is effectively a sample average of the counts.) When choosing the random variables $X^n(l,m)$ IID according to the pruned distribution, the expectation of the random counting function $C_{x^n}(\mathcal{L},\mathcal{M})$ is equal to the probability of the sequence x^n:

$$\mathbb{E}\{C_{x^n}(\mathcal{L},\mathcal{M})\} = p_{X'^n}(x^n).$$

Thus, for any sequence $x^n \in T_\delta^{X^n}$, the expectation of the counting function has the following lower bound:

$$\begin{aligned}\mathbb{E}\{C_{x^n}(\mathcal{L},\mathcal{M})\} &= p_{X'^n}(x^n)\\ &= \frac{1}{S}p_{X^n}(x^n)\\ &\geqslant \min\left\{p_{X^n}(x^n) : x^n \in T_\delta^{X^n}\right\}\\ &\geqslant \gamma \equiv 2^{-n[H(X)+\delta]},\end{aligned}$$

where S, recall, is the probability that a random sequence X^n is typical.

In order to appeal to the operator Chernoff bound (we could just use the classical one, but we instead choose to exploit the operator one), we define $\hat{P}$ as a diagonal density operator of dimension $\left|T_\delta^{X^n}\right| \times \left|T_\delta^{X^n}\right|$, whose diagonal entries are just the entries of the pruned distribution $p_{X'^n}(x^n)$. Similarly, for a particular realization of the set $\{x^n(l,m)\}_{l\in\mathcal{L},m\in\mathcal{M}}$, we can define $\hat{C}$ as a diagonal density operator of the same dimension, whose diagonal entries are just the entries of the counting functions $c_{x^n}(\mathcal{L},\mathcal{M})$. From the above reasoning, it is then clear that the expectation of $\hat{C}$ under a random choice of the $x^n(l,m)$ sequences is just $\hat{P}$:

$$\mathbb{E}\{\hat{C}\} = \hat{P}.$$

Furthermore, (again by the above reasoning), we can establish the following lower bound on $\hat{P}$:

$$\hat{P} \geqslant \gamma\ \Pi^n_{p_X,\delta},$$

where $\Pi^n_{p_X,\delta}$ is a typical projector corresponding to the distribution $p_X(x)$.

Let E_0 be the event that the operator $\hat{C}$ is within ϵ of its mean $\hat{P}$:

$$(1-\epsilon)\hat{P} \leqslant \hat{C} \leqslant (1+\epsilon)\hat{P}. \tag{30}$$

By appealing to the operator Chernoff bound, we can bound the probability that the above event does not occur when choosing the sequences $x^n(l,m)$ randomly as prescribed above:

$$\begin{aligned}\Pr\{\neg E_0\} &\leqslant 2\operatorname{rank}\left(\Pi^n_{p_X,\delta}\right)\exp\left(-\frac{|\mathcal{L}||\mathcal{M}|\epsilon^2\gamma}{4\ln 2}\right)\\ &= 2\cdot 2^{n[H(X)+\delta]}\exp\left(-\frac{2^{n[I(X;R)+3\delta]}2^{n[H(X|R)+\delta]}\epsilon^2 2^{-n[H(X)+\delta]}}{4\ln 2}\right)\\ &= 2\cdot 2^{n[H(X)+\delta]}\exp\left(-\frac{2^{n3\delta}}{4\ln 2}\right)\\ &= 2\exp\left(-\frac{2^{n3\delta}}{4\ln 2}+n[H(X)+\delta]\ln 2\right).\end{aligned}$$

Thus, by choosing $|\mathcal{L}||\mathcal{M}|$ as we did in (16)–(17), it is possible to make this probability be doubly-exponentially small in n.

We want to ensure that it is possible for all of the events E_m and E_0 to occur simultaneously. We can guarantee this by applying DeMorgan's law, the union bound, and the above estimates:

$$\begin{aligned}\Pr\left\{\neg\left[E_0\cap\left(\bigcap_m E_m\right)\right]\right\} &= \Pr\left\{\neg E_0\cup\left(\bigcup_m \neg E_m\right)\right\}\\ &\leqslant \Pr\{\neg E_0\}+\sum_m \Pr\{\neg E_m\}\\ &\leqslant 2\exp\left(-\frac{2^{n3\delta}}{4\ln 2}+n[H(X)+\delta]\ln 2\right)\\ &\quad + |\mathcal{M}|2\exp\left(-\frac{2^{n\delta}\epsilon^3}{4\ln 2}+n[H(R)+\delta]\ln 2\right),\end{aligned}\tag{31}$$

which becomes arbitrarily small as $n\to\infty$. (Thus, it is in fact overwhelmingly likely for our desired conditions to hold if we choose $|\mathcal{L}|$ and $|\mathcal{M}|$ as we did in (16)–(17).)

So, assume now that we have a set $\{x^n(l,m)\}_{l\in\mathcal{L},m\in\mathcal{M}}$ such that the corresponding operators $\{\xi_{x^n(l,m)}\}$ and $\hat{C}$ satisfy the conditions in (29) and (30). We can now construct from them a set of POVMs $\{\Gamma^{(m)}\}$ that will perform a faithful measurement simulation. We define the POVM elements $\Gamma^{(m)}_{x^n}$ of $\Gamma^{(m)}$ as follows:

$$\begin{aligned}\Gamma^{(m)}_{x^n} &\equiv \frac{S}{1+\epsilon}\omega^{-1/2}\left(\frac{1}{|\mathcal{L}|}\sum_{l\,:\,x^n(l,m)=x^n}\xi_{x^n(l,m)}\right)\omega^{-1/2}\\ &= \frac{S}{1+\epsilon}\,\frac{|\{l:x^n(l,m)=x^n\}|}{|\mathcal{L}|}\,\omega^{-1/2}\,\xi_{x^n}\,\omega^{-1/2}.\end{aligned}\tag{32}$$

We check that for each value m of the common randomness that these operators form a sub-POVM (a set of positive operators whose sum is upper bounded by the identity). Indeed, we can appeal to the fact that the operators satisfy the condition in (29):

$$\begin{aligned}\sqrt{\omega}\sum_{x^n\in\mathcal{X}^n}\Gamma^{(m)}_{x^n}\sqrt{\omega} &= \frac{S}{1+\epsilon}\frac{1}{|\mathcal{L}|}\sum_{x^n\in\mathcal{X}^n}\left(\sum_{l\,:\,x^n(l,m)=x^n}\xi_{x^n(l,m)}\right)\\ &= \frac{S}{1+\epsilon}\frac{1}{|\mathcal{L}|}\sum_l \xi_{x^n(l,m)}\\ &\leqslant \frac{S}{1+\epsilon}\,\Omega(1+\epsilon)\\ &= S\,\Omega,\end{aligned}$$

where the inequality appeals to (29). Continuing with the definition of Ω in (27), we have

$$\begin{aligned}
&= S \sum_{x^n \in \mathcal{X}^n} p_{X^n}(x^n) \Pi\ \Pi^n_{\rho,\delta}\ \Pi_{\hat{\rho}_{x^n},\delta}\ \hat{\rho}_{x^n}\ \Pi_{\hat{\rho}_{x^n},\delta}\ \Pi^n_{\rho,\delta}\ \Pi \\
&\leqslant \sum_{x^n \in T_\delta^{X^n}} p_{X^n}(x^n) \Pi\ \Pi^n_{\rho,\delta}\ \hat{\rho}_{x^n}\ \Pi^n_{\rho,\delta}\ \Pi \\
&\leqslant \sum_{x^n \in \mathcal{X}^n} p_{X^n}(x^n) \Pi\ \Pi^n_{\rho,\delta}\ \hat{\rho}_{x^n}\ \Pi^n_{\rho,\delta}\ \Pi \\
&= \Pi\ \Pi^n_{\rho,\delta}\ \rho^{\otimes n}\ \Pi^n_{\rho,\delta}\ \Pi \\
&\leqslant \rho^{\otimes n} = \omega.
\end{aligned}$$

The first inequality follows from the operator inequality $\Pi_{\hat{\rho}_{x^n},\delta}\ \hat{\rho}_{x^n}\ \Pi_{\hat{\rho}_{x^n},\delta} \leqslant \hat{\rho}_{x^n}$ (the projectors $\Pi_{\hat{\rho}_{x^n},\delta}$ are defined with respect to the eigenbasis of $\hat{\rho}_{x^n}$). We can then conclude that these operators form a sub-POVM because

$$\sqrt{\omega} \sum_{x^n \in \mathcal{X}^n} \Gamma^{(m)}_{x^n} \sqrt{\omega} \leqslant \omega \implies \sum_{x^n \in \mathcal{X}^n} \Gamma^{(m)}_{x^n} \leqslant I.$$

By filling up the rest of the space with some extra operator

$$\Gamma^{(m)}_0 \equiv I - \sum_{x^n \in \mathcal{X}^n} \Gamma^{(m)}_{x^n},$$

we then have a valid POVM.

Note that we have chosen the measurement operators $\Gamma^{(m)}_{x^n}$ so that there is a correspondence between a sequence x^n and a measurement outcome. In the communication paradigm, though, we would like to have the measurement output some index l that Alice can send over noiseless classical channels to Bob, so that he can subsequently construct the sequence $x^n(l, m)$ from the value of l, the common randomness m, and the codebook $\{x^n(l, m)\}$. So, for the communication paradigm, we can also consider the measurement to be of the form

$$\Upsilon^{(m)}_l \equiv \frac{S}{1+\epsilon} \frac{1}{|\mathcal{L}|}\ \omega^{-1/2}\ \xi_{x^n(l,m)}\ \omega^{-1/2}. \tag{33}$$

The POVM in (32) then just results by computing x^n from the codeword $x^n(l, m)$.

We define the operators $\widetilde{\Lambda}_{x^n}$ as

$$\widetilde{\Lambda}_{x^n} \equiv \frac{1}{|\mathcal{M}|} \sum_m \Gamma^{(m)}_{x^n},$$

or equivalently as

$$\widetilde{\Lambda}_{x^n} = \frac{1}{|\mathcal{M}|} \sum_{l,m} \mathcal{I}(x^n = x^n(l, m)) \Upsilon^{(m)}_l,$$

where $\mathcal{I}(x^n = x^n(l, m))$ is an indicator function. We now check that the constructed POVM satisfies the condition in (7) for a faithful simulation:

$$\begin{aligned}
&\sum_{x^n} \|\sqrt{\omega}(\Lambda_{x^n} - \widetilde{\Lambda}_{x^n})\sqrt{\omega}\|_1 \\
&= \sum_{x^n} \left\| p_{X^n}(x^n)\hat{\rho}_{x^n} - \frac{S}{1+\epsilon} \frac{|\{l, m : x^n(l, m) = x^n\}|}{|\mathcal{L}||\mathcal{M}|}\ \xi_{x^n} \right\|_1 \\
&= \sum_{x^n \notin T_\delta^{X^n}} \|p_{X^n}(x^n)\hat{\rho}_{x^n}\|_1 + \sum_{x^n \in T_\delta^{X^n}} \left\| p_{X^n}(x^n)\hat{\rho}_{x^n} - \frac{S}{1+\epsilon} \frac{|\{l, m : x^n(l, m) = x^n\}|}{|\mathcal{L}||\mathcal{M}|}\ \xi_{x^n} \right\|_1 \\
&\leqslant \epsilon + \sum_{x^n \in T_\delta^{X^n}} \left\| p_{X^n}(x^n)\hat{\rho}_{x^n} - p_{X^n}(x^n)\xi_{x^n} + p_{X^n}(x^n)\xi_{x^n} - \frac{S}{1+\epsilon} \frac{|\{l, m : x^n(l, m) = x^n\}|}{|\mathcal{L}||\mathcal{M}|}\ \xi_{x^n} \right\|_1 .
\end{aligned}$$

The third equality above follows because the operators ξ_{x^n} are defined to be zero when $x^n \notin T_\delta^{X^n}$. Then, the bound in the last line follows from typicality ($\Pr\{X^n \notin T_\delta^{X^n}\} \leqslant \epsilon$). Continuing, we upper bound as

$$\begin{aligned}
&\leqslant \epsilon + \sum_{x^n \in T_\delta^{X^n}} p_{X^n}(x^n) \|\hat{\rho}_{x^n} - \xi_{x^n}\|_1 + \sum_{x^n \in T_\delta^{X^n}} \left\| p_{X^n}(x^n) \xi_{x^n} - \frac{S}{1+\epsilon} \frac{|\{l,m : x^n(l,m) = x^n\}|}{|\mathcal{L}||\mathcal{M}|} \xi_{x^n} \right\|_1 \\
&\leqslant \epsilon + \sum_{x^n \in T_\delta^{X^n}} \frac{p_{X^n}(x^n)}{S} \|\hat{\rho}_{x^n} - \xi_{x^n}\|_1 + \sum_{x^n \in T_\delta^{X^n}} \left| \frac{1}{S} p_{X^n}(x^n) - \frac{1}{1+\epsilon} \frac{|\{l,m : x^n(l,m) = x^n\}|}{|\mathcal{L}||\mathcal{M}|} \right| \\
&= \epsilon + \sum_{x^n \in T_\delta^{X^n}} \frac{p_{X^n}(x^n)}{S} \|\hat{\rho}_{x^n} - \xi_{x^n}\|_1 + \left\| \hat{P} - \frac{1}{1+\epsilon} \hat{C} \right\|_1 .
\end{aligned}$$

The first inequality is the triangle inequality, and the second inequality follows by dividing the rightmost two terms by S. The equality follows by invoking the definitions of the operators $\hat{P}$ and $\hat{C}$. We handle these two remaining terms individually. Consider that

$$\begin{aligned}
\left\| \hat{P} - \frac{1}{1+\epsilon} \hat{C} \right\|_1 &= \frac{1}{1+\epsilon} \|(1+\epsilon)\hat{P} - \hat{C}\|_1 \\
&\leqslant \frac{1}{1+\epsilon} (\|\epsilon \hat{P}\|_1 + \|\hat{P} - \hat{C}\|_1) \\
&\leqslant \frac{2\epsilon}{1+\epsilon} \\
&\leqslant 2\epsilon, \qquad (34)
\end{aligned}$$

which follows from the triangle inequality, the fact that $\hat{P}$ is a density operator, and that $\hat{P}$ and $\hat{C}$ satisfy (30). Consider the other term:

$$\begin{aligned}
\sum_{x^n \in T_\delta^{X^n}} \frac{p_{X^n}(x^n)}{S} \|\hat{\rho}_{x^n} - \xi_{x^n}\|_1 &= \sum_{x^n \in T_\delta^{X^n}} p_{X^n}(x^n) \|\hat{\rho}_{x^n} - \xi'_{x^n} + \xi'_{x^n} - \xi_{x^n}\|_1 \\
&\leqslant \sum_{x^n \in T_\delta^{X^n}} p_{X^n}(x^n) \|\hat{\rho}_{x^n} - \xi'_{x^n}\|_1 + \sum_{x^n \in T_\delta^{X^n}} p_{X^n}(x^n) \|\xi'_{x^n} - \xi_{x^n}\|_1 \\
&\leqslant 2\sqrt{\epsilon'} + 2\sqrt{\epsilon''}, \qquad (35)
\end{aligned}$$

where we apply the triangle inequality in the third line. For the first bound with ϵ', we apply the Gentle Operator lemma (lemma 15) to the condition in (23), with $\epsilon' \equiv \epsilon + 2\sqrt{\epsilon}$. For the second bound with ϵ'', we exploit the equality $\xi_{x^n} = \Pi \xi'_{x^n} \Pi$ and apply the Gentle Operator lemma for ensembles (lemma 16) to the condition

$$\begin{aligned}
\sum_{x^n \in T_\delta^{X^n}} p_{X^n}(x^n) \mathrm{Tr}\{\Pi \xi_{x^n} \Pi\} &= \sum_{x^n \in T_\delta^{X^n}} p_{X^n}(x^n) \mathrm{Tr}\{\xi'_{x^n}\} \\
&= \mathrm{Tr}\{\Omega\} \\
&\geqslant 1 - \epsilon'',
\end{aligned}$$

which we proved before in (28) (with $\epsilon'' \equiv 2\epsilon + 2\sqrt{\epsilon}$). This concludes the proof of the achievability part of the measurement compression theorem with feedback.

2.4. Converse theorem for measurement compression

This section provides a proof of a version of the converse theorem, which states that the only achievable rates R and S of classical communication and common randomness consumption, respectively, are in the rate region given in theorem 5. We note that Winter proved a strong version of the converse theorem [65], which states that the error probability converges

exponentially to one as n becomes large. Winter's strong converse implies that the boundary of the rate region in theorem 5 is a very sharp dividing line. Here, for the sake of simplicity, we stick to the proof of the 'weak' converse, which only bounds the error probability away from zero. The reader can consult section IV of [65] for details of Winter's strong converse proof.

The converse theorem states that the 'single-letter' quantities in the rate region in theorem 5 are optimal. A nice consequence is that there is no need to evaluate an intractable regularization of the associated region, as is often the case for many coding theorems in quantum Shannon theory [60]. The theorem truly provides a complete understanding of the measurement compression task from an information-theoretic perspective.

We now prove the weak converse. Figure 2 depicts the most general protocol for measurement compression with feedback, and it proves to be useful here to consider a purification of the original input state. The protocol begins with the reference and Alice possessing the joint system $R^n A^n$ and Alice sharing the common randomness M with Bob. She then performs a simulation of the measurement, outputting a random variable L and another random variable X'^n that acts as the measurement output on her side. She sends L to Bob, and Bob produces $\hat{X}^n$ from L and the common randomness M. If the protocol is any good for measurement compression with feedback, then the resulting state $\omega^{R^n \hat{X}^n X'^n}$ should be ϵ-close in trace distance to the ideal state $\sigma^{R^n X^n \overline{X}^n}$ (the state resulting from the ideal protocol in figure 1), where $\overline{X}^n$ is a copy of the variable X^n:

$$\left\| \omega^{R^n \hat{X}^n X'^n} - \sigma^{R^n X^n \overline{X}^n} \right\|_1 \leqslant \epsilon. \tag{36}$$

We now prove the first lower bound on the classical communication rate R:

$$\begin{aligned} nR &\geqslant H(L) \\ &\geqslant I(L; MR^n) \\ &= I(LM; R^n) + I(L; M) - I(R^n; M) \\ &\geqslant I(LM; R^n) \\ &\geqslant I(\hat{X}^n; R^n)_\omega \\ &\geqslant I(X^n; R^n)_\sigma - n\epsilon' \\ &= nI(X; R) - n\epsilon'. \end{aligned}$$

The first inequality follows because the entropy of a uniform random variable is larger than the entropy of any other random variable. The second inequality follows because $I(L; MR^n) = H(L) - H(L|MR^n)$ and $H(L|MR^n) \geqslant 0$ for a classical variable L. The first equality is an easily verified identity for mutual information. The third inequality follows because the common randomness M is not correlated with the reference R^n (and hence $I(R^n; M) = 0$) and because $I(L; M) \geqslant 0$. The fourth inequality is from quantum data processing (Bob processes L and M to get $\hat{X}^n$). The fifth inequality is from (36) and continuity of quantum mutual information (the Alicki–Fannes' inequality [3]), where ϵ' is some function $f(\epsilon)$ such that $\lim_{\epsilon \to 0} f(\epsilon) = 0$. The final equality follows because the ideal state σ is a tensor-power state, and thus the mutual information $I(X^n; R^n)_\sigma$ is additive.

A proof for the lower bound on the sum rate $R + S$ goes as follows:

$$\begin{aligned} n(R+S) &\geqslant H(LM) \\ &\geqslant I(X'^n; LM) \\ &\geqslant I(X'^n; \hat{X}^n)_\omega \\ &\geqslant I(\overline{X}^n; X^n)_\sigma - n\epsilon' \\ &= H(X^n) - n\epsilon' \\ &= nH(X) - n\epsilon'. \end{aligned}$$

The first two inequalities follow for the same reasons as the first two above (we are assuming that copies of L and M are available since they are classical). The third inequality is quantum data processing. The fourth inequality follows from (36) and continuity of entropy. The first equality follows because the mutual information between a variable and a copy of it is equal to its entropy. The final equality follows because the entropy is additive for a tensor power state.

Optimality of the bound $R + S \geqslant H(X)$ for negative S follows by considering a protocol whereby Alice uses classical communication alone in order to simulate the measurement output X^n and generate common randomness M with Bob. The converse in this case proceeds as follows:

$$\begin{aligned} nR &\geqslant H(L) \\ &= I(\overline{L}; L) \\ &\geqslant I(X'^n M'; \hat{X}^n M) \\ &\geqslant I(\overline{X}^n \overline{M}; X^n M) - n\epsilon' \\ &= I(\overline{X}^n; X^n) + I(\overline{M}; M) - n\epsilon' \\ &= nH(X) + n|S| - n\epsilon'. \end{aligned}$$

The second inequality follows because Bob and Alice have to process L and its copy $\overline{L}$ in order to recover the approximate $\hat{X}^n M$ and $X'^n M'$, respectively. The third inequality follows because these systems should be close to the ideal ones for a good protocol (and applying continuity of entropy). The next equalities follow because the information quantities factor as above for the ideal state.

2.5. Extension to quantum instruments

We now briefly review Winter's argument for extending the above protocol from POVMs to quantum instruments. A quantum instrument is the most general model for quantum measurement that includes both a classical output and a post-measurement quantum state [20, 21, 48]. Our goal is now to simulate the action of a given quantum instrument on many copies of an input state ρ using as few resources as possible. The simulation should be such that Bob possesses the classical output at the end of the protocol (as in the case of POVM compression), and, as an additional requirement, Alice possesses the quantum output.

In the present setting, we can conveniently treat a quantum instrument as a completely positive, trace-preserving (CPTP) map $\mathcal{N}_{\text{instr}}$ of the form

$$\mathcal{N}_{\text{instr}}(\rho) \equiv \sum_x \mathcal{N}_x(\rho) \otimes |x\rangle\langle x|, \tag{37}$$

where each $\mathcal{N}_x$ is a completely positive, trace-non-increasing map of the form

$$\mathcal{N}_x(\rho) \equiv \sum_y N_{x,y} \rho N_{x,y}^\dagger,$$

such that

$$\sum_y N_{x,y}^\dagger N_{x,y} \leqslant I.$$

The simulation, implemented by a sequence of maps $\widetilde{\mathcal{N}^n_{\text{instr}}}$, is defined to be faithful if the following condition holds:

Definition 7 (Faithful instrument simulation)**.** *A sequence of maps $\widetilde{\mathcal{N}^n_{\text{instr}}}$ provides a faithful simulation of the quantum instrument $\mathcal{N}_{\text{instr}}$ on the state ρ if for all $\epsilon > 0$ and sufficiently*

large n, the action of the approximation channel on many copies of a purification ϕ_ρ of ρ is indistinguishable from the true quantum instrument, up to a factor of ϵ:

$$\left\| \left(\mathrm{id} \otimes \mathcal{N}_{\mathrm{instr}}^{\otimes n}\right)\left(\phi_\rho^{\otimes n}\right) - \left(\mathrm{id} \otimes \widetilde{\mathcal{N}_{\mathrm{instr}}^{n}}\right)\left(\phi_\rho^{\otimes n}\right) \right\|_1 \leqslant \epsilon. \tag{38}$$

In this case, we have the following theorem:

Theorem 8 (Instrument simulation)**.** *Let ρ be a source state and $\mathcal{N}_{\mathrm{instr}}$ an instrument to simulate on this state. A protocol for a faithful feedback simulation of $\mathcal{N}_{\mathrm{instr}}$ with classical communication rate R and common randomness rate S exists if and only if*

$$\begin{aligned} R &\geqslant I(X;R), \\ R + S &\geqslant H(X), \end{aligned}$$

where the entropies are with respect to a state of the following form:

$$\sum_x |x\rangle\langle x|^X \otimes \mathrm{Tr}_A\left\{\left(I^R \otimes \mathcal{N}_x^A\right)\left(\phi^{RA}\right)\right\}, \tag{39}$$

and ϕ^{RA} is some purification of the state ρ. The simulation is such that Alice possesses the quantum output of the channel and Bob possesses the classical output.

Proof. We just prove achievability because the converse theorem from the previous section applies to this case as well. We start by considering the case in which every map $\mathcal{N}_x$ can be written as

$$\mathcal{N}_x(\rho) = N_x \rho N_x^\dagger. \tag{40}$$

(The general case, stated in section V–G of [26] though lacking a formal proof, will also be addressed.)

We construct an approximation instrument using the protocol in the achievability proof from section 2.3. Let us set

$$\Lambda_x = N_x^\dagger N_x,$$

and construct the operators $\Gamma_{x^n}^{(m)}$ from Λ_x and the state ρ as in (32), such that they satisfy all of the properties that we had before. Define the distribution $p_X(x)$ and the states $\hat{\rho}_x$ as we did before:

$$\begin{aligned} p_X(x) &= \mathrm{Tr}\{\Lambda_x \rho\}, \\ \hat{\rho}_x &= \frac{1}{p_X(x)} \sqrt{\rho}\Lambda_x\sqrt{\rho}. \end{aligned}$$

We construct the approximation instrument $\widetilde{\mathcal{N}_{\mathrm{instr}}^{n}}$ from $\Gamma_{x^n}^{(m)}$, $p_X(x)$, $\hat{\rho}_x$, and the Kraus operators N_x. First, consider that the approximation instrument $\widetilde{\mathcal{N}_{\mathrm{instr}}^{n}}$ will be a convex combination of some other instruments:

$$\widetilde{\mathcal{N}_{\mathrm{instr}}^{n}}(\sigma) \equiv \frac{1}{|\mathcal{M}|} \sum_m \mathcal{E}_{\mathrm{instr}}^{(m)}(\sigma), \tag{41}$$

where

$$\mathcal{E}_{\mathrm{instr}}^{(m)}(\sigma) \equiv \sum_{x^n} F_{x^n}^{(m)} \sigma F_{x^n}^{(m)\dagger} \otimes |x^n\rangle\langle x^n|, \qquad \sum_{x^n} F_{x^n}^{(m)\dagger} F_{x^n}^{(m)} \leqslant I.$$

We now construct the operators $F_{x^n}^{(m)}$. Define the conditional distribution $p_{\widetilde{X}^n|M}(x^n|m)$ as follows:

$$p_{\widetilde{X}^n|M}(x^n|m) \equiv \frac{1}{|\mathcal{L}|} |\{l : x^n(l,m) = x^n\}|,$$

as in (32), and let $p_M(m) = 1/|\mathcal{M}|$, so that the marginal distribution $p_{\widetilde{X}^n}(x^n)$ is as follows:

$$\begin{aligned} p_{\widetilde{X}^n}(x^n) &= \frac{1}{|\mathcal{M}|}\sum_m p_{\widetilde{X}^n|M}(x^n|m) \\ &= \frac{1}{|\mathcal{L}||\mathcal{M}|}|\{l, m : x^n(l, m) = x^n\}|. \end{aligned}$$

Take a left polar decomposition of the operator $N_x\sqrt{\rho}$ and use it to define the unitary operator U_x:

$$N_x\sqrt{\rho} = U_x\sqrt{\sqrt{\rho}N_x^\dagger N_x\sqrt{\rho}} = U_x\sqrt{p_X(x)\hat{\rho}_x}. \tag{42}$$

Let U_{x^n} be as follows:

$$U_{x^n} \equiv U_{x_1} \otimes \cdots \otimes U_{x_n}.$$

We define the Kraus operators $F_{x^n}^{(m)}$ for the instruments $\mathcal{E}_{\text{instr}}^{(m)}$ as follows:

$$F_{x^n}^{(m)} \equiv U_{x^n}\sqrt{p_{\widetilde{X}^n|M}(x^n|m)\,\frac{S}{1+\epsilon}\,\xi_{x^n}}(\omega^{-1/2}). \tag{43}$$

One can check that these define completely positive trace-non-increasing instruments $\mathcal{E}_{\text{instr}}^{(m)}$—this follows from the fact that the operators $\Gamma_{x^n}^{(m)}$ in (32) form a sub-POVM. The instruments $\mathcal{E}_{\text{instr}}^{(m)}$ in turn form the instrument $\widetilde{\mathcal{N}}_{\text{instr}}^n$ via the relation in (41).

We can now check that this construction satisfies the condition in (38) for a faithful simulation. Consider a purification

$$\phi_\rho^{\otimes n} = (I \otimes \sqrt{\omega})|I\rangle\langle I|(I \otimes \sqrt{\omega}),$$

where $|I\rangle$ is the vector obtained by 'flipping the bra' of the identity channel $\sum_{x^n}|x^n\rangle\langle x^n|$:

$$|I\rangle \equiv \sum_{x^n}|x^n\rangle|x^n\rangle.$$

We have the following bound on the instrument simulation performance:

$$\begin{aligned} &\left\|\left(\mathrm{id}\otimes\mathcal{N}_{\text{instr}}^{\otimes n}\right)\left(\phi_\rho^{\otimes n}\right) - \left(\mathrm{id}\otimes\widetilde{\mathcal{N}_{\text{instr}}^n}\right)\left(\phi_\rho^{\otimes n}\right)\right\|_1 \\ &= \sum_{x^n}\left\|(I\otimes N_{x^n}\sqrt{\omega})|I\rangle\langle I|\left(I\otimes\sqrt{\omega}N_{x^n}^\dagger\right) - \frac{1}{|\mathcal{M}|}\sum_m\left(I\otimes F_{x^n}^{(m)}\sqrt{\omega}\right)|I\rangle\langle I|\left(I\otimes\sqrt{\omega}F_{x^n}^{(m)\dagger}\right)\right\|_1 \\ &= \sum_{x^n}\left\|\begin{array}{r} p_{X^n}(x^n)\left(I\otimes U_{x^n}\sqrt{\hat{\rho}_{x^n}}\right)|I\rangle\langle I|\left(I\otimes\sqrt{\hat{\rho}_{x^n}}U_{x^n}^\dagger\right) \\ -p_{\widetilde{X}^n}(x^n)\,\frac{S}{1+\epsilon}\,(I\otimes U_{x^n}\sqrt{\xi_{x^n}})|I\rangle\langle I|\left(I\otimes\sqrt{\xi_{x^n}}U_{x^n}^\dagger\right)\end{array}\right\|_1 \\ &= \sum_{x^n}\left\|p_{X^n}(x^n)(I\otimes\sqrt{\hat{\rho}_{x^n}})|I\rangle\langle I|(I\otimes\sqrt{\hat{\rho}_{x^n}}) - p_{\widetilde{X}^n}(x^n)\,\frac{S}{1+\epsilon}\,(I\otimes\sqrt{\xi_{x^n}})|I\rangle\langle I|(I\otimes\sqrt{\xi_{x^n}})\right\|_1 \\ &\leqslant \sum_{x^n}\left\|p_{X^n}(x^n)(I\otimes\sqrt{\hat{\rho}_{x^n}})|I\rangle\langle I|(I\otimes\sqrt{\hat{\rho}_{x^n}}) - p_{X^n}(x^n)(I\otimes\sqrt{\xi_{x^n}})|I\rangle\langle I|(I\otimes\sqrt{\xi_{x^n}})\right\|_1 \\ &\quad + \sum_{x^n}\left\|p_{X^n}(x^n)(I\otimes\sqrt{\xi_{x^n}})|I\rangle\langle I|(I\otimes\sqrt{\xi_{x^n}}) - p_{\widetilde{X}^n}(x^n)\frac{S}{1+\epsilon}(I\otimes\sqrt{\xi_{x^n}})|I\rangle\langle I|(I\otimes\sqrt{\xi_{x^n}})\right\|_1. \end{aligned}$$

The first equality follows from the block structure of the instruments with respect to the classical flags $|x^n\rangle\langle x^n|$. The second equality follows by substituting the polar decomposition in (42) and the definition in (43). The third equality follows from the unitary invariance of the

trace norm. The last inequality is the triangle inequality. Continuing, we have

$$\leqslant \sum_{x^n} p_{X^n}(x^n) \|(I \otimes \sqrt{\hat{\rho}_{x^n}})|I\rangle\langle I|(I \otimes \sqrt{\hat{\rho}_{x^n}}) - (I \otimes \sqrt{\xi_{x^n}})|I\rangle\langle I|(I \otimes \sqrt{\xi_{x^n}})\|_1$$

$$+ \sum_{x^n} \left| p_{X^n}(x^n) - p_{\widetilde{X}^n}(x^n) \frac{S}{1+\epsilon} \right|$$

$$\leqslant 2\sqrt{2} \sum_{x^n} p_{X^n}(x^n) \sqrt[4]{\|\hat{\rho}_{x^n} - \xi_{x^n}\|_1} + 2\epsilon$$

$$\leqslant 2\sqrt{2} \sqrt[4]{\sum_{x^n} p_{X^n}(x^n) \|\hat{\rho}_{x^n} - \xi_{x^n}\|_1} + 2\epsilon$$

$$\leqslant 2\sqrt{2} \sqrt[4]{\epsilon + 2\sqrt{\epsilon'} + 2\sqrt{\epsilon''}} + 2\epsilon.$$

The first inequality follows by factoring out the distribution $p_{X^n}(x^n)$ and because the positive operator $(I \otimes \sqrt{\xi_{x^n}})|I\rangle\langle I|(I \otimes \sqrt{\xi_{x^n}})$ has trace less than one. The second inequality follows from Winter's lemma 14:

$$\|(I \otimes \sqrt{\hat{\rho}_{x^n}})|I\rangle\langle I|(I \otimes \sqrt{\hat{\rho}_{x^n}}) - (I \otimes \sqrt{\xi_{x^n}})|I\rangle\langle I|(I \otimes \sqrt{\xi_{x^n}})\|_1 \leqslant \sqrt[4]{\|\hat{\rho}_{x^n} - \xi_{x^n}\|_1},$$

and our previous bound in (34). The third inequality is from concavity of the quartic-root function, and the final one follows from our previous bounds in (35) and the fact that the probability mass of the atypical set is upper bounded by ϵ.

We are now ready to consider the general case, in which one wants to simulate an instrument of the form (37). For this purpose, we require a slightly different coding strategy that combines ideas from section 2.3 and the above development.

First, consider that it is possible to implement a quantum instrument of the form in (37) by tracing over an auxiliary register Y:

$$\mathcal{N}_{\text{instr}}(\rho) = \text{Tr}_Y \left\{ \sum_{x,y} N_{x,y} \rho N_{x,y}^\dagger \otimes |x\rangle\langle x|^X \otimes |y\rangle\langle y|^Y \right\}.$$

So, Alice and Bob will simulate the following instrument

$$\sum_{x,y} N_{x,y} \rho N_{x,y}^\dagger \otimes |x\rangle\langle x|^X \otimes |y\rangle\langle y|^Y, \tag{44}$$

in such a way that Bob does not receive the outcome y, and thus they effectively implement the instrument $\mathcal{N}_{\text{instr}}$. The idea is that they will exploit a code with the following structure.

(1) Alice communicates $nI(X;R)$ bits of classical communication to Bob (enough for him to reconstruct the x output).
(2) Alice keeps $nI(Y;R|X)$ bits of the output to herself.
(3) Alice exploits $nH(X|R)$ bits of common randomness shared with Bob in the simulation.
(4) Alice uses $nH(Y|XR)$ bits of local, uniform randomness (not shared with Bob).

The entropies are with respect to the following classical-quantum state:

$$\sum_{x,y} \text{Tr}_A \left\{ \left(I^R \otimes \left(N_{x,y}^\dagger N_{x,y} \right)^A \right) \left(\phi_\rho^{RA} \right) \right\} \otimes |x\rangle\langle x|^X \otimes |y\rangle\langle y|^Y,$$

and observe that $I(X;R)$ and $H(X|R)$ are invariant with respect to the choice of Kraus operators $\{N_{x,y}\}$ for each map $\mathcal{N}_x$.

More precisely, the measurements used in the simulation are chosen randomly as in the proof in section 2.3, with the following modifications. Choose $|\mathcal{L}_1||\mathcal{M}_1|$ codewords $x^n(l_1, m_1)$ independently and randomly according to a pruned version of the distribution

$$p_X(x) \equiv \text{Tr}\{\mathcal{N}_x(\rho)\} = \sum_y \text{Tr}\left\{ N_{x,y}^\dagger N_{x,y} \rho \right\},$$

with

$$|\mathcal{L}_1| \approx 2^{nI(X;R)},$$
$$|\mathcal{M}_1| \approx 2^{nH(X|R)}.$$

For each pair (l_1, m_1), choose $|\mathcal{L}_2||\mathcal{M}_2|$ codewords $y^n(l_2, m_2|l_1, m_1)$ independently and randomly according to a distribution:

$$p_{Y^n|X^n}(y^n|x^n(l_1, m_1)),$$

which is a pruned version of the conditional distribution

$$p_{Y|X}(y|x) = \frac{1}{p_X(x)}\mathrm{Tr}\{N_{x,y}^\dagger N_{x,y}\rho\},$$

where

$$|\mathcal{L}_2| \approx 2^{nI(Y;R|X)},$$
$$|\mathcal{M}_2| \approx 2^{nH(Y|XR)}.$$

After choosing these codewords, we have a codebook $\{x^n(l_1, m_1), y^n(l_2, m_2|l_1, m_1)\}$. Divide all of these codewords into $|\mathcal{M}_1||\mathcal{M}_2|$ sets of the form $\{x^n(l_1, m_1), y^n(l_2, m_2|l_1, m_1)\}_{l_1,l_2}$. In order to have a faithful simulation, we require several conditions analogous to (29) and (30) to hold (except that the first average similar to that in (29) is over just l_1 and there is another over both l_1 and l_2, and the other operators like $\hat{C}$ in (30) are with respect to both l_1 and m_1 and all of l_1, l_2, m_1, and m_2). Choosing the sizes of the sets as we do above and applying the operator Chernoff bound several times guarantees that there exists a choice of the codebook $\{x^n(l_1, m_1), y^n(l_2, m_2|l_1, m_1)\}$ such that these conditions hold. By the development at the end of section 2.3 and the result for instruments of the special form (40), it follows that these conditions lead to a faithful simulation.

The simulation then operates by having the variable m_1 be common randomness shared with Bob, m_2 as additional local, uniform randomness that Alice uses for picking the measurement, and all of the measurements have outcomes l_1 and l_2. After performing the measurement simulation, Alice sends the outcome l_1 to Bob, which he can subsequently use to reconstruct the codeword $x^n(l_1, m_1)$ by combining with his share m_1 of the common randomness. The proof as we had it before goes through—the only difference is in constructing the codebook in such a way that the sequences x^n and y^n are separated out. The simulated instrument has a form like that in (44), and if Alice discards y^n, it follows, by applying the monotonicity of trace distance to the condition in (38), that Alice and Bob simulate the original instrument. □

As a closing note for this section, we would like to mention that the quantity $I(X; R)$, appearing in theorem 8, has a long history. Since $I(X; R)$ measures the amount of data *created* by the quantum measurement, contrarily to the shared randomness that exists before the measurement itself, it seems natural to consider it as a measure of the *information gain* produced by the quantum measurement. In this connection, the 1971 paper of Groenewold was the first to put forward the problem of measuring the information gain in 'quantal' measurements by means of information-theoretic quantities [33]. Groenewold considered the following quantity (reformulated according to our notation):

$$G(\rho, \mathcal{N}_{\mathrm{instr}}) := H(\rho) - \sum_x p_X(x) H\{\mathcal{N}_x(\rho)/p_X(x)\},$$

and he conjectured its positivity for von Neumann–Lüders measurements. Keep in mind that, at that time, the theory of quantum instruments was in its infancy, and the von Neumann–Lüders state reduction postulate, according to which the initial state is projected onto the

eigenspace corresponding to the observed outcome, was the only model of state reduction usually considered. Subsequently, Groenewold's conjecture was proved by Lindblad [42]. As the theory of quantum instruments advanced [48], quantum instruments with negative Groenewold's information gain appeared to be the rule, rather than the exception, until Ozawa finally settled the problem by proving that $G(\rho, \mathcal{N}_{\text{instr}})$ is nonnegative for all states ρ if and only if the quantum instrument has the special form in (40) [49].

The point is that, for quantum instruments of the form in (40), Groenewold's information gain $G(\rho, \mathcal{N}_{\text{instr}})$ is equal to $I(X; R)$ [12]. This is a consequence of the fact that, for any matrix K, $K^\dagger K$ and $KK^\dagger$ have the same eigenvalues (i.e. the squares of the singular values of K), so that

$$H\{\mathcal{N}_x(\rho)/p_X(x)\} \equiv H\big(N_x \rho N_x^\dagger / p_X(x)\big) = H\big(\sqrt{\rho} N_x^\dagger N_x \sqrt{\rho}/p_X(x)\big).$$

This coincidence retroactively strengthens the interpretation of $G(\rho, \mathcal{N}_{\text{instr}})$ as the information gain due to a quantum measurement, at least in the special case of instruments satisfying (40). In those cases, due to Winter's measurement compression theorem, $G(\rho, \mathcal{N}_{\text{instr}})$ truly is the rate at which the instrument generates information. More generally, however, $I(X; R)$ is the better measure of information gain both because it is nonnegative and because it *always* has the full strength of Winter's theorem behind it.

2.5.1. Application to channels. As already noticed in [65], with theorem 8 at hand, it is easy to consider the case in which one wants to simulate the action of some CPTP map $\mathcal{N}$ on many copies of the state ρ. The idea is that, for every Kraus representation [40] of the map $\mathcal{N}$ as

$$\mathcal{N}(\sigma) \equiv \sum_x N_x \sigma N_x^\dagger, \tag{45}$$

where N_x are a set of Kraus operators satisfying

$$\sum_x N_x^\dagger N_x = I,$$

one can apply theorem 8 and simulate the corresponding quantum instrument

$$\mathcal{N}_{\text{instr}}(\rho) = \sum_x N_x \rho N_x^\dagger \otimes |x\rangle\langle x|.$$

Then, any protocol faithfully simulating the above instrument automatically leads, by monotonicity of trace distance, to a faithful simulation of the channel $\mathcal{N}$, in the sense that it provides a sequence of maps $\widetilde{\mathcal{N}}^n$ such that:

$$\big\| (\mathrm{id} \otimes \mathcal{N}^{\otimes n})\big(\phi_\rho^{\otimes n}\big) - (\mathrm{id} \otimes \widetilde{\mathcal{N}}^n)\big(\phi_\rho^{\otimes n}\big) \big\|_1 \leqslant \epsilon,$$

for any $\epsilon > 0$ and sufficiently large n.

An important thing to stress is that the rates obtained in this way *depend* on the particular Kraus representation used to construct the instrument $\mathcal{N}_{\text{instr}}$. The rates of consumption of classical resources can hence be minimized over all possible Kraus representations of a given channel. However, such an optimization turns out to be difficult in general, as the following example shows.

Let us consider the case of a channel, which can be written as a mixture of unitaries, i.e.

$$\mathcal{N}(\rho) = \sum_x p(x) U_x \rho U_x^\dagger, \tag{46}$$

where $U_x^\dagger U_x = I$. Such a channel can be simulated without the need for classical communication. This follows simply from the fact that the quantum instrument constructed from (46) corresponds to measuring the POVM $\Lambda_x = p(x)I$, whose outcomes are completely

random and uncorrelated with the reference, so that $I(X;R) = 0$. In fact, the converse is also true: if a given channel admits a Kraus decomposition for which $I(X;R) = 0$, then its action on the state ρ can be written as a mixture of unitaries as in (46) [11]. In order to show this, suppose that we find a Kraus decomposition $\mathcal{N}(\rho) = \sum_x N_x \rho N_x^\dagger$ such that the quantum mutual information $I(X;R) = 0$, where it is calculated with respect to the following classical-quantum state

$$\sum_x |x\rangle\langle x|^X \otimes \mathrm{Tr}_A\left\{\left(I^R \otimes N_x^\dagger N_x^A\right)\phi^{RA}\right\},$$

and ϕ^{RA} is a purification of ρ. Adopting the same notation used at the beginning of section 2.3, we know that $I(X;R) = 0$ if and only if the sub-normalized states $\theta_x^R = \sqrt{\rho}\left(N_x^\dagger N_x\right)^T\sqrt{\rho}$ are all proportional to ρ. This is possible if and only if the operators $(N_x^\dagger N_x)^T$ are all proportional to the identity (on the support of ρ), thus proving the claim.

Hence, as the above example shows, to minimize the rate of classical communication needed to simulate a quantum channel constitutes a task of complexity comparable to that of deciding whether a given channel possesses a random-unitary Kraus decomposition or not, for which numerical methods are known [4] but a general analytical solution has yet to be found.

3. Non-feedback measurement compression

We now discuss an extension of Winter's measurement compression theorem in which the sender is not required to obtain the outcome of the measurement simulation (known as a 'non-feedback' simulation). Achieving a feedback simulation is more demanding than one without feedback, so we should expect the non-feedback problem to show some reduction in the resources required. To get a sense of where the improvement comes from will require considering a more general type of POVM decomposition than that in (3). Suppose that it is possible to decompose a POVM $\{\Lambda_x\}$ in terms of a random selection according to a random variable M, an 'internal' POVM $\{\Gamma_w^{(m)}\}$ with outcomes w, and a classical post-processing map $p_{X|W}(x|w)$ [45, 13]:

$$\Lambda_x = \sum_{m,w} p_M(m)\Gamma_w^{(m)} p_{X|W}(x|w). \tag{47}$$

In that case, Alice and Bob could proceed with a protocol along the following lines: they use Winter's measurement compression protocol to simulate the POVM $\{\sum_m p_M(m)\Gamma_w^{(m)}\}_w$ and Bob locally simulates the classical postprocessing map $p_{X|W}(x|w)$. (This is essentially how a 'non-feedback' simulation will proceed, but there are some details to be worked out.)

We should compare the performance of the above protocol against one that exploits a feedback simulation for $\{\Lambda_x\}$. The classical communication cost will increase from $I(X;R)$ to $I(W;R)$ (the data-processing inequality $I(W;R) \geqslant I(X;R)$ holds because W is 'closer' to R than is X), but the common randomness cost will be cheaper because the non-feedback protocol requires only $nI(W;X|R)$ bits of common randomness rather than $nH(X|R)$ bits (essentially because Bob can find a clever way to simulate the local map $p_{X|W}(x|w)$). Thus, 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. In general, decomposing a POVM in many different ways according to (47) leads to a non-trivial curve characterizing the trade-off between classical communication and common randomness.

In this connection, it is important to remark that the decomposition (sometimes referred to as a *refinement*) of a POVM according to the post-processing relation:

$$\Lambda_x = \sum_w \Xi_w p_{X|W}(x|w), \tag{48}$$

of which (47) is a special case, is different from the convex decomposition described in (3). In particular, while the conditions for a POVM to be extremal (i.e. not non-trivially decomposable) with respect to (3) are known to be rather involved [18], it turns out that POVMs which are extremal with respect to (48) can be neatly characterized as those (and only those) whose elements are all rank-1 operators [45]. Hence, if the POVM that Alice and Bob want to simulate is rank-1 (i.e. all its elements are rank-1 operators), then there is nothing to gain from implementing a non-feedback simulation instead of a feedback simulation. Notice, however, that the two decompositions (3) and (48) are completely independent: POVMs which are extremal with respect to (3) need not also be extremal with respect to (48), and vice versa [13]. This is the reason why there is plenty of room for non-trivial trade-off relations between classical communication and common randomness if the POVM to be simulated is not rank-1.

Theorem 9 below gives a full characterization of the trade-off for a nonfeedback measurement compression protocol, in the sense that the protocol summarized above has a matching single-letter converse proof for its optimality. Thus, we can claim to have a complete understanding of this task from an information-theoretic perspective.

We should mention that some of the above ideas regarding non-feedback simulation are already present in prior works [26, 17, 5], and indeed, these works are what led us to pursue a non-feedback measurement compression protocol. In [26], Devetak *et al* observed in their remarks around equations (43)–(45) of their paper that a protocol in which the sender also receives the outcomes of the simulation is optimal, but 'examples are known in which less randomness is necessary' for protocols that do not have this restriction. They did not state any explicit examples, however, nor did they state that there would be a general theorem characterizing the trade-off in the non-feedback case. Cuff's theorem [17] regarding the trade-off between classical communication and common randomness for a non-feedback reverse Shannon theorem is a special case of theorem 9 below, essentially because a noisy classical channel is a special case of a quantum measurement and thus the simulation task is a special case. Reference [5] characterized the trade-off between quantum communication and entanglement for a non-feedback simulation of a quantum channel. Thus, theorem 9 below 'sits in between' the communication tasks considered in [17] and [5]. We should also remark that [5] stated that it is possible to reduce the common randomness cost in the non-feedback reverse Shannon theorem either with randomness recycling or by derandomizing some of it, and we should be able to employ these approaches in a non-feedback measurement compression protocol. Though, our approach below is to modify Winter's original protocol directly by changing the structure of the code.

3.1. Non-feedback measurement compression theorem

Theorem 9 (Non-feedback measurement compression)**.** *Let ρ be a source state and $\mathcal{N}$ a quantum instrument to simulate on this state:*

$$\mathcal{N}(\rho) = \sum_x \mathcal{N}_x(\rho) \otimes |x\rangle\langle x|^X.$$

A protocol for faithful non-feedback simulation of the quantum instrument with classical communication rate R and common randomness rate S exists if and only if R and S are in the rate region given by the union of the following regions:

$$R \geqslant I(W;R),$$
$$R + S \geqslant I(W;XR),$$

where the entropies are with respect to a state of the following form:

$$\sum_{x,w} p_{X|W}(x|w)|w\rangle\langle w|^W \otimes |x\rangle\langle x|^X \otimes Tr_A\left\{\left(I^R \otimes \mathcal{M}_w^A\right)\left(\phi_\rho^{RA}\right)\right\}, \tag{49}$$

ϕ_ρ^{RA} is some purification of the state ρ, and the union is with respect to all decompositions of the original instrument $\mathcal{N}$ of the form:

$$\mathcal{N}(\rho) = \sum_{x,w} p_{X|W}(x|w)\mathcal{M}_w(\rho) \otimes |x\rangle\langle x|^X. \tag{50}$$

Observe that the systems R, W, and X in (49) form a quantum Markov chain: $R - W - X$.

The information quantity $I(W; XR)$ appearing in the above theorem generalizes Wyner's well-known 'common information' between dependent random variables [67].

3.2. Achievability for non-feedback measurement compression

We now prove the achievability part of the above theorem. Suppose for simplicity that we are just trying to simulate the POVM $\Lambda = \{\Lambda_x\}$ where each Λ_x is a positive operator such that $\Lambda_x = \sum_w p_{X|W}(x|w)M_w$ and each M_w is a positive operator. The case for a general quantum instrument follows by considering this case and by extending it similarly to how we extended POVM compression to instrument compression in theorem 8. So, the relevant overall classical-quantum state to consider when building codes for a non-feedback simulation is as follows:

$$\sum_{w,x} p_{X|W}(x|w)\mathrm{Tr}_A\left\{M_w^A \phi_\rho^{RA}\right\} \otimes |w\rangle\langle w|^W \otimes |x\rangle\langle x|^X,$$

which simplifies to

$$\sum_{w,x} p_{X|W}(x|w)\ \sqrt{\rho}M_w\sqrt{\rho} \otimes |w\rangle\langle w|^W \otimes |x\rangle\langle x|^X,$$

after realizing that $\mathrm{Tr}_A\left\{M_w^A \phi_\rho^{RA}\right\} = \sqrt{\rho}M_w\sqrt{\rho}$ (in the above and what follows, we ignore the transpose in the eigenbasis of ρ on M_w because it is irrelevant for the result). Let τ denote the state obtained by tracing over the W register of the above state:

$$\tau \equiv \sum_w \sqrt{\rho}M_w\sqrt{\rho} \otimes \sigma_w^X,$$

where the classical state σ_w^X is as follows:

$$\sigma_w^X \equiv \sum_x p_{X|W}(x|w)\ |x\rangle\langle x|^X.$$

Consider the following ensemble:

$$p_W(w) \equiv \mathrm{Tr}\{M_w\rho\},$$
$$\hat{\rho}_w \equiv \frac{1}{p_W(w)}\sqrt{\rho}M_w\sqrt{\rho}.$$

Observe that ρ is the expected state of this ensemble:

$$\sum_w p_W(w)\hat{\rho}_w = \sum_w \sqrt{\rho}M_w\sqrt{\rho} = \rho.$$

Also, the state τ is as follows:

$$\tau = \sum_w p_W(w)\hat{\rho}_w \otimes \sigma_w^X.$$

Our approach is similar to Winter's approach detailed in section 2.3: choose $|\mathcal{L}||\mathcal{M}|$ codewords $w^n(l, m)$ according to the pruned version of the distribution $p_{W^n}(w^n)$. As long as

$$\begin{aligned} |\mathcal{L}| &\approx 2^{nI(W;R)}, \\ |\mathcal{M}| &\approx 2^{nI(W;X|R)}, \\ |\mathcal{L}||\mathcal{M}| &\approx 2^{nI(W;XR)}, \end{aligned}$$

the operator Chernoff bound (lemma 6) guarantees that there exists a choice of the codewords $w^n(l, m)$ such that the following conditions are true:

$$\frac{1}{|\mathcal{L}|}\sum_l \hat{\rho}'_{w^n(l,m)} \in [(1 \pm \epsilon)\rho'^n], \tag{51}$$

$$\frac{1}{|\mathcal{L}||\mathcal{M}|}\sum_{l,m} \kappa_{w^n(l,m)} \in [(1 \pm \epsilon)\tau'^n], \tag{52}$$

where each $\hat{\rho}'_{w^n}$ is a typical projected version of $\hat{\rho}_{w^n} \equiv \hat{\rho}_{w_1} \otimes \cdots \otimes \hat{\rho}_{w_n}$:

$$\hat{\rho}'_{w^n} \equiv \Pi\ \Pi^n_{\rho,\delta}\ \Pi_{\hat{\rho}_{w^n},\delta}\ \hat{\rho}_{w^n}\ \Pi_{\hat{\rho}_{w^n},\delta}\ \Pi^n_{\rho,\delta}\ \Pi.$$

In the above, $\Pi_{\hat{\rho}_{w^n},\delta}$ is the conditionally typical projector for $\hat{\rho}_{w^n}$, $\Pi^n_{\rho,\delta}$ is the average typical projector for $\rho^{\otimes n}$, and Π is the eigenvalue cutoff projector as before. We define each $\kappa_{w^n(l,m)}$ as a typical projected version of the state $\hat{\rho}_{w^n(l,m)} \otimes \sigma_{w^n(l,m)}$:

$$\kappa_{w^n} \equiv \Pi'\ \Pi^n_{\tau,\delta}\ (\Pi_{\hat{\rho}_{w^n},\delta} \otimes \Pi_{\sigma_{w^n},\delta})\ \hat{\rho}_{w^n} \otimes \sigma_{w^n}\ (\Pi_{\hat{\rho}_{w^n},\delta} \otimes \Pi_{\sigma_{w^n},\delta})\ \Pi^n_{\tau,\delta}\ \Pi'.$$

In the above, $\Pi_{\hat{\rho}_{w^n},\delta} \otimes \Pi_{\sigma_{w^n},\delta}$ is the conditionally typical projector for $\hat{\rho}_{w^n} \otimes \sigma_{w^n}$, $\Pi^n_{\tau,\delta}$ is an average typical projector for τ, and Π' is another eigenvalue cutoff projector. The states ρ'^n and τ'^n are the expectations of the states $\hat{\rho}'_{w^n}$ and κ_{w^n}, respectively, with respect to the pruned version of the distribution $p_{W^n}(w^n)$. Recall that the operator Chernoff bound guarantees with high probability that the sample averages $\frac{1}{|\mathcal{L}|}\sum_l \hat{\rho}'_{w^n(l,m)}$ and $\frac{1}{|\mathcal{L}||\mathcal{M}|}\sum_{l,m} \kappa_{w^n(l,m)}$ are within ϵ (in the operator interval sense) of their true expectations ρ'^n and τ'^n, respectively. Thus there exist particular values of the $w^n(l, m)$ such that the above conditions are all true. We can use the condition in (51) to guarantee that the following defines a legitimate POVM (just as in Winter's approach in section 2.3):

$$\Upsilon^{(m)}_l \equiv \frac{S}{1+\epsilon}\frac{1}{|\mathcal{L}|}\omega^{-1/2}\hat{\rho}'_{w^n(l,m)}\omega^{-1/2},$$

where S is the mass of the typical set corresponding to the distribution $p_{W^n}(w^n)$ and $\omega = \rho^{\otimes n}$. Also, observe that the following states are close in trace distance for sufficiently large n, due to quantum typicality and the Gentle Operator lemma:

$$\left\|\hat{\rho}'_{w^n} - \hat{\rho}_{w^n}\right\|_1 \leqslant f_1(\epsilon), \tag{53}$$

$$\|\kappa_{w^n} - \hat{\rho}_{w^n} \otimes \sigma_{w^n}\|_1 \leqslant f_2(\epsilon). \tag{54}$$

Here and in what follows, $f_i(\epsilon)$ is some polynomial in ϵ so that $\lim_{\epsilon\to 0} f_i(\epsilon) = 0$.

We use the conditions in (51) and (52) to guarantee that the simulation is faithful. The protocol proceeds as follows: Alice and Bob use the common randomness M to select a POVM $\Upsilon^{(m)}$. Alice performs a measurement and gets the outcome l (corresponding to the operator $\Upsilon^{(m)}_l$). She sends the index l to Bob, who then prepares the classical state $\sigma_{w^n(l,m)}$ based on

the common randomness m and the measurement outcome l. Consider the following chain of equalities:

$$\begin{aligned}\sum_{x^n} \mathrm{Tr}_{A^n}\left\{(\mathrm{id}\otimes\Lambda_{x^n})\left(\phi_\rho^{\otimes n}\right)\right\}\otimes|x^n\rangle\langle x^n| &= \sum_{w^n,x^n} \mathrm{Tr}_{A^n}\left\{(\mathrm{id}\otimes M_{w^n})\left(\phi_\rho^{\otimes n}\right)\right\}\otimes p_{X^n|W^n}(x^n|w^n)|x^n\rangle\langle x^n| \\ &= \sum_{w^n,x^n} \sqrt{\omega}M_{w^n}\sqrt{\omega}\otimes p_{X^n|W^n}(x^n|w^n)|x^n\rangle\langle x^n| \\ &= \tau^{\otimes n},\end{aligned}$$

$$\sum_{m,l}\frac{1}{|\mathcal{M}|}\mathrm{Tr}_{A^n}\left\{\left(\mathrm{id}\otimes\Upsilon_l^{(m)}\right)\left(\phi_\rho^{\otimes n}\right)\right\}\otimes\sigma_{w^n(l,m)} = \frac{1}{|\mathcal{M}||\mathcal{L}|}\sum_{m,l}\frac{S}{1+\epsilon}\hat{\rho}'_{w^n(l,m)}\otimes\sigma_{w^n(l,m)}.$$

By exploiting (52), that $\|\tau^{\otimes n}-\tau'^n\|_1\leqslant f_3(\epsilon)$, and that $\|\rho^{\otimes n}-\rho'^n\|_1\leqslant f_4(\epsilon)$, we have that

$$\left\|\tau^{\otimes n}-\frac{1}{|\mathcal{M}||\mathcal{L}|}\sum_{l,m}\kappa_{w^n(l,m)}\right\|_1\leqslant f_5(\epsilon).$$

Also, we have that

$$\left\|\frac{1}{|\mathcal{M}||\mathcal{L}|}\sum_{m,l}\frac{S}{1+\epsilon}\hat{\rho}'_{w^n(l,m)}\otimes\sigma_{w^n(l,m)}-\frac{1}{|\mathcal{M}||\mathcal{L}|}\sum_{m,l}\hat{\rho}'_{w^n(l,m)}\otimes\sigma_{w^n(l,m)}\right\|_1\leqslant f_6(\epsilon).$$

From (53) and (54) and convexity of trace distance, we have that

$$\left\|\frac{1}{|\mathcal{M}||\mathcal{L}|}\sum_{m,l}\kappa_{w^n(l,m)}-\frac{1}{|\mathcal{M}||\mathcal{L}|}\sum_{m,l}\hat{\rho}'_{w^n(l,m)}\otimes\sigma_{w^n(l,m)}\right\|_1\leqslant f_7(\epsilon).$$

Putting all of these together with the triangle inequality gives an upper bound on the trace distance between the ideal output of the measurement and the state resulting from the simulation:

$$\left\|\sum_{x^n}\mathrm{Tr}_{A^n}\left\{(\mathrm{id}\otimes\Lambda_{x^n})\left(\phi_\rho^{\otimes n}\right)\right\}\otimes|x^n\rangle\langle x^n|-\sum_{m,l}\frac{1}{|\mathcal{M}|}\mathrm{Tr}_{A^n}\left\{\left(\mathrm{id}\otimes\Upsilon_l^{(m)}\right)\left(\phi_\rho^{\otimes n}\right)\right\}\otimes\sigma_{w^n(l,m)}\right\|_1 \leqslant f_8(\epsilon).$$

The case for a general quantum instrument follows by similar reasoning as that in the proof of theorem 8.

3.3. Converse for non-feedback measurement compression

We now prove the converse part of theorem 9. Our proof is similar to Cuff's converse proof for the non-feedback version of the classical reverse Shannon theorem [17]. Figure 2 can serve as a depiction of the most general protocol for a non-feedback simulation, if we ignore the decoding on Alice's side to produce X'^n. The non-feedback protocol begins with Alice and the reference sharing many copies of the state ϕ_ρ^{RA} and Alice sharing common randomness M with Bob. She then chooses a quantum instrument $\Upsilon^{(m)}$ based on the common randomness M and performs it on her systems A^n. The measurement returns outcome L, and the overall state is as follows:

$$\sum_{l,m}\frac{1}{|\mathcal{M}|}\left(\Upsilon_l^{(m)}\right)^{A^n}\left(\left(\phi_\rho^{RA}\right)^{\otimes n}\right)\otimes|l\rangle\langle l|^L\otimes|m\rangle\langle m|^M,$$

where $\Upsilon_l^{(m)}$ is a completely positive, trace non-increasing map. Alice sends the register L to Bob. Based on L and M, he performs some stochastic map $p_{\hat{X}^n|L,M}(\hat{x}^n|l,m)$ to give his estimate $\hat{x}^n$ of the measurement outcome. The resulting state is as follows:

$$\omega^{R^nA^nLM\hat{X}^n} \equiv \sum_{l,m,\hat{x}^n} \frac{1}{|\mathcal{M}|} p_{\hat{X}^n|L,M}(\hat{x}^n|l,m)\left(\Upsilon_l^{(m)}\right)^{A^n}\left(\left(\phi_\rho^{RA}\right)^{\otimes n}\right) \otimes |l\rangle\langle l|^L \otimes |m\rangle\langle m|^M \otimes |\hat{x}^n\rangle\langle\hat{x}^n|^{\hat{X}^n}.$$

The following condition should hold for all $\epsilon > 0$ and sufficiently large n for a faithful non-feedback simulation:

$$\left\| \omega^{R^n\hat{X}^n} - \sum_{x^n} \mathrm{Tr}_{A^n}\left\{(I\otimes\mathcal{N}_{x^n})\left(\phi_\rho^{RA}\right)^{\otimes n}\right\} \otimes |x^n\rangle\langle x^n|^{X^n} \right\|_1 \leqslant \epsilon.$$

We prove the first bound as follows:

$$\begin{aligned}
nR &\geqslant H(L)_\omega \\
&\geqslant I(L;MR^n)_\omega \\
&= I(LM;R^n)_\omega + I(L;M)_\omega - I(M;R^n)_\omega \\
&\geqslant I(LM;R^n)_\omega \\
&= H(R^n)_\omega - H(R^n|LM)_\omega \\
&\geqslant \sum_k [H(R_k)_\omega - H(R_k|LM)_\omega] \\
&= \sum_k I(LM;R_k)_\omega \\
&= nI(LM;R|K)_\sigma \\
&\geqslant nI(LM;R|K)_\sigma + nI(R;K)_\sigma - n\epsilon' \\
&= nI(LMK;R)_\sigma - n\epsilon'.
\end{aligned}$$

The first two inequalities follow for reasons similar to the first few steps of our previous converse. The first equality is an identity for quantum mutual information. The third inequality follows because there are no correlations between R^n and M so that $I(M;R^n)_\omega = 0$. The second equality is an identity for quantum mutual information. The fourth inequality follows from subadditivity of quantum entropy:

$$H(R^n|LM)_\omega \leqslant \sum_k H(R_k|LM)_\omega,$$

and because the state on R^n is a tensor-power state so that

$$H(R^n)_\omega = \sum_k H(R_k)_\omega.$$

The third equality is another identity. The fourth equality comes about by defining the state σ as follows:

$$\begin{aligned}
\sigma^{RALM\hat{X}K} \equiv \sum_{l,m,k,\hat{x}} \frac{1}{n|\mathcal{M}|} p_{\hat{X}_k|L,M}(\hat{x}|l,m)\mathrm{Tr}_{R_1^{k-1}R_{k+1}^nA_1^{k-1}A_{k+1}^n}\left\{\left(\Upsilon_l^{(m)}\right)^{A^n}\left(\left(\phi_\rho^{RA}\right)^{\otimes n}\right)\right\} \otimes \\
|l\rangle\langle l|^L \otimes |m\rangle\langle m|^M \otimes |\hat{x}\rangle\langle\hat{x}|^{\hat{X}} \otimes |k\rangle\langle k|^K, \qquad (55)
\end{aligned}$$

and exploiting the fact that K is a uniform classical random variable, with distribution $1/n$, determining which systems $R_kA_k\hat{X}_k$ to select. (The notation $\mathrm{Tr}_{R_i^j}$ with $i \leqslant j$ indicates to trace over systems $R_i \ldots R_j$.) From the fact that the measurement simulation is faithful, we can apply the Alicki–Fannes' inequality to conclude that

$$I(R\hat{X};K)_\sigma = |I(R\hat{X};K)_\sigma - I(RX;K)_\tau| \leqslant \epsilon', \qquad (56)$$

where τ is a state like σ but resulting from the tensor-power state for ideal measurement compression (and due to its IID structure, it has no correlations with any particular system k so that $I(RX;K)_\tau = 0$). The above also implies that

$$I(R;K)_\sigma \leqslant \epsilon',$$

by quantum data processing. The final equality is an application of the chain rule for quantum mutual information. The state σ for the final information term has the form in (50) with $LMK = W$, the distribution $p_{X|W}(x|w)$ as

$$p_{\hat{X}_k|L,M}(\hat{x}|l,m),$$

and the completely positive, trace non-increasing maps $\mathcal{M}_w$ defined by

$$\varrho^A \to \frac{1}{n|\mathcal{M}|}\mathrm{Tr}_{A_1^{k-1}A_{k+1}^n}\left\{\left(\Upsilon_l^{(m)}\right)^{A^n}\left(\left(\phi_\rho^A\right)^{\otimes k-1}\otimes\varrho^A\otimes\left(\phi_\rho^A\right)^{\otimes n-k}\right)\right\}.$$

Also, observe that $R-(LMK)-\hat{X}$ forms a quantum Markov chain.

We now prove the second bound:

$$\begin{aligned}
n(R+S) &\geqslant H(LM)_\omega \\
&\geqslant I(LM;\hat{X}^nR^n)_\omega \\
&= H(\hat{X}^nR^n)_\omega - H(\hat{X}^nR^n|LM)_\omega \\
&\geqslant \sum_k [H(\hat{X}_kR_k)_\omega - H(\hat{X}_kR_k|LM)_\omega] - n\epsilon' \\
&= \sum_k I(LM;\hat{X}_kR_k)_\omega - n\epsilon' \\
&= nI(LM;\hat{X}R|K)_\sigma - n\epsilon' \\
&\geqslant nI(LM;\hat{X}R|K)_\sigma + nI(K;\hat{X}R)_\sigma - n2\epsilon' \\
&= nI(LMK;\hat{X}R)_\sigma - n2\epsilon'.
\end{aligned}$$

The first two inequalities follow from similar reasons as our previous inequalities. The first equality is an identity. The third inequality follows from subadditivity of entropy:

$$H(\hat{X}^nR^n|LM)_\omega \leqslant \sum_k H(\hat{X}_kR_k|LM)_\omega,$$

and from the fact that the measurement simulation is faithful so that

$$\left|H(\hat{X}^nR^n)_\omega - \sum_k H(\hat{X}_kR_k)_\omega\right| \leqslant n\epsilon',$$

where we have applied lemma 10 below. The second equality is an identity. The third equality follows by considering the state σ as defined in (55). The fourth inequality follows from (56). The final equality is the chain rule for quantum mutual information. Similarly as stated above, the state σ has the form in (50).

Lemma 10. *Suppose that a state ρ^{A^n} is ϵ-close in trace distance to an IID state $(\sigma^A)^{\otimes n}$:*

$$\|\rho^{A^n} - (\sigma^A)^{\otimes n}\|_1 \leqslant \epsilon. \tag{57}$$

Then the entropy $H(A^n)_\rho$ of ρ^{A^n} and the entropy $\sum_k H(A_k)_\rho$ are close in the following sense:

$$\left|H(A^n)_\rho - \sum_k H(A_k)_\rho\right| \leqslant 2n\epsilon\log|A| + (n+1)H_2(\epsilon).$$

Proof. Apply the Fannes–Audenart inequality to (57) to obtain

$$|H(A^n)_\rho - H(A^n)_{\sigma^{\otimes n}}| \leqslant \epsilon n \log |A| + H_2(\epsilon). \tag{58}$$

The following inequality also follows by applying monotonicity of trace distance to (57):

$$\|\rho^{A_k} - \sigma^A\|_1 \leqslant \epsilon,$$

which then gives that

$$|H(A_k)_\rho - H(A)_\sigma| \leqslant \epsilon \log |A| + H_2(\epsilon),$$

by again applying the Fannes–Audenart inequality. Summing these over all k then gives that

$$\sum_{k=1}^{n} |H(A)_\sigma - H(A_k)_\rho| \leqslant n\epsilon \log |A| + nH_2(\epsilon). \tag{59}$$

Applying the triangle inequality to (58) and (59) gives the desired result:

$$\begin{aligned}
2n\epsilon \log |A| + (n+1)H_2(\epsilon) &\geqslant |H(A^n)_\rho - H(A^n)_{\sigma^{\otimes n}}| + \sum_{k=1}^{n} |H(A)_\sigma - H(A_k)_\rho| \\
&\geqslant \left| H(A^n)_\rho - H(A^n)_{\sigma^{\otimes n}} + \sum_{k=1}^{n} [H(A)_\sigma - H(A_k)_\rho] \right| \\
&= \left| H(A^n)_\rho - H(A^n)_{\sigma^{\otimes n}} + H(A^n)_{\sigma^{\otimes n}} - \sum_{k=1}^{n} H(A_k)_\rho \right| \\
&= \left| H(A^n)_\rho - \sum_{k=1}^{n} H(A_k)_\rho \right|.
\end{aligned}$$

□

4. Classical data compression with quantum side information

We now turn to the third protocol of this review: classical data compression with quantum side information. We discuss this protocol in detail because it is a step along the way to constructing our protocol for measurement compression with quantum side information (and we have a particular way that we construct this latter protocol). Devetak and Winter first proved achievability and optimality of a protocol for this task [27]. They proved this result by appealing to Winter's proof of the classical capacity theorem [63] and a standard recursive code construction argument of Csiszàr and Körner [16]. Renes *et al* later gave a proof of this protocol by exploiting two-universal hash functions and a square-root measurement [51, 54] (the first paper proved the IID version and the latter the 'one-shot' case). Renes *et al* further explored a connection between this protocol and privacy amplification by considering entropic uncertainty relations [50, 53].

Our development here contains a review of this information processing task and the statement of the theorem, in addition to providing novel proofs of both the achievability part and the converse that are direct quantum generalizations of the well-known approaches in [15, 30] for the Slepian–Wolf problem [58]. The encoder in our achievability proof bears some similarities with those in [15, 30, 54]—the protocol has the sender first hash the received sequence and send the hash along to the receiver. The receiver then employs a sequential quantum decoder—he searches sequentially among all the possible quantum states that are consistent with the hash in order to determine the sequence emitted by the source. The main tool employed in the error analysis is Sen's non-commutative union bound [56]. A potential advantage of a sequential decoding approach is that it might lead to physical implementations of these protocols for small block sizes, along the lines discussed in [61].

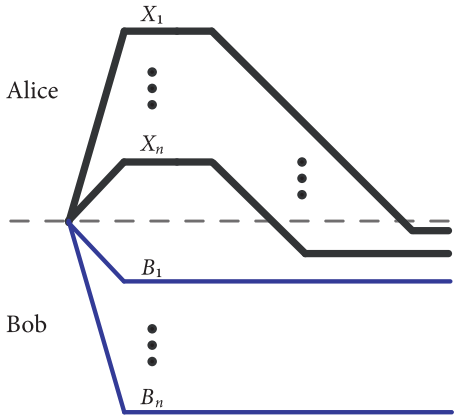

Figure 4. *Ideal protocol for classical data compression with quantum side information.* In this protocol, we assume that a quantum information source distributes many copies of a classical-quantum state to Alice and Bob, such that Alice receives the classical part and Bob receives the quantum part. The goal is for Alice to communicate the classical sequence received from the source to Bob. In an ideal case, she would simply transmit this sequence to Bob. Though, it is possible to obtain a significant savings in communication by allowing for an asymptotically vanishing error and for Bob to infer something about the classical sequence from his correlated quantum states.

4.1. Information processing task for CDC with QSI

We now discuss the general information processing task. Consider an ensemble $\{p_X(x), \rho_x\}$. Suppose that a source issues a random sequence X^n to Alice, distributed according to the IID distribution $p_{X^n}(x^n)$, while also issuing the correlated quantum state ρ_{X^n} to Bob. Their joint state is described by the ensemble $\{p_{X^n}(x^n), \rho_{x^n}\}$, or equivalently, by a classical-quantum state of the following form:

$$\sum_{x^n} p_{X^n}(x^n)|x^n\rangle\langle x^n|^{X^n} \otimes \rho_{x^n}^{B^n}.$$

The goal is for Alice to communicate the particular sequence x^n that the source issues, by using as few bits as possible. Figure 4 depicts the ideal result of a protocol for CDC–QSI.

One potential strategy is to exploit Shannon compression—just compress the sequence to $nH(X)$ bits, keeping only the typical set according to the distribution $p_{X^n}(x^n)$. But they can actually do much better in general if Bob exploits his quantum side information in the form of the correlated state ρ_{x^n}.

The most general protocol has Alice hash her sequence x^n to some variable $L \in \mathcal{L}$ (this is just some many-to-one mapping $f : \mathcal{X}^n \to \mathcal{L}$). She transmits the variable L to Bob over a noiseless classical channel using $\log_2 |\mathcal{L}|$ bits. Bob then exploits the hashed variable L and his quantum side information ρ_{x^n} to distinguish between all of the possible states that are consistent with the hash L (i.e. his action will be some quantum measurement depending on the hash L). The output of his measurement is some approximation sequence $\hat{X}^n$. The protocol has one parameter that characterizes its quality. We demand that the state $\sigma^{X^n\hat{X}^nB^n}$ after Alice and Bob's actions should be close in trace distance to an ideal state $\rho^{X^n\overline{X}^nB^n}$, where $\overline{X}^n$ is a copy of X^n (this would be the ideal output if Alice were to just send a copy of the variable X^n to Bob):

$$\|\rho^{X^n\overline{X}^nB^n} - \sigma^{X^n\hat{X}^nB^n}\|_1 \leqslant \epsilon. \tag{60}$$

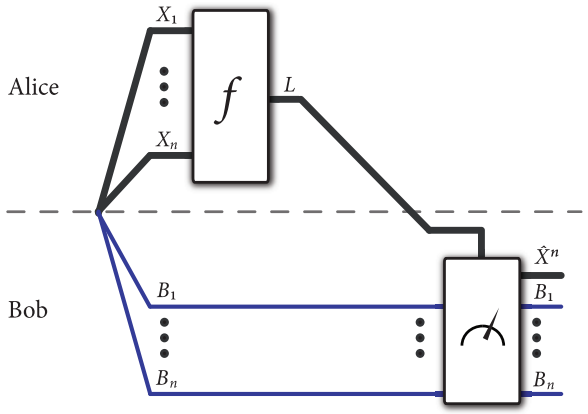

Figure 5. *Classical data compression with QSI protocol*. The protocol begins with the source distributing a random classical sequence to Alice and a correlated quantum state to Bob. Alice begins by hashing the sequence to a variable L with some hash function f. She then transmits the variable L to Bob, using $\log_2 |\mathcal{L}|$ noiseless classical bit channels. Bob receives the hash, and he then enumerates all of the sequences x^n that are consistent with the hash (so that $f(x^n) = l$). He performs a 'quantum scan' over all of the quantum states ρ_{x^n} that are consistent with the received hash. This quantum scan amounts to a sequence of binary quantum measurements, effectively asking, 'Does my quantum state correspond to the first sequence consistent with the hash? To the second? etc.' After receiving the answer 'yes' to one of these questions, he declares the 'yes' sequence to be the one sent from Alice. This strategy has asymptotically vanishing error as long as the size of the hash is at least $nH(X|B)$ bits.

The above specifies an (n, R, ϵ) code for this task, where $R \equiv \log_2 |\mathcal{L}|/n$. Figure 5 depicts an actual implementation of CDC–QSI along the lines discussed in the paragraph.

A rate R is achievable if there exists an (n, R, ϵ) code for all $\epsilon > 0$ and sufficiently large n.

4.2. Classical data compression with QSI theorem

Theorem 11 (Classical data compression with quantum side information)**.** *Suppose that*

$$\sum_x p_X(x)|x\rangle\langle x|^X \otimes \rho_x^B$$

is a classical-quantum state that characterizes a classical-quantum source. Then the conditional von Neumann entropy $H(X|B)$ is the smallest possible achievable rate for classical data compression with quantum side information for this source:

$$\inf\{R \mid R \text{ is achievable}\} = H(X|B).$$

4.3. Achievability proof for CDC with QSI

The resource inequality for this communication task is as follows:

$$\langle \rho^{XB} \rangle + H(X|B)[c \to c] \geqslant \langle \rho^{XX_BB} \rangle,$$

the meaning being that if Alice and Bob share many copies of the state ρ^{XB} and she communicates at a rate $H(X|B)$ to Bob, then they can construct the state ρ^{XX_BB}, so that Bob has a copy of the variable X.

The strategy for achievability is for Alice to hash her sequence X^n to some variable L. Bob then receives the variable L after Alice communicates it to him. He then 'scans' over all of the quantum states ρ_{x^n} that are consistent with the hash and such that the sequence x^n is typical (the strategy essentially disregards the atypical sequences x^n since their total probability mass is asymptotically negligible). He can accomplish this 'scan' by performing a sequential decoding strategy [32, 56], which consists of binary tests of the form, 'Is this state consistent with the hash? Or this one? etc.' He performs these tests until he receives a 'yes' answer in one of his measurements.

The intuition for why $H(X|B)$ should be the ultimate rate of communication is that there are $\approx 2^{nH(X)}$ sequences of the source to account for (the typical ones). From the HSW theorem [35, 55], we know that the maximal number of sequences that Bob can distinguish is $\approx 2^{nI(X;B)}$. Thus, if Alice divides the source sequences into $\approx 2^{nH(X)}/2^{nI(X;B)} = 2^{nH(X|B)}$ groups and sends the label of the group, then Bob should be able to determine which sequence x^n is the one that the source issued.

Detailed strategy. More formally, the encoding strategy is as follows. Alice and Bob are allowed to have an agreed-upon hash function $f : \mathcal{X}^n \to \mathcal{L}$, selected at random from a two-universal family. A hash function f has a collision if two differing sequences $x_1^n, x_2^n \in \mathcal{X}^n$ hash to the same value:

$$x_1^n \neq x_2^n \implies f(x_1^n) = f(x_2^n).$$

A two-universal family has the property that the probability of a collision is the same as that for a uniformly random function (where the probability is with respect to the random choice of the hash function):

$$x_1^n \neq x_2^n \implies \Pr_f \{f(x_1^n) = f(x_2^n)\} \leqslant \frac{1}{|\mathcal{L}|} = 2^{-nR}. \tag{61}$$

Such a strategy is equivalent to the 'random binning' strategy often discussed in information theory texts [15, 30].

Upon receiving the hash value l, Bob performs a sequence of binary measurements $\{\Pi_{x^n}, I - \Pi_{x^n}\}$ for all the sequences x^n that are consistent with the hash value (so that $f(x^n) = l$) and such that they are strongly typical ($x^n \in T_\delta^{X^n}$—see appendix A for details). We define the set $\mathcal{A}(f, l)$ to capture these sequences:

$$\mathcal{A}(f, l) \equiv \{x^n : f(x^n) = l, \ \ x^n \in T_\delta^{X^n}\}. \tag{62}$$

The projector Π_{x^n} is a strong conditionally typical projector (see appendix A), with the property that

$$\mathrm{Tr}\{\Pi_{x^n} \rho_{x^n}\} \geqslant 1 - \epsilon,$$

for all $\epsilon > 0$ and sufficiently large n. From the above property, we would expect these measurements to perform well in identifying the actual state transmitted.

Error analysis. We define the error probability as follows:

$$\Pr\{\text{'error @ decoder'}\} = \sum_{x^n} p_{X^n}(x^n) \ \Pr\{\text{'error @ decoder'} | \ x^n\}.$$

It is then clear that we can focus on the typical sequences x^n because the above error probability is equal to

$$\sum_{x^n \in T_\delta^{X^n}} p_{X^n}(x^n) \ \Pr\{\text{'error @ decoder'} | \ x^n\} + \sum_{x^n \notin T_\delta^{X^n}} p_{X^n}(x^n) \ \Pr\{\text{'error @ decoder'} \mid x^n\}$$

$$\leqslant \sum_{x^n \in T_\delta^{X^n}} p_{X^n}(x^n) \ \Pr\{\text{'error @ decoder'} \mid x^n\} + \epsilon. \tag{63}$$

Now we consider the error term Pr{'error @ decoder' | x^n}. Let $a_1^{(l)}, \ldots, a_{|\mathcal{A}|}^{(l)}$ enumerate all of the sequences in the set $\mathcal{A}(f, l)$ defined in (62) (those sequences consistent with the hash l). Let $a_m^{(l)}$ be the actual sequence x^n that the source issues. Then the probability for a correct decoding for Bob is as follows:

$$\mathrm{Tr}\{\Pi_{a_m^{(l)}} \hat{\Pi}_{a_{m-1}^{(l)}} \cdots \hat{\Pi}_{a_1^{(l)}} \rho_{a_m^{(l)}} \hat{\Pi}_{a_1^{(l)}} \cdots \hat{\Pi}_{a_{m-1}^{(l)}} \Pi_{a_m^{(l)}}\},$$

where $\hat{\Pi}_{x^n} \equiv I - \Pi_{x^n}$, so that the binary tests give a response of 'no' until the test for $a_m^{(l)}$ gives a response of 'yes.' The probability for incorrectly decoding with this strategy is

$$1 - \mathrm{Tr}\{\Pi_{a_m^{(l)}} \hat{\Pi}_{a_{m-1}^{(l)}} \cdots \hat{\Pi}_{a_1^{(l)}} \rho_{a_m^{(l)}} \hat{\Pi}_{a_1^{(l)}} \cdots \hat{\Pi}_{a_{m-1}^{(l)}} \Pi_{a_m^{(l)}}\},$$

so that we can write the error probability in (63) as

$$\sum_{l \in \mathcal{L}} \sum_{a_m^{(l)} \in \mathcal{A}(f,l)} p_{X^n}(a_m^{(l)}) \left[1 - \mathrm{Tr}\{\Pi_{a_m^{(l)}} \hat{\Pi}_{a_{m-1}^{(l)}} \cdots \hat{\Pi}_{a_1^{(l)}} \rho_{a_m^{(l)}} \hat{\Pi}_{a_1^{(l)}} \cdots \hat{\Pi}_{a_{m-1}^{(l)}} \Pi_{a_m^{(l)}}\}\right]. \tag{64}$$

We can rewrite this error probability as

$$\sum_{l \in \mathcal{L}} \sum_{a_m^{(l)} \in \mathcal{A}(f,l)} p_{X^n}(a_m^{(l)}) \, \mathrm{Tr}\{(I - \Theta_{a_m^{(l)}}) \rho_{a_m^{(l)}}\},$$

where we define the POVM element $\Theta_{a_m^{(l)}}$ as

$$\Theta_{a_m^{(l)}} \equiv \hat{\Pi}_{a_1^{(l)}} \cdots \hat{\Pi}_{a_{m-1}^{(l)}} \Pi_{a_m^{(l)}} \hat{\Pi}_{a_{m-1}^{(l)}} \cdots \hat{\Pi}_{a_1^{(l)}}. \tag{65}$$

Using the facts that (see appendix A)

$$1 = \mathrm{Tr}\{\rho_{a_m^{(l)}}\} = \mathrm{Tr}\{\Pi \, \rho_{a_m^{(l)}}\} + \mathrm{Tr}\{(I - \Pi)\rho_{a_m^{(l)}}\} \leqslant \mathrm{Tr}\{\Pi \rho_{a_m^{(l)}}\} + \epsilon,$$

where Π is the typical projector for the average state $\sum_x p_X(x)\rho_x$, and

$$\begin{aligned}
&\mathrm{Tr}\{\Pi_{a_m^{(l)}} \hat{\Pi}_{a_{m-1}^{(l)}} \cdots \hat{\Pi}_{a_1^{(l)}} \rho_{a_m^{(l)}} \hat{\Pi}_{a_1^{(l)}} \cdots \hat{\Pi}_{a_{m-1}^{(l)}} \Pi_{a_m^{(l)}}\} \\
&\quad = \mathrm{Tr}\{\hat{\Pi}_{a_1^{(l)}} \cdots \hat{\Pi}_{a_{m-1}^{(l)}} \Pi_{a_m^{(l)}} \hat{\Pi}_{a_{m-1}^{(l)}} \cdots \hat{\Pi}_{a_1^{(l)}} \rho_{a_m^{(l)}}\} \\
&\quad \geqslant \mathrm{Tr}\{\hat{\Pi}_{a_1^{(l)}} \cdots \hat{\Pi}_{a_{m-1}^{(l)}} \Pi_{a_m^{(l)}} \hat{\Pi}_{a_{m-1}^{(l)}} \cdots \hat{\Pi}_{a_1^{(l)}} \Pi \rho_{a_m^{(l)}} \Pi\} - \|\rho_{a_m^{(l)}} - \Pi \rho_{a_m^{(l)}} \Pi\|_1 \\
&\quad \geqslant \mathrm{Tr}\{\hat{\Pi}_{a_1^{(l)}} \cdots \hat{\Pi}_{a_{m-1}^{(l)}} \Pi_{a_m^{(l)}} \hat{\Pi}_{a_{m-1}^{(l)}} \cdots \hat{\Pi}_{a_1^{(l)}} \Pi \rho_{a_m^{(l)}} \Pi\} - 2\sqrt{\epsilon},
\end{aligned}$$

we can bound the expression in (64) from above by

$$\sum_{l \in \mathcal{L}} \sum_{a_m^{(l)} \in \mathcal{A}(f,l)} p_{X^n}(a_m^{(l)}) \left[\mathrm{Tr}\{\Pi \rho_{a_m^{(l)}} \Pi\} - \mathrm{Tr}\{\Pi_{a_m^{(l)}} \hat{\Pi}_{a_{m-1}^{(l)}} \cdots \hat{\Pi}_{a_1^{(l)}} \Pi \rho_{a_m^{(l)}} \Pi \hat{\Pi}_{a_1^{(l)}} \cdots \hat{\Pi}_{a_{m-1}^{(l)}} \Pi_{a_m^{(l)}}\}\right], \tag{66}$$

(with the other terms $\epsilon + 2\sqrt{\epsilon}$ omitted for simplicity). We now apply Sen's non-commutative union bound [56] (lemma 18 in appendix B) along with concavity of square-root to obtain the following upper bound:

$$2\sqrt{\sum_{l \in \mathcal{L}} \sum_{a_m^{(l)} \in \mathcal{A}(f,l)} p_{X^n}(a_m^{(l)}) \left[\mathrm{Tr}\{(I - \Pi_{a_m^{(l)}}) \Pi \rho_{a_m^{(l)}} \Pi\} + \sum_{i=1}^{m-1} \mathrm{Tr}\{\Pi_{a_i^{(l)}} \Pi \rho_{a_m^{(l)}} \Pi\}\right]}.$$

For the first term in the square-root, we have that

$$\begin{aligned}
\mathrm{Tr}\{(I - \Pi_{a_m^{(l)}}) \Pi \rho_{a_m^{(l)}} \Pi\} &\leqslant \mathrm{Tr}\{(I - \Pi_{a_m^{(l)}}) \rho_{a_m^{(l)}}\} + \|\rho_{a_m^{(l)}} - \Pi \rho_{a_m^{(l)}} \Pi\|_1 \\
&\leqslant \epsilon + 2\sqrt{\epsilon}.
\end{aligned} \tag{67}$$

For the second term in the square-root, we have

$$\begin{aligned}
&\sum_{l\in\mathcal{L}}\sum_{a_m^{(l)}\in\mathcal{A}(f,l)} p_{X^n}\left(a_m^{(l)}\right)\sum_{i=1}^{m-1}\operatorname{Tr}\left\{\Pi_{a_i^{(l)}}\Pi\rho_{a_m^{(l)}}\Pi\right\}\\
&\qquad\leqslant \sum_{l\in\mathcal{L}}\sum_{a_m^{(l)}\in\mathcal{A}(f,l)} p_{X^n}\left(a_m^{(l)}\right)\sum_{a_i^{(l)}\in\mathcal{A}(f,l)\,:\,i\neq m}\operatorname{Tr}\left\{\Pi_{a_i^{(l)}}\Pi\rho_{a_m^{(l)}}\Pi\right\}\\
&\qquad= \sum_{x^n\in T_\delta^{X^n}} p_{X^n}(x^n)\sum_{x'^n\in T_\delta^{X^n}\,:\,x'^n\neq x^n}\mathcal{I}(f(x'^n)=f(x^n))\operatorname{Tr}\left\{\Pi_{x'^n}\Pi\rho_{x^n}\Pi\right\}\\
&\qquad\leqslant \sum_{x^n} p_{X^n}(x^n)\sum_{x'^n\in T_\delta^{X^n}\,:\,x'^n\neq x^n}\mathcal{I}(f(x'^n)=f(x^n))\operatorname{Tr}\left\{\Pi_{x'^n}\Pi\rho_{x^n}\Pi\right\}.
\end{aligned}$$

The first inequality follows by summing over all the indices not equal to m. The equality follows by introducing the indicator function $\mathcal{I}\left(f\left(x'^n\right)=f\left(x^n\right)\right)$, and the last inequality follows by summing over all sequences x^n.

We now analyze the expectation of the error probability, with respect to the random hash function f. (We can imagine that this expectation was there from the beginning of the analysis, and apply concavity of square-root to bring it over this second term.) This leads to

$$\begin{aligned}
&\mathbb{E}_f\left\{\sum_{x^n} p_{X^n}(x^n)\sum_{x'^n\in T_\delta^{X^n}\,:\,x'^n\neq x^n}\mathcal{I}(f(x'^n)=f(x^n))\operatorname{Tr}\{\Pi_{x^n}\,\Pi\,\rho_{x^n}\,\Pi\}\right\}\\
&\qquad= \sum_{x^n} p_{X^n}(x^n)\sum_{x'^n\in T_\delta^{X^n}\,:\,x'^n\neq x^n}\mathbb{E}_f\{\mathcal{I}(f(x'^n)=f(x^n))\}\operatorname{Tr}\{\Pi_{x^n}\,\Pi\,\rho_{x^n}\,\Pi\}\\
&\qquad= \sum_{x^n} p_{X^n}(x^n)\sum_{x'^n\in T_\delta^{X^n}\,:\,x'^n\neq x^n}\Pr_f\{f(x'^n)=f(x^n)\}\operatorname{Tr}\{\Pi_{x^n}\,\Pi\,\rho_{x^n}\,\Pi\}\\
&\qquad\leqslant 2^{-nR}\sum_{x^n} p_{X^n}(x^n)\sum_{x'^n\in T_\delta^{X^n}}\operatorname{Tr}\{\Pi_{x'^n}\,\Pi\,\rho_{x^n}\,\Pi\},
\end{aligned}$$

where the inequality follows from the two-universal property in (61), the fact that $R=\log_2|\mathcal{L}|/n$, and by summing over all sequences x'^n in the typical set. Continuing, we have

$$\begin{aligned}
&= 2^{-nR}\sum_{x'^n\in T_\delta^{X^n}}\operatorname{Tr}\left\{\Pi_{x'^n}\,\Pi\,\left(\sum_{x^n} p_{X^n}(x^n)\rho_{x^n}\right)\,\Pi\right\}\\
&= 2^{-nR}\sum_{x'^n\in T_\delta^{X^n}}\operatorname{Tr}\{\Pi_{x'^n}\,\Pi\,\rho^{\otimes n}\,\Pi\}\\
&\leqslant 2^{-nR}\,2^{-n[H(B)-\delta]}\sum_{x'^n\in T_\delta^{X^n}}\operatorname{Tr}\{\Pi_{x'^n}\,\Pi\}\\
&\leqslant 2^{-nR}\,2^{-n[H(B)-\delta]}\sum_{x'^n\in T_\delta^{X^n}}\operatorname{Tr}\{\Pi_{x'^n}\}\\
&\leqslant 2^{-nR}\,2^{-n[H(B)-\delta]}\,2^{n[H(B|X)+\delta]}\,2^{n[H(X)+\delta]}\\
&= 2^{-n[R-H(X|B)-3\delta]}.
\end{aligned}$$

The first inequality follows from the operator inequality $\Pi\,\rho^{\otimes n}\,\Pi\leqslant 2^{-n[H(B)-\delta]}\,\Pi$, and the second from $\Pi\leqslant I$. The final inequality follows from the bounds $\operatorname{Tr}\{\Pi_{x'^n}\}\leqslant 2^{n[H(B|X)+\delta]}$ and $|T_\delta^{X^n}|\leqslant 2^{n[H(X)+\delta]}$. Collecting everything, the overall error probability is bounded by

$$\epsilon'\equiv\epsilon+2\sqrt{\epsilon}+2\sqrt{\epsilon+2\sqrt{\epsilon}+2^{-n[R-H(X|B)-3\delta]}}. \tag{68}$$

Since the expectation of the above error probability is small (where the expectation is with respect to the random choice of hash function), there exists some particular hash function from the family such that the above inequality is true. Thus, as long as $R = H(X|B) + 4\delta$, we can guarantee that the error probability of the scheme is arbitrarily small.

We now argue that it is possible to make the state after the decoding be arbitrarily close to the initial state, so that the condition in (60) holds. After recovering the sequence $x^n = a_m^{(l)}$ issued by the source, Bob can place it in a classical register, and the post-measurement state from sequential decoding has the following form:

$$\frac{1}{\operatorname{Tr}\{\Theta_{a_m^{(l)}}\rho_{a_m^{(l)}}\}}\Pi_{a_m^{(l)}}\hat{\Pi}_{a_{m-1}^{(l)}}\cdots\hat{\Pi}_{a_1^{(l)}}\rho_{a_m^{(l)}}\hat{\Pi}_{a_1^{(l)}}\cdots\hat{\Pi}_{a_{m-1}^{(l)}}\Pi_{a_m^{(l)}},$$

with $\Theta_{a_m^{(l)}}$ defined in (65) and assuming a correct decoding. The operators $\Pi_{a_m^{(l)}}\hat{\Pi}_{a_{m-1}^{(l)}}\cdots\hat{\Pi}_{a_1^{(l)}}$ and $\sqrt{\Theta_{a_m^{(l)}}}$ are related by a left polar decomposition:

$$\Pi_{a_m^{(l)}}\hat{\Pi}_{a_{m-1}^{(l)}}\cdots\hat{\Pi}_{a_1^{(l)}} = U_{a_m^{(l)}}\sqrt{\Theta_{a_m^{(l)}}},$$

for some unitary $U_{a_m^{(l)}}$. So after Bob recovers the sequence $a_m^{(l)}$, he applies the unitary $U_{a_m^{(l)}}^\dagger$, and the state becomes as follows:

$$\frac{1}{\operatorname{Tr}\{\Theta_{a_m^{(l)}}\rho_{a_m^{(l)}}\}}\sqrt{\Theta_{a_m^{(l)}}}\,\rho_{a_m^{(l)}}\sqrt{\Theta_{a_m^{(l)}}}.$$

We can now show that the condition in (60) holds for this decoding procedure (including the unitaries $U_{a_m^{(l)}}^\dagger$). Consider that for all typical sequences $x^n \in T_\delta^{X^n}$, the trace distance between the initial state and the post-measurement state has the following bound:

$$\|\rho_{x^n} - \operatorname{Tr}\{\Theta_{x^n}\rho_{x^n}\}^{-1}\sqrt{\Theta_{x^n}}\,\rho_{x^n}\,\sqrt{\Theta_{x^n}}\|_1 \leqslant 2\sqrt{\operatorname{Tr}\{(I-\Theta_{x^n})\rho_{x^n}\}},$$

which follows from the Gentle Measurement lemma (lemma 9.4.1 of [60]). Combining this bound with the fact that there is no measurement when $x^n \notin T_\delta^{X^n}$ and defining $\Theta_{x^n} = I$ in this case, averaging over $p_{X^n}(x^n)$, and then applying our bound in (68) and concavity of square-root, we obtain the following upper bound:

$$\sum_{x^n} p_{X^n}(x^n)\|\rho_{x^n} - \operatorname{Tr}\{\Theta_{x^n}\rho_{x^n}\}^{-1}\sqrt{\Theta_{x^n}}\,\rho_{x^n}\,\sqrt{\Theta_{x^n}}\|_1 \leqslant 2\sqrt{\epsilon'}.$$

It then follows that the condition in (60) is satisfied for this protocol.

4.4. Converse theorem for CDC with QSI

This section provides a simple proof of the converse theorem for CDC–QSI. The converse demonstrates that the single-letter rate in theorem 11 is optimal. An inspection of the proof reveals a close similarity with the Slepian–Wolf converse in [30].

In the most general protocol for this task, Alice receives the sequence X^n from the source. She then hashes it to a random variable L where $f(X^n) = L$ and sends it over to Bob via some noiseless classical bit channels. Bob receives L, processes it and B^n to obtain an estimate $\hat{X}^n$ of X^n. If the protocol is any good for this task, then the actual state $\omega^{X^n\hat{X}^nB^n}$ at the end should be ϵ-close in trace distance to the ideal state $\sigma^{X^n\overline{X}^nB^n}$, where $\overline{X}^n$ is a copy of the variable X^n:

$$\|\omega^{X^n\hat{X}^nB^n} - \sigma^{X^n\overline{X}^nB^n}\|_1 \leqslant \epsilon. \tag{69}$$

A proof of the converse goes as follows:

$$\begin{aligned} nR &\geqslant H(L) \\ &\geqslant H(L|B^n) \\ &= I(X^n; L|B^n) + H(L|B^n X^n) \\ &\geqslant I(X^n; L|B^n) \\ &= H(X^n|B^n) - H(X^n|LB^n) \\ &\geqslant H(X^n|B^n)_\omega - H(X^n|\hat{X}^n)_\omega \\ &\geqslant H(X^n|B^n)_\sigma - n\epsilon' \\ &= nH(X|B) - n\epsilon'. \end{aligned}$$

The first two inequalities follow for reasons similar to those in the previous converse in section 2.4. The first equality is an entropy identity, and the third inequality follows because $H(L|B^n X^n) \geqslant 0$ for a classical variable L. The second equality is another entropy identity. The fourth inequality follows from quantum data processing of L and B^n to obtain the estimate $\hat{X}^n$. The final inequality follows from the condition in (69), continuity of entropy (Alicki–Fannes' inequality [3]), and the fact that $H(X^n|\overline{X}^n)_\sigma = 0$ since $\overline{X}^n$ is copy of X^n, with ϵ' being some function $g(\epsilon)$ such that $\lim_{\epsilon\to 0} g(\epsilon) = 0$. The final equality follows because the entropy is additive on a tensor-power state.

5. Measurement compression with quantum side information

We now discuss another new protocol: measurement compression with quantum side information (MC–QSI). The information processing task for this protocol is similar to that in measurement compression (section 2), with the exception that they are to perform the protocol on the A system of some bipartite state, and Bob is allowed to use his system B in order to reduce the communication resources needed for the simulation. The protocol discussed in this section is a 'feedback' simulation, in which the sender also obtains the outcome of the measurement simulation. After reviewing the information processing task, we state the MC–QSI with feedback theorem and prove achievability of the protocol and its converse. Finally, we discuss several applications of it.

5.1. Information processing task for MC–QSI

The information processing task in this case is a straightforward extension of that for measurement compression with feedback. As such, we leave the discussion of it to the captions of figures 6 and 7. One point to observe from the figures is that in the ideal implementation of the measurement, the side information in system B is left untouched. As a result, the measurement compression protocol will be permitted to use system B, but only in ways that do not significantly disturb it.

5.2. Measurement compression with quantum side information theorem

Theorem 12 (Measurement compression with QSI)**.** *Let ρ^{AB} be a source state shared between a sender A and a receiver B, and let Λ be a POVM to simulate on the A system of this state. A protocol for faithful feedback simulation of the POVM with classical communication rate R and common randomness rate S exists if and only if the following inequalities hold:*

$$\begin{aligned} R &\geqslant I(X; R|B), \\ R + S &\geqslant H(X|B), \end{aligned}$$

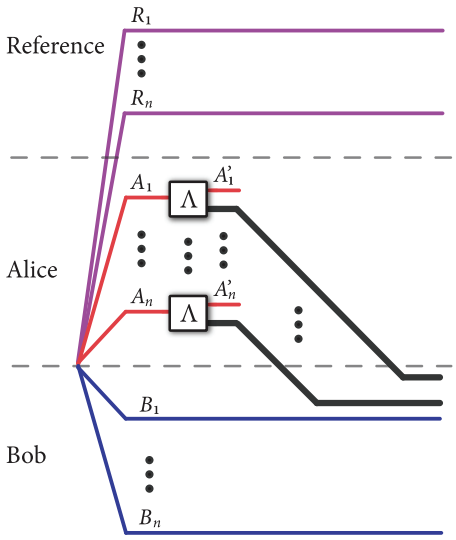

Figure 6. *Ideal measurement compression with quantum side information*. The ideal protocol to which we should compare performance of any actual protocol. The sender and receiver share many copies of some bipartite state ρ^{AB}. Alice performs the measurement $\Lambda \equiv \{\Lambda_x\}$ locally on each of her shares and sends the outcomes to Bob. A simulation of this measurement would have the sender and receiver operate according to some procedure that is statistically indistinguishable from this ideal case.

where the entropies are with respect to a state of the following form:

$$\sum_x |x\rangle\langle x|^X \otimes Tr_A\left\{\left(I^R \otimes \Lambda_x^A\right)\phi^{RAB}\right\},$$

and ϕ^{RAB} is some purification of the state ρ^{AB}.

The achievable rate region closely resembles figure 3, except that all of the information quantities should be conditioned on the system B since, in the new task, B is available as quantum side information.

It is instructive to see how the second example of section 2.2.1 changes if quantum side information is available. Suppose that Bob now possesses the purification of the maximally mixed state, so that Alice and Bob share a Bell state before communication begins. This means that there is no purification system R because the state on A and B is already pure. In this case, the state on X and B after the measurement in (15) is as follows:

$$\tfrac{1}{4}(|0\rangle\langle 0|^X \otimes |0\rangle\langle 0|^B + |1\rangle\langle 1|^X \otimes |1\rangle\langle 1|^B + |2\rangle\langle 2|^X \otimes |+\rangle\langle +|^B + |3\rangle\langle 3|^X \otimes |-\rangle\langle -|^B).$$

The conditional mutual information $I(X;R|B)$ is zero because the reference system is trivial. The conditional entropy $H(X|RB) = H(X|B)$ is equal to one bit. A simple interpretation of this result is that the measurement in (15) just requires one bit of common randomness in order to pick the X or Z Pauli measurement at random. Bob then performs the selected measurement locally, and the effect is the same as if Alice were to perform it on her share of the state because the state is maximally entangled.

5.3. Achievability proof for MC–QSI

The resource inequality corresponding to MC–QSI is as follows:

$$\langle \rho^{AB} \rangle + I(X;R|B)[\,c \to c\,] + H(X|RB)[cc] \geqslant \langle \Lambda^A(\rho^{AB}) \rangle.$$

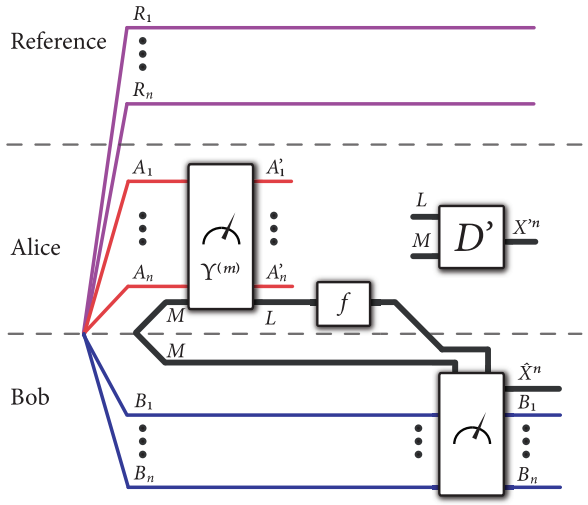

Figure 7. *Measurement compression with quantum side information protocol*. The figure depicts the most general protocol for this task when both the sender and receiver are to obtain the outcome of the measurement simulation. Assuming that Alice and Bob share many copies of some bipartite state ρ^{AB} and have common randomness M available, Alice simulates the measurement $\Lambda^{\otimes n}$ by performing some POVM conditional on the value of the common randomness. Rather than send the full output L of the measurement to Bob, Alice hashes it to $f(L)$ using some hash function f, and she sends the hash $f(L)$ to Bob. Bob performs a measurement on his systems B^n, conditional on the hash $f(L)$ and his share of the common randomness M. From this measurement, he can recover the full value of L and then reconstruct the sequence x^n using L and M. The protocol is also a 'feedback' simulation, such that Alice recovers the outcome of the simulation by processing the classical registers L and M. The protocol performs well if the output of this simulated measurement is statistically indistinguishable from the output of the true measurement $\Lambda^{\otimes n}$ (from the perspective of someone holding the reference systems and the measurement outcomes).

The meaning of this resource inequality is that the sender and receiver can simulate the action of the POVM $\Lambda^{\otimes n}$ on n copies of the state ρ^{AB}, by exploiting $nI(X;R|B)$ bits of classical communication and $nH(X|RB)$ bits of common randomness, and the simulation becomes exact as n becomes large.

One might think that it would be possible to concatenate the protocols of measurement compression and CDC–QSI according to the rules of the resource calculus [26] in order to have a protocol for MC–QSI. The scheme that we develop below certainly does exploit features of both protocols, but a direct concatenation is not possible because Alice and Bob need to exploit the same codebook for both the measurement compression part and the CDC–QSI part of the protocol. We note that this is similar to the way that the protocol for channel simulation with quantum side information operates [44].

The basic strategy for MC–QSI is as follows. Alice simulates the measurement on the A^n systems of the IID state $(\rho^{AB})^{\otimes n}$, with the systems R^nB^n acting as a purification of A^n, by first selecting the variable m according to the common randomness shared with Bob, and then by performing a POVM $\{\Upsilon_l^{(m)}\}$ chosen according to a codebook $\mathcal{C} \equiv \{x^n(l,m)\}$. (The codebook is of the form discussed in section 2.3.) Alice and Bob both know the codebook $\mathcal{C}$ used in the measurement compression strategy. Bob shares the common randomness variable m with Alice, and thus he already has this as side information to help in determining the variable l. Alice hashes the variable l according to some hash function f and sends the hash. Bob receives the hash $k \equiv f(l)$, and then he 'scans' over all of the post-measurement states (corresponding

to codewords $x^n(l', m)$) that are consistent with the hash k and his common randomness value m. We define the set $\mathcal{A}(f, k, m)$ to denote the set of all such codewords:

$$\mathcal{A}(f, k, m) \equiv \{x^n(l, m) : f(l) = k,\ x^n(l, m) \in \mathcal{C}\}. \tag{70}$$

Observe that this set cannot be any larger than $\mathcal{L}$ (the set of all possible l):

$$|\mathcal{A}(f, k, m)| \leqslant |\mathcal{L}| = 2^{n[I(X;RB)+3\delta]}.$$

The intuition behind the protocol is that measurement compression proceeds as before using $nI(X; BR)$ bits for the outcome of the measurement and $nH(X|RB)$ for the common randomness, because the systems RB act as a purification for A. But in this case, Bob has the quantum systems B^n available and should be able to determine $nI(X; B)$ bits about X^n by performing a collective measurement on his systems (following from the HSW theorem [35, 55]). So, Alice should only need to send the difference of these amounts, $n(I(X; BR) - I(X; B)) = nI(X; R|B)$, to Bob.

Detailed Strategy. The encoding strategy for this scenario is as follows. Alice and Bob are allowed to have an agreed-upon hash function $f : \mathcal{L} \to \mathcal{K}$, selected at random from a two-universal family (as described in section 4.3). Alice's message to Bob will be an element of $\mathcal{K}$. The collision probability for some $l \neq l'$ in this case is as follows:

$$\Pr_f\{f(l) = f(l')\} \leqslant \frac{1}{|\mathcal{K}|} = 2^{-nR}.$$

Upon receiving the hash value k and having a particular value m for the common randomness, Bob performs a sequence of binary measurements $\{\Pi_{x^n(l,m)}, I - \Pi_{x^n(l,m)}\}$ for all the codewords $x^n(l, m) \in \mathcal{C}$ that are consistent with the hash value (so that $f(l) = k$). Note that $\Pi_{x^n(l,m)}$ is a conditionally typical projector for the tensor-product state $\rho^{B^n}_{x^n(l,m)}$, the conditional state on Bob's system after performing the ideal measurement $\Lambda^{\otimes n}$ and receiving the outcome $x^n(l, m)$. Recall from (70) that $\mathcal{A}(f, k, m)$ is the set of all such codewords. In the following, we will show that, by choosing

$$|\mathcal{L}| = 2^{n[I(X;RB)+3\delta]}, \tag{71}$$

$$|\mathcal{M}| = 2^{n[H(X|RB)+\delta]}, \tag{72}$$

$$|\mathcal{K}| = 2^{n[I(X;R|B)+11\delta]}, \tag{73}$$

the error probability will approach zero as n goes to infinity.

Error Analysis. The error probability for this decoder is then as follows:

$$\Pr\{\text{'error @ decoder'}\} = \frac{1}{|\mathcal{M}|}\sum_{l,m} q(x^n(l, m))\ \Pr\{\text{'error @ decoder'} \mid l, m\}, \tag{74}$$

where

$$q(x^n(l, m)) = \mathrm{Tr}\left\{\left(\left(\Upsilon^{(m)}_l\right)^{A^n} \otimes I^{B^n}\right)(\rho^{AB})^{\otimes n}\right\}$$

is the probability of receiving outcome l when performing the simulated measurement in (33). The post-measurement states on B^n for the POVM $\{\Upsilon^{(m)}_l\}$ are as follows:

$$\widetilde{\rho}^{B^n}_{x^n(l,m)} \equiv \frac{1}{q(x^n(l, m))}\mathrm{Tr}_{A^n}\left\{\left(\Upsilon^{(m)}_l\right)^{A^n}(\rho^{AB})^{\otimes n}\right\}.$$

Note that the probability masses $q(x^n(l, m))$ and $q(x^n(l', m'))$ and the states $\widetilde{\rho}^{B^n}_{x^n(l,m)}$ and $\widetilde{\rho}^{B^n}_{x^n(l',m')}$ are equivalent, respectively, if two different codewords have the same value (i.e. if $l \neq l'$ or

$m \neq m'$ but $x^n(l,m) = x^n(l',m')$, then $q(x^n(l,m)) = q(x^n(l',m'))$ and $\widetilde{\rho}^{B^n}_{x^n(l,m)} = \widetilde{\rho}^{B^n}_{x^n(l',m')}$—this is due to the way that we choose the measurement operators $\Upsilon^{(m)}_l$ in (33) for the measurement simulation).

Now we consider the error term $\Pr\{\text{'error@decoder'} \mid l, m\}$. Let $a^{(km)}_1, \ldots, a^{(km)}_{|\mathcal{A}|}$ enumerate all of the codewords $x^n(l,m)$ in the set $\mathcal{A}(f,k,m)$ defined in (70) (those codewords $x^n(l,m)$ consistent with the hash k). Let $a^{(km)}_j$ denote the actual codeword $x^n(l,m)$ produced by the simulated measurement. The probability for a correct decoding for Bob is as follows:

$$\operatorname{Tr}\left\{\Pi_{a^{(km)}_j}\ \hat{\Pi}_{a^{(km)}_{j-1}} \cdots \hat{\Pi}_{a^{(km)}_1}\ \widetilde{\rho}^{B^n}_{a^{(km)}_j}\ \hat{\Pi}_{a^{(km)}_1} \cdots \hat{\Pi}_{a^{(km)}_{j-1}}\ \Pi_{a^{(km)}_j}\right\},$$

so that the binary tests give a response of 'no' until the test for $a^{(km)}_j$ gives a response of 'yes.' Then the probability for incorrectly decoding is

$$1 - \operatorname{Tr}\left\{\Pi_{a^{(km)}_j}\ \hat{\Pi}_{a^{(km)}_{j-1}} \cdots \hat{\Pi}_{a^{(km)}_1}\ \widetilde{\rho}^{B^n}_{a^{(km)}_j}\ \hat{\Pi}_{a^{(km)}_1} \cdots \hat{\Pi}_{a^{(km)}_{j-1}}\ \Pi_{a^{(km)}_j}\right\},$$

so that we can write the error probability in (74) as follows (for this decoding strategy):

$$\sum_{\substack{m\in\mathcal{M},\\ k\in\mathcal{K}}} \sum_{a^{(km)}_j \in \mathcal{A}(k,f,m)} \frac{1}{|\mathcal{M}|} q\left(a^{(km)}_j\right) \left[1 - \operatorname{Tr}\left\{\Pi_{a^{(km)}_j}\ \hat{\Pi}_{a^{(km)}_{j-1}} \cdots \hat{\Pi}_{a^{(km)}_1}\ \widetilde{\rho}^{B^n}_{a^{(km)}_j}\ \hat{\Pi}_{a^{(km)}_1} \cdots \hat{\Pi}_{a^{(km)}_{j-1}}\ \Pi_{a^{(km)}_j}\right\}\right]. \tag{75}$$

Observe that the above error probability is equal to

$$\frac{1}{|\mathcal{M}|} \sum_{l,m} q(x^n(l,m)) \operatorname{Tr}\left\{(I - \Theta_{x^n(l,m)})\ \widetilde{\rho}^{B^n}_{x^n(l,m)}\right\}, \tag{76}$$

if we define the POVM element $\Theta_{x^n(l,m)}$ as

$$\Theta_{x^n(l,m)} \equiv \hat{\Pi}_{a^{(km)}_1} \cdots \hat{\Pi}_{a^{(km)}_{j-1}}\ \Pi_{a^{(km)}_j}\ \hat{\Pi}_{a^{(km)}_{j-1}} \cdots \hat{\Pi}_{a^{(km)}_1},$$

where we recall that $a^{(km)}_j = x^n(l,m)$.

Now, we can further express the error probability in (76) as follows, by employing an indicator function:

$$\begin{aligned} &= \sum_{x^n \in \mathcal{X}^n} \frac{1}{|\mathcal{M}|} \sum_{l,m} q(x^n(l,m))\, \mathcal{I}(x^n = x^n(l,m)) \operatorname{Tr}\left\{(I - \Theta_{x^n(l,m)})\ \widetilde{\rho}^{B^n}_{x^n(l,m)}\right\} \\ &\leqslant \sum_{x^n \in \mathcal{X}^n} \frac{1}{|\mathcal{M}|} \sum_{l,m} q(x^n(l,m))\, \mathcal{I}(x^n = x^n(l,m)) \operatorname{Tr}\left\{\left(I - \Theta'_{x^n}\right)\ \widetilde{\rho}^{B^n}_{x^n(l,m)}\right\}, \end{aligned} \tag{77}$$

where we define Θ'_{x^n} as a POVM element corresponding to a worst-case decoding over the states $\widetilde{\rho}^{B^n}_{x^n(l,m)}$ with the same codeword value x^n:

$$\Theta'_{x^n} \equiv \arg \max_{\substack{\Theta_{x^n(l,m)}:\\ x^n = x^n(l,m)}} \operatorname{Tr}\left\{(I - \Theta_{x^n(l,m)})\ \widetilde{\rho}^{B^n}_{x^n(l,m)}\right\}.$$

Let $\mathcal{C}'$ be a pruning of the original codebook $\mathcal{C}$ containing no duplicate entries (it contains only the codewords with worst-case error probabilities as given above). Then the last line in (77) is equivalent to the following one:

$$\sum_{x^n \in \mathcal{C}'} \operatorname{Tr}\left\{\left(I - \Theta'_{x^n}\right) \frac{1}{|\mathcal{M}|} \sum_{l,m} q(x^n(l,m))\, \mathcal{I}(x^n = x^n(l,m))\ \widetilde{\rho}^{B^n}_{x^n(l,m)}\right\}. \tag{78}$$

This decoding scheme will only work well if the states $\widetilde{\rho}^{B^n}_{x^n(l,m)}$ are close to the tensor product states $\rho^{B^n}_{x^n(l,m)}$ that would result from the ideal measurement. We expect that this should hold if the measurement compression part of the protocol is successful, and we prove this in detail in what follows.

The quantity characterizing a faithful measurement simulation in (11) is equivalent to the following (one can show this by exploiting the definitions of the measurement maps and the post-measurement states given above, after tracing out the reference systems):

$$\begin{aligned}\Delta(\mathcal{C}) &\equiv \sum_{x^n\in\mathcal{X}^n}\left\| p_{X^n}(x^n)\rho_{x^n}^{B^n} - \frac{1}{|\mathcal{M}|}\sum_{l,m} q(x^n(l,m))\,\mathcal{I}(x^n = x^n(l,m))\,\widetilde{\rho}^{B^n}_{x^n(l,m)}\right\|_1 \\ &= \sum_{x^n\notin\mathcal{C}'}\left\| p_{X^n}(x^n)\rho_{x^n}^{B^n}\right\|_1 + \sum_{x^n\in\mathcal{C}'}\left\| p_{X^n}(x^n)\rho_{x^n}^{B^n}\right. \\ &\qquad \left. - \frac{1}{|\mathcal{M}|}\sum_{l,m} q(x^n(l,m))\,\mathcal{I}(x^n = x^n(l,m))\,\widetilde{\rho}^{B^n}_{x^n(l,m)}\right\|_1,\end{aligned}$$

where $\mathcal{C}'$ is the pruned codebook containing no duplicate entries (observe that the indicator function $\mathcal{I}(x^n = x^n(l,m))$ captures all of the duplicates). Applying the trace inequality $\mathrm{Tr}\{\Lambda\sigma\} \leqslant \mathrm{Tr}\{\Lambda\rho\} + \|\rho-\sigma\|_1$ from lemma 17 to the expression in (78), we then obtain the following upper bound on it:

$$\begin{aligned}&\leqslant \sum_{x^n\in\mathcal{C}'}\mathrm{Tr}\left\{\left(I-\Theta'_{x^n}\right)p_{X^n}(x^n)\rho_{x^n}^{B^n}\right\} \\ &\qquad + \sum_{x^n\in\mathcal{C}'}\left\| p_{X^n}(x^n)\rho_{x^n}^{B^n} - \frac{1}{|\mathcal{M}|}\sum_{l,m} q(x^n(l,m))\,\mathcal{I}(x^n = x^n(l,m))\,\widetilde{\rho}^{B^n}_{x^n(l,m)}\right\|_1 \\ &\leqslant \sum_{x^n\in\mathcal{C}'} p_{X^n}(x^n)\mathrm{Tr}\left\{\left(I-\Theta'_{x^n}\right)\rho_{x^n}^{B^n}\right\} + \Delta(\mathcal{C}).\end{aligned}$$

We can now focus on bounding the term on the LHS of the last line above. Expanding it again leads to

$$\begin{aligned}\sum_{x^n\in\mathcal{C}'} p_{X^n}(x^n)\mathrm{Tr}\left\{\left(I-\Theta'_{x^n}\right)\rho_{x^n}^{B^n}\right\} &= \sum_{\substack{m\in\mathcal{M},\\ k\in\mathcal{K}}}\sum_{a_j^{(km)}\in\mathcal{A}'(k,f,m)} p_{X^n}\left(a_j^{(km)}\right)\mathrm{Tr}\left\{\left(I-\Theta'_{a_j^{(km)}}\right)\rho_{a_j^{(km)}}^{B^n}\right\} \\ &= \sum_{\substack{m\in\mathcal{M},\\ k\in\mathcal{K}}}\sum_{a_j^{(km)}\in\mathcal{A}'(k,f,m)} p_{X^n}\left(a_j^{(km)}\right)\Big[1 - \mathrm{Tr}\Big\{\Pi_{a_j^{(km)}} \\ &\qquad \times \hat{\Pi}_{a_{j-1}^{(km)}}\cdots\hat{\Pi}_{a_1^{(km)}}\,\rho^{B^n}_{a_j^{(km)}}\,\hat{\Pi}_{a_1^{(km)}}\cdots\hat{\Pi}_{a_{j-1}^{(km)}}\,\Pi_{a_j^{(km)}}\Big\}\Big],\end{aligned}$$

where

$$\mathcal{A}'(f,k,m) \equiv \{x^n(l,m) : f(l) = k,\ \ x^n(l,m)\in\mathcal{C}'\}.$$

We can then insert the average state typical projector as we did before in (66), in order to bound the last line from above as

$$\begin{aligned}&\sum_{\substack{m\in\mathcal{M},\\ k\in\mathcal{K}}}\sum_{a_j^{(km)}\in\mathcal{A}'(k,f,m)} p_{X^n}\left(a_j^{(km)}\right) \\ &\qquad \times\left[\mathrm{Tr}\left\{\Pi\rho^{B^n}_{a_j^{(km)}}\Pi\right\} - \mathrm{Tr}\left\{\Pi_{a_j^{(km)}}\hat{\Pi}_{a_{j-1}^{(km)}}\cdots\hat{\Pi}_{a_1^{(km)}}\Pi\rho^{B^n}_{a_j^{(km)}}\Pi\hat{\Pi}_{a_1^{(km)}}\cdots\hat{\Pi}_{a_{j-1}^{(km)}}\Pi_{a_j^{(km)}}\right\}\right].\end{aligned}$$

The error accumulated in doing so is $\epsilon + 2\sqrt{\epsilon}$ as before. At this point, we can apply Sen's non-commutative union bound (lemma 18 in appendix B) and concavity of square root to obtain the upper bound

$$2\sqrt{\sum_{\substack{m\in\mathcal{M},\\ k\in\mathcal{K}}}\sum_{a_j^{(km)}\in\mathcal{A}'(k,f,m)} p_{X^n}\left(a_j^{(km)}\right)\left[\mathrm{Tr}\left\{\left(I-\Pi_{a_j^{(km)}}\right)\Pi\,\rho^{B^n}_{a_j^{(km)}}\,\Pi\right\} + \sum_{i=1}^{j-1}\mathrm{Tr}\left\{\Pi_{a_i^{(km)}}\,\Pi\,\rho^{B^n}_{a_j^{(km)}}\,\Pi\right\}\right]}.$$

The first term inside the square root we can bound from above by $\epsilon + 2\sqrt{\epsilon}$ as we did before in (67), using properties of quantum typicality. We continue bounding the second term as follows:

$$\begin{aligned}
&\sum_{\substack{m\in\mathcal{M},\\ k\in\mathcal{K}}} \sum_{a_j^{(km)}\in\mathcal{A}'(k,f,m)} p_{X^n}\big(a_j^{(km)}\big) \sum_{i=1}^{j-1} \mathrm{Tr}\big\{\Pi_{a_i^{(km)}}\ \Pi\ \rho^{B^n}_{a_j^{(km)}}\ \Pi\big\} \\
&\qquad \leqslant \sum_{\substack{m\in\mathcal{M},\\ k\in\mathcal{K}}} \sum_{a_j^{(km)}\in\mathcal{A}'(k,f,m)} p_{X^n}(a_j^{(km)}) \sum_{i\neq j\,:\, a_i^{(km)}\in\mathcal{A}(k,f,m)} \mathrm{Tr}\{\Pi_{a_i^{(km)}}\ \Pi\ \rho^{B^n}_{a_j^{(km)}}\ \Pi\} \\
&\qquad = \sum_{l,m\,:\, x^n(l,m)\in\mathcal{C}'} p_{X^n}(x^n(l,m)) \sum_{l'\in\mathcal{L}\,:\, l'\neq l} \mathcal{I}(f(l)=f(l'))\mathrm{Tr}\big\{\Pi_{x^n(l',m)}\ \Pi\ \rho^{B^n}_{x^n(l,m)}\ \Pi\big\},
\end{aligned}$$

where the two steps follow by including all indices in the sum not equal to j and rewriting the sum with indicator functions. We now take an expectation with respect to the random hash (realizing that we could have done this the whole time):

$$\begin{aligned}
&\mathbb{E}_f \left\{ \sum_{l,m\,:\, x^n(l,m)\in\mathcal{C}'} p_{X^n}(x^n(l,m)) \sum_{l'\in\mathcal{L}\,:\, l'\neq l} \mathcal{I}(f(l)=f(l'))\mathrm{Tr}\big\{\Pi_{x^n(l',m)}\ \Pi\ \rho^{B^n}_{x^n(l,m)}\ \Pi\big\} \right\} \\
&\quad = \sum_{l,m\,:\, x^n(l,m)\in\mathcal{C}'} p_{X^n}(x^n(l,m)) \sum_{l'\in\mathcal{L}\,:\, l'\neq l} \mathbb{E}_f\{\mathcal{I}(f(l)=f(l'))\}\mathrm{Tr}\big\{\Pi_{x^n(l',m)}\ \Pi\ \rho^{B^n}_{x^n(l,m)}\ \Pi\big\} \\
&\quad = \sum_{l,m\,:\, x^n(l,m)\in\mathcal{C}'} p_{X^n}(x^n(l,m)) \sum_{l'\in\mathcal{L}\,:\, l'\neq l} \Pr_f\{f(l)=f(l')\}\mathrm{Tr}\big\{\Pi_{x^n(l',m)}\ \Pi\ \rho^{B^n}_{x^n(l,m)}\ \Pi\big\} \\
&\quad \leqslant 2^{-nR} \sum_{l,m\,:\, x^n(l,m)\in\mathcal{C}'} p_{X^n}(x^n(l,m)) \sum_{l'\in\mathcal{L}\,:\, l'\neq l} \mathrm{Tr}\big\{\Pi_{x^n(l',m)}\ \Pi\ \rho^{B^n}_{x^n(l,m)}\ \Pi\big\} \\
&\quad \leqslant 2^{-nR} \sum_{l,m} \sum_{l'\in\mathcal{L}\,:\, l'\neq l} \mathrm{Tr}\big\{\Pi_{x^n(l',m)}\ \Pi\ p_{X^n}(x^n(l,m))\rho^{B^n}_{x^n(l,m)}\ \Pi\big\} \\
&\quad \leqslant 2^{-nR}\, 2^{-n[H(X)-\delta]} \sum_{l,m} \sum_{l'\in\mathcal{L}\,:\, l'\neq l} \mathrm{Tr}\big\{\Pi_{x^n(l',m)}\ \Pi\ \rho^{B^n}_{x^n(l,m)}\ \Pi\big\}. \qquad (79)
\end{aligned}$$

The first inequality follows from the two-universal hashing property. The second inequality follows from summing over all of the codewords in $\mathcal{C}$, not just the non-duplicate entries in $\mathcal{C}'$. The third inequality follows because all of the sequences in $x^n(l,m)$ are chosen to be strongly typical (recall the construction in section 2.3) and by upper bounding their probabilities by $2^{-n[H(X)-\delta]}$. From here, we exploit the fact that the codewords $x^n(l,m)$ were chosen randomly as specified in section 2.3. So we now consider $X^n(l,m)$ as random variables and take the expectation with respect to them (realizing again that we could have done this the whole time and focusing on the rightmost term above):

$$\begin{aligned}
&\mathbb{E}_{X^n} \left\{ \sum_{l,m} \sum_{l'\in\mathcal{L}\,:\, l'\neq l} \mathrm{Tr}\big\{\Pi_{X^n(l',m)}\ \Pi\ \rho^{B^n}_{X^n(l,m)}\ \Pi\big\} \right\} \\
&\qquad = \sum_{l,m} \sum_{l'\in\mathcal{L}\,:\, l'\neq l} \mathrm{Tr}\big\{\mathbb{E}_{X^n}\{\Pi_{X^n(l',m)}\}\ \Pi\ \mathbb{E}_{X^n}\big\{\rho^{B^n}_{X^n(l,m)}\big\}\ \Pi\big\} \\
&\qquad = \sum_{l,m} \sum_{l'\in\mathcal{L}\,:\, l'\neq l} \mathrm{Tr}\left\{\mathbb{E}_{X^n}\{\Pi_{x^n(l',m)}\}\ \Pi\ \sum_{x^n} p_{X'^n}(x^n)\rho^{B^n}_{x^n}\ \Pi\right\} \\
&\qquad \leqslant [1-\epsilon]^{-1} 2^{-n[H(B)-\delta]} \sum_{l,m} \sum_{l'\in\mathcal{L}\,:\, l'\neq l} \mathrm{Tr}\{\mathbb{E}_{X^n}\{\Pi_{x^n(l',m)}\}\Pi\}.
\end{aligned}$$

The first equality follows because the indices l' and l are different, implying that the random variables $X^n(l', m)$ and $X^n(l, m)$ are independent so that we can distribute the expectation. The inequality follows by applying the operator inequality (A.10) from appendix A and properties of quantum typicality. Continuing, we have

$$\begin{aligned}
&\leqslant [1-\epsilon]^{-1}\, 2^{-n[H(B)-\delta]} \sum_{l,m} \sum_{l' \in \mathcal{L}\,:\, l' \neq l} \mathbb{E}_{X^n}\{\mathrm{Tr}\{\Pi_{x^n(l',m)}\}\} \\
&\leqslant [1-\epsilon]^{-1}\, 2^{-n[H(B)-\delta]} \sum_{l,m} \sum_{l' \in \mathcal{L}\,:\, l' \neq l} 2^{n[H(B|X)+\delta]} \\
&\leqslant [1-\epsilon]^{-1}\, 2^{-n[H(B)-\delta]}\, 2^{n[H(B|X)+\delta]}\, |\mathcal{L} \times \mathcal{M}||\mathcal{L}| \\
&\leqslant [1-\epsilon]^{-1}\, 2^{-n[H(B)-\delta]}\, 2^{n[H(B|X)+\delta]}\, 2^{n[H(X)+4\delta]}\, 2^{n[I(X;RB)+3\delta]}.
\end{aligned}$$

The first inequality is from $\Pi \leqslant I$ and the second is from the bound $\mathrm{Tr}\{\Pi_{x^n(l',m)}\} \leqslant 2^{n[H(B|X)+\delta]}$. The final inequality follows from the selection for the sizes of $\mathcal{L}$ and $\mathcal{M}$ in (71)–(72). Combining this bound with the one in (79), our final upper bound is

$$[1-\epsilon]^{-1}\, 2^{-n[R-I(X;R|B)-10\delta]}.$$

Collecting everything together, we arrive at the following upper bound on the decoding error probability in (74):

$$\epsilon''' \equiv \Delta(\mathcal{C}) + \epsilon + 2\sqrt{\epsilon} + 2\sqrt{\epsilon + 2\sqrt{\epsilon} + [1-\epsilon]^{-1}\, 2^{-n[R-I(X;R|B)-10\delta]}}.$$

Thus, as along as we choose $R = I(X; R|B) + 11\delta$, the expectation of this error with respect to the hash function and the random choice of code vanishes in the asymptotic limit.

We now complete our achievability proof by demonstrating that there exists a choice of the $\{X^n(l, m)\}$ codewords such that the measurement simulation error and the decoding error become arbitrarily small. Let F be the event that the decoding error probability is less than $\sqrt{\epsilon'''}$. Then we have the following upper bound on the complement of this event by invoking Markov's inequality:

$$\Pr\{F^c\} \leqslant \frac{\mathbb{E}_{\mathcal{C},f}\{\text{'decoding error'}\}}{\sqrt{\epsilon'''}} \leqslant \sqrt{\epsilon'''}.$$

Thus, by choosing $|\mathcal{L}|$, $|\mathcal{M}|$, and $|\mathcal{K}|$ appropriately, we can have all of the events E_m, E_0, and F be true for some choice of the codebook $\{x^n(l, m)\}$ and the hash f (similar to the 'union bound' argument in (31)), so that both the measurement simulation error and the decoding error probability are arbitrarily small for sufficiently large n.

After determining the sequence x^n resulting from the measurement simulation, Bob can place it in a classical register. By using the fact that the measurement simulation and the decoding are successful and employing an argument similar to that at the end of section 4.3, we know that the disturbance of the state is asymptotically negligible, so that the condition in (80) for a good protocol is satisfied.

A proof similar to that in theorem 8 implies that Alice and Bob can exploit quantum side information and a protocol similar to the above to simulate a general quantum instrument, in such a way that Alice possesses the quantum output and Bob obtains the classical output. The resulting resource inequality is stated in (81) below.

5.4. Converse for measurement compression with QSI

The converse proof for measurement compression with QSI demonstrates the optimality of the protocol from the previous section. Specifically, it shows that the single-letter rates in theorem 12 are optimal for the case of a feedback simulation.

The most general protocol for this task has Alice combine her shares A^n of the state with her share of the common randomness M and perform some quantum operation with quantum outputs A'^n and classical output L. Alice then processes this variable L to produce another random variable L', which she sends to Bob over some noiseless classical bit channels. Bob feeds L', his share of the common randomness, and his systems B^n into some quantum operation with classical outputs $\hat{X}^n$ and quantum outputs B'^n. If the protocol is any good for this task, then the actual state $\omega^{R^n A'^n X'^n \hat{X}^n B'^n}$ should be ϵ-close in trace distance to the ideal state $\sigma^{R^n A'^n X^n \overline{X}^n B^n}$, where $\overline{X}^n$ is a copy of the variable X^n:

$$\|\omega^{R^n A'^n X'^n \hat{X}^n B'^n} - \sigma^{R^n A'^n X^n \overline{X}^n B^n}\|_1 \leqslant \epsilon. \tag{80}$$

A proof for the first bound in theorem 12 goes as follows:

$$\begin{aligned}
nR &\geqslant H(L') \\
&\geqslant I(L'; MB^n R^n) \\
&= I(L'MB^n; R^n) + I(L'; MB^n) - I(R^n; MB^n) \\
&\geqslant I(L'MB^n; R^n) - I(R^n; B^n) \\
&\geqslant I(\hat{X}^n B'^n; R^n)_\omega - I(R^n; B^n)_\sigma \\
&\geqslant I(X^n B^n; R^n)_\sigma - I(R^n; B^n)_\sigma - n\epsilon' \\
&= I(X^n; B^n | R^n)_\sigma - n\epsilon' \\
&= nI(X; B|R) - n\epsilon'.
\end{aligned}$$

The first two inequalities are straightforward (similar to steps in our previous converse proofs). The first equality is an identity for quantum mutual information. The third inequality follows because $I(L'; MB^n) \geqslant 0$ and the common randomness M is uncorrelated with systems R^n and B^n (so that $I(R^n; MB^n) = I(R^n; B^n)$). The fourth inequality follows from quantum data processing of the systems $L'MB^n$ to produce the systems $\hat{X}^n B'^n$. The fifth inequality follows from the condition in (80) and continuity of quantum mutual information (the Alicki–Fannes' inequality [3]), where ϵ' is some function $g(\epsilon)$ such that $\lim_{\epsilon \to 0} g(\epsilon) = 0$. The second equality follows from the chain rule for quantum mutual information: $I(X^n B^n; R^n)_\sigma = I(X^n; R^n | B^n)_\sigma + I(B^n; R^n)_\sigma$. The final equality follows because conditional quantum mutual information is additive on tensor-power states.

The argument justifying the other bound in theorem 12 goes as follows:

$$\begin{aligned}
n(R+S) &\geqslant H(L'M) \\
&\geqslant H(L'M|B^n) \\
&= I(X'^n; L'M|B^n) + H(L'M|B^n X'^n) \\
&\geqslant I(X'^n; L'M|B^n) \\
&= H(X'^n|B^n) - H(X'^n|L'MB^n) \\
&\geqslant H(X'^n|B^n)_\omega - H(X'^n|\hat{X}^n)_\omega \\
&\geqslant H(X^n|B^n)_\sigma - n\epsilon' \\
&= nH(X|B) - n\epsilon'.
\end{aligned}$$

The first two inequalities are straightforward. The first equality is an identity for quantum mutual information. The third inequality follows because the entropy $H(L'M|B^n X'^n) \geqslant 0$ for classical systems L' and M. The second equality is an identity for quantum mutual information. The fourth inequality follows from quantum data processing of the systems $L'MB^n$. The last inequality follows from the condition in (80), continuity of entropy, and the fact that $H(X^n|\overline{X}^n)_\sigma = 0$ since $\overline{X}^n$ is a copy of X^n. The final equality follows because the entropy is additive for tensor-power states.

Optimality of the bound $R+S \geqslant H(X|B)$ for negative S follows by considering a protocol whereby Alice uses classical communication alone in order to simulate X^n and generate common randomness M with Bob. The converse in this case proceeds as follows:

$$\begin{aligned} nR &\geqslant H(L') \\ &\geqslant H(L'|B^n) \\ &\geqslant I(X'^n M; L'|B^n) \\ &= H(X'^n M|B^n) - H(X'M|L'B^n) \\ &\geqslant H(X'^n M|B^n)_\omega - H(X'M|\hat{X}^n M)_\omega \\ &\geqslant H(X^n M|B^n)_\sigma - n\epsilon' \\ &= nH(X|B) + H(M) - n\epsilon' \\ &= nH(X|B) + n|S| - n\epsilon'. \end{aligned}$$

The fourth inequality follows because Bob has to process L' and B^n in order to recover the approximate $\hat{X}^n$ and M. The fifth inequality follows because these systems should be close to the ideal ones for a good protocol (and applying continuity of entropy). The next equalities follow because the information quantities factor as above for the ideal state.

5.5. Relation of MC–QSI to other protocols

We remark on the connection between measurement compression with quantum side information and two other protocols: channel simulation with quantum side information [44] and state redistribution [29, 68]. MC–QSI lies somewhere in between both of these protocols—it generalizes channel simulation with QSI but is not 'fully quantum,' in contrast to state redistribution, which is. Channel simulation with QSI is a protocol whereby a sender and receiver share many copies of a classical-quantum state $\sum_y p_Y(y)|y\rangle\langle y|^Y \otimes \rho_y^B$ distributed to them by a source, with the sender holding the classical systems and the receiver holding the quantum systems. The goal is for the sender and receiver to simulate the action of a noisy classical channel $p_{X|Y}(x|y)$ on the sender's classical systems by using as few noiseless bit channels and common randomness bits as possible. Luo and Devetak found that this is possible by using a classical communication rate of $I(X;Y|B)$ and a common randomness rate of $H(X|YB)$ (compare with $I(X;R|B)$ and $H(X|RB)$ for MC–QSI), where the entropies are with respect to a state of the following form:

$$\sum_{y,x} p_Y(y) p_{X|Y}(x|y)|y\rangle\langle y|^Y \otimes |x\rangle\langle x|^X \otimes \rho_y^B.$$

(They actually found the rates to be $I(Y;X) - I(B;X)$ and $H(X|Y)$, but combining these rates with the fact that $I(X;B|Y) = 0$ for a state of the above form gives the rates we state above.) This protocol exploits aspects of CDC–QSI and the classical reverse Shannon theorem in its proof. It has applications to rate distortion theory with quantum side information and in devising a simpler proof of the distillable common randomness from quantum states [28]. The completely classical version of this protocol has further applications to multi-terminal problems in classical rate distortion theory [43]. Clearly, our protocol generalizes channel simulation with QSI because a classical channel, a classical-to-classical map, is a special case of a quantum measurement, a quantum-to-classical map.

State redistribution is a protocol that generalizes MC–QSI to the setting where one would like to simulate the action of a noisy quantum channel on some bipartite state ρ^{AB}. That is, state redistribution leads to a quantum reverse Shannon theorem in the presence of quantum side information which we call QRST–QSI (the authors of [29, 68] did not emphasize this aspect

of their protocol). Indeed, supposing that the goal is to simulate the action of a channel $\mathcal{N}^{A\to B'}$ on the bipartite state, they could proceed by Alice locally performing the isometric extension $U_{\mathcal{N}}^{A\to B'E}$ of the channel $\mathcal{N}^{A\to B'}$ on the A system of the state ρ^{AB}. Including the reference R as a purification of ρ^{AB}, there are four systems $RB'EB$ after she does so, where Alice possesses B' and E and Bob possesses B. Alice and Bob then operate according to the state redistribution protocol in order for Alice to transfer the B' system to Bob (this effects the channel simulation of $\mathcal{N}^{A\to B'}$ on the state ρ^{AB}). Transferring the state requires some rate Q of noiseless quantum communication and some rate E of noiseless entanglement, and according to the main theorem of [29, 68], this is possible as long as

$$\begin{aligned} Q &\geqslant \tfrac{1}{2} I(B'; R|B), \\ Q + E &\geqslant H(B'|B). \end{aligned}$$

Comparing the above rate region with theorem 12 of this review reveals a close analogy between noiseless quantum communication / entanglement in QRST–QSI and noiseless classical communication / common randomness in MC–QSI, with the factor of 1/2 above accounting for the fact that the communication in QRST–QSI is quantum. Though, one should be aware that this connection is only formally similar—in QRST–QSI, sometimes the protocol can generate entanglement rather than consume it, depending on the channel and the state on which the channel acts (this can never happen in MC–QSI because the entropy $H(X|RB)$ is always positive for a classical X system).

5.6. Applications of MC–QSI

We now discuss three applications of MC–QSI. The first application is one that two of us announced in [38], the second involves developing a quantum reverse Shannon theorem for a quantum instrument, and the third is in reducing the classical communication cost of the local purity distillation protocol outlined in [41].

5.6.1. Classically-assisted state redistribution. For the first application, the setting is that Alice and Bob share many copies of some bipartite state ρ^{AB}, and we would like to know how the resources of classical communication, quantum communication, and entanglement can combine with the state ρ^{AB} for different information processing tasks. Let $|\psi\rangle^{RAB}$ be a purification of ρ^{AB}. We found a general protocol, called 'classically-assisted state redistribution,' that when combined with teleportation, super-dense coding, and entanglement distribution can generate all of the known 'static' protocols in the literature and is furthermore optimal for these tasks according to a multi-letter converse theorem [38]. In the first step of classically-assisted state redistribution, Alice and Bob employ the MC–QSI protocol in order to implement the following resource inequality:

$$\langle \rho^{AB} \rangle + I(X_B; R|B)[\, c \to c\,] + H(X_B|RB)[cc] \geqslant \langle \overline{\Delta}^{X\to X_A X_B} \circ T^{A\to A'XE'} : \rho^{AB} \rangle. \tag{81}$$

In the above, the resource on the RHS is a remote instrument, such that the map $T^{A\to A'XE'}$ is first simulated in such a way that Alice possesses the environment E' of the instrument $T^{A\to A'X}$, followed by a copying of the classical output X to one for Alice (X_A) and one for Bob (X_B) (the notation $\overline{\Delta}^{X\to X_A X_B}$ indicates a classical copying channel). Let $\sigma^{A'X_A X_B E'B}$ denote the post-measurement state. Conditional on the classical variable X, the parties then perform the state redistribution protocol [29, 68], in which Alice redistributes the share A' of the post-measurement state to Bob. The resource inequality for this task is as follows:

$$\langle \sigma^{A'E'X_A|BX_B} \rangle + \tfrac{1}{2} I(A'; R|BX_B)[\, q \to q\,] + \tfrac{1}{2}(I(A'; E'|X_B) - I(A'; B|X_B))[qq] \geqslant \langle \sigma^{E'X_A|A'BX_B} \rangle,$$

where the vertical divider | for the states above indicates who possesses what systems and the information quantities are all conditioned on X since this classical variable is available to both parties. The above resource inequality is equivalent to the following one, after applying the identity $I(A';R|BX_B) = I(A';R|E'X_B)$ [29, 68] and moving the entanglement consumption to the RHS along with a sign inversion (so that it now corresponds to an entanglement generation rate):

$$\langle \sigma^{A'X_AE'|BX_B} \rangle + \tfrac{1}{2} I(A';R|E'X_B)[\,q \to q\,] \geqslant \tfrac{1}{2}(I(A';B|X_B) - I(A';E'|X_B))[qq] + \langle \sigma^{E'X_A|A'BX_B} \rangle.$$

Overall, we then have the following resource inequality

$$\langle \rho^{AB} \rangle + I(X_B;R|B)[\,c \to c\,] + H(X_B|RB)[cc] + \tfrac{1}{2} I(A';R|E'X_B)[\,q \to q\,] \geqslant \tfrac{1}{2}(I(A';B|X_B) - I(A';E'|X_B))[qq],$$

if we are not concerned with the 'state redistribution' aspect of the protocol and merely its abilities for entanglement distillation. Finally, since the goal of the protocol is entanglement distillation and not actually simulating the measurement, we can exploit the common randomness to agree upon a particular protocol in the ensemble of these protocols for the task of entanglement distillation and it is not necessary to have common randomness as a resource (it can be derandomized and this is the content of corollary 4.8 of [26]). The final resource inequality is then

$$\langle \rho^{AB} \rangle + I(X_B;R|B)[c \to c] + \tfrac{1}{2} I(A';R|E'X_B)[q \to q] \geqslant \tfrac{1}{2}(I(A';B|X_B) - I(A';E'|X_B))[qq].$$

Combining the above protocol with teleportation, super-dense coding, and entanglement distribution then gives all of the known protocols on the 'static branch' of quantum information theory.

5.6.2. Quantum reverse Shannon theorem for a quantum instrument. The quantum reverse Shannon theorem in its simplest form makes a statement regarding the ability of noiseless quantum communication and entanglement to simulate the action of some channel $\mathcal{N}^{A\to B'}$ on many copies of a state ρ. A simple extension of the theorem that we discussed in the introduction is QRST–QSI, which simulates the channel on a bipartite state ρ^{AB}. The resource inequality for this protocol is as follows:

$$\tfrac{1}{2} I(R;B'|B)_\omega[q \to q] + \tfrac{1}{2}(I(B';E)_\omega - I(B';B)_\omega)\,[qq] \geqslant \langle U_{\mathcal{N}}^{A\to B'E} : \rho^{AB} \rangle,$$

where the information quantities are with respect to a state ω^{RBE} of the following form:

$$\omega^{RB'EB} \equiv U_{\mathcal{N}}^{A\to B'E} |\psi_\rho\rangle^{RAB}.$$

In the above, $U_{\mathcal{N}}^{A\to B'E}$ is an isometric extension of the channel $\mathcal{N}$ and $|\psi_\rho\rangle^{RAB}$ is a purification of the state ρ^{AB}. The protocol employs state redistribution [29, 68]. In this case, if $\frac{1}{2}(I(B';E)_\omega - I(B';B)_\omega)$ is negative, then the protocol is generating entanglement rather than consuming it. A special case of the above theorem is when there is no quantum side information (when B is trivial), in which case the resource inequality becomes the usual quantum reverse Shannon theorem [24, 1, 5, 10]:

$$\tfrac{1}{2} I(R;B')_\omega[q \to q] + \tfrac{1}{2} I(B';E)_\omega[qq] \geqslant \langle U_{\mathcal{N}}^{A\to B'E} : \rho^{A} \rangle.$$

Suppose that we instead would like to simulate the action of a quantum instrument $\mathcal{N}^{A\to XB'}$ with classical output X and quantum output B' on the bipartite state ρ^{AB}. A quantum instrument is the most general model for quantum measurement that includes both a classical output and a post-measurement quantum state [21]. A quantum instrument always admits the following decomposition

$$\mathcal{N}^{A\to XB'}(\rho) \equiv \sum_x |x\rangle\langle x|^X \otimes \mathcal{N}_x^{A\to B'}(\rho),$$

in terms of the completely positive trace-nonincreasing maps $\mathcal{N}_x^{A\to B'}$, such that the overall quantum map after tracing over the classical system X is a completely positive trace-preserving map:

$$\mathcal{N}^{A\to B'}(\rho) \equiv \sum_x \mathcal{N}_x^{A\to B'}(\rho),$$
$$\mathrm{Tr}\{\mathcal{N}^{A\to B'}(\rho)\} = 1.$$

$\mathcal{N}^{A\to B'}$ has the following isometric extension:

$$U_{\mathcal{N}}^{A\to XX_EB'E} = \sum_x |x\rangle^X |x\rangle^{X_E} U_{\mathcal{N}_x}^{A\to B'E},$$

where $U_{\mathcal{N}_x}^{A\to B'E}$ is an extension of the map $\mathcal{N}_x^{A\to B'}$. Thus, tracing over X_E and E recovers the action of the original instrument.

If we are interested in simulating this channel on the state ρ^{AB}, we could straightforwardly apply the quantum reverse Shannon theorem to show that the following resource inequality exists

$$\tfrac{1}{2}I(R;B'X|B)_\omega[q\to q] + \tfrac{1}{2}(I(B'X;E)_\omega - I(B'X;B)_\omega)\,[qq] \geqslant \langle U_{\mathcal{N}}^{A\to XX_EB'E} : \rho^{AB}\rangle. \tag{82}$$

Though, we could perform this task by using less quantum communication and entanglement if we exploit MC–QSI first followed by state redistribution (as we do in the classically-assisted state redistribution protocol). The first step of the protocol implements the following resource inequality

$$\langle \rho^{AB}\rangle + I(X;R|B)[c\to c] + H(X|RB)[cc] \geqslant \langle \overline{\Delta}^{X\to X_AX_B} \circ \mathcal{N}^{A\to B'XE} : \rho^{AB}\rangle,$$

while the second is as follows:

$$\langle \sigma^{B'X_AE|BX_B}\rangle + \tfrac{1}{2}I(B';R|BX)[q\to q] + \tfrac{1}{2}(I(B';E|X) - I(B';B|X))[qq] \geqslant \langle \sigma^{X_AE|B'BX_B}\rangle.$$

Overall, the resource inequality for simulating the quantum instrument is as follows:

$$\langle \rho^{AB}\rangle + I(X;R|B)[c\to c] + H(X|RB)[cc] + \tfrac{1}{2}I(B';R|BX)[q\to q] + \tfrac{1}{2}(I(B';E|X) - I(B';B|X))[qq] \geqslant \langle U_{\mathcal{N}}^{A\to XX_EB'E} : \rho^{AB}\rangle,$$

which is a cheaper simulation than in (82) because we are using classical communication and common randomness to achieve part of the task, rather than quantum communication and entanglement for the whole protocol. One would expect to have such a savings, since a quantum instrument has both a classical and quantum output. We remark that this approach is very similar to classically-assisted state redistribution from the previous section, with the exception that we require the common randomness since the goal is to simulate the instrument in full, rather than to distill entanglement. A special case of the above reverse Shannon theorem occurs when there is no quantum side information available, in which the resource inequality reduces to

$$I(X;R)[c\to c] + H(X|R)[cc] + \tfrac{1}{2}I(B';R|X)[q\to q] + \tfrac{1}{2}I(B';E|X)[qq] \geqslant \langle U_{\mathcal{N}}^{A\to XX_EB'E} : \rho^A\rangle.$$

5.6.3. Classical communication cost in local purity distillation. We can also exploit MC–QSI to improve upon the classical communication cost in local purity distillation [41]. This leads to the following improvement of theorem 1 of [41]:

Theorem 13. *The one-way distillable local purity of the state ρ^{AB} is given by $\kappa_\to = \kappa_\to^*$, where*

$$\kappa_\to^*(\rho^{AB},R) = \kappa(\rho^A) + \kappa(\rho^B) + P_\to(\rho^{AB},R).$$

In the above, we have the definitions

$$\kappa(\omega^C) \equiv \log d_C - H(C)_\omega,$$
$$P_{\rightarrow}(\rho^{AB}, R) \equiv \lim_{k\to\infty} \frac{1}{k} P^{(1)}((\rho^{AB})^{\otimes k}, kR),$$

and

$$P^{(1)}(\rho^{AB}, R) \equiv \max_{\Lambda} \{I(Y;B)_\sigma : I(Y;E|B) \leqslant R\},$$
$$\sigma^{YBE} \equiv (\mathcal{M}_\Lambda \otimes I^{BE})(\psi^{ABE}),$$

where ψ^{ABE} is a purification of ρ^{AB}, $\mathcal{M}_\Lambda$ is a measurement map corresponding to the POVM Λ, and the maximization is over all POVMs mapping Alice's system A to a classical system Y.

The improvement of theorem 1 of [41] comes about by reducing the classical communication rate from $I(Y;EB)$ to $I(Y;E|B)$ by employing the MC–QSI protocol in the achievability part. The converse part of the theorem (in (19) of [41]) gets improved as follows:

$$nR = \log d_Y \geqslant H(Y) \geqslant H(Y|B^n) \geqslant I(Y;E^n|B^n).$$

It is apparent that the multi-letter nature of the converse theorem is what led to the ability to improve the theorem, so that for any finite k, the above revision of the theorem improves upon the previous one, but they are both optimal in the regularized limit. This leads us to believe that even further improvements might be possible.

5.7. Entropic uncertainty relation with QSI

We close by relating the MC–QSI protocol to recent work on an entropic uncertainty relation in the presence of quantum memory [52, 9, 59, 14, 31]. This uncertainty relation characterizes the ability of two parties to predict the outcomes of measurements on another system, by exploiting the quantum systems in their possession. The formal statement of the uncertainty relation applies to a tripartite state ρ^{ABC} and is as follows:

$$H(X|B) + H(Z|C) \geqslant \log_2(1/c_1). \tag{83}$$

The two entropies are with respect to the following states resulting from applying measurement maps for Λ and Γ to the A system:

$$\sum_x |x\rangle\langle x|^X \otimes \mathrm{Tr}_{AC}\{\Lambda_x^A \rho^{ABC}\},$$
$$\sum_z |z\rangle\langle z|^Z \otimes \mathrm{Tr}_{AB}\{\Gamma_z^A \rho^{ABC}\},$$

and c characterizes the non-commutativity of the measurements:

$$c_1 \equiv \max_{x,z} \|\sqrt{\Lambda_x}\sqrt{\Gamma_z}\|_\infty^2.$$

This uncertainty relation is useful conceptually, but it also has operational applications to quantum key distribution [9, 59], in relating data compression to privacy amplification [50, 54], and in constructing capacity-achieving quantum error correction codes (for certain channels) [62] because it is formulated in terms of entropies. Another statement of the above entropic uncertainty relation is as follows [9, 31]:

$$H(X|B) + H(Z|B) \geqslant \log_2(1/c_2) + H(A|B), \tag{84}$$

where

$$c_2 \equiv \max_{x,z} \sqrt{\mathrm{Tr}\{\Lambda_x \Gamma_z\}}.$$

Here, we show how the above uncertainty relations apply in bounding from below the nonlocal classical resources required in two different MC–QSI protocols. First, suppose that Alice would like to simulate the measurement $\{\Lambda_x^A\}$ on the state ρ^{ABC} and send the outcomes to Bob. Let R be a system that purifies the state ρ^{ABC}. Then the MC–QSI protocol corresponds to the following resource inequality:

$$\langle \rho^{ABC} \rangle + I(X; RC|B)[c \to c] + H(X|RBC)[cc] \geqslant \langle \Lambda^A : \rho^{ABC} \rangle,$$

where the entropies are with respect to the state

$$\sum_x |x\rangle\langle x|^X \otimes \mathrm{Tr}_A\left\{\Lambda_x^A \, \psi^{RABC}\right\}.$$

The total classical cost of the above protocol is $H(X|B) = I(X; RC|B) + H(X|RBC)$. For the second protocol, suppose that Alice would like to simulate the measurement $\{\Gamma_z^A\}$ on the state ρ^{ABC} and send the outcomes to Charlie. Then the MC–QSI protocol in this case corresponds to the following resource inequality:

$$\langle \rho^{ABC} \rangle + I(Z; RB|C)[c \to c] + H(Z|RBC)[cc] \geqslant \langle \Gamma^A : \rho^{ABC} \rangle,$$

where the entropies are with respect to the state

$$\sum_z |z\rangle\langle z|^Z \otimes \mathrm{Tr}_A\left\{\Gamma_z^A \, \psi^{RABC}\right\}.$$

The total classical cost of the above protocol is $H(Z|C) = I(Z; RB|C) + H(Z|RBC)$.

Using the entropic uncertainty relation in (83), we can then bound from below the total classical cost of the above protocols as follows:

$$I(X; RC|B) + H(X|RBC) + I(Z; RB|C) + H(Z|RBC) = H(X|B) + H(Z|C) \geqslant \log_2(1/c_1).$$

We can also apply the uncertainty relation in (84) to bound from below the total common randomness cost:

$$\begin{aligned} H(X|RBC) + H(Z|RBC) &\geqslant \log_2(1/c_2) + H(A|RBC) \\ &= \log_2(1/c_2) - H(A), \end{aligned}$$

where the last equality follows because the state on $RABC$ is pure. Since this lower bound might sometimes be negative but the entropies $H(X|RBC)$ and $H(Z|RBC)$ are always positive, we can revise the above lower bound to be as follows:

$$H(X|RBC) + H(Z|RBC) \geqslant \max\{\log_2(1/c_2) - H(A), 0\}.$$

Given that lower bounds on the total classical cost and the total common randomness exist, one might be tempted to think that a lower bound on the total information should exist as well. One might conjecture it to be of the following form:

$$I(X; RC|B) + I(Z; RB|C) \geqslant l,$$

where l is some non-negative parameter that depends only on the measurements and not on the state. Such a universal, state-independent lower bound cannot hold in general, however. A simple counterexample demonstrates that the following lower bound for strong subadditivity is the best that one might hope for:

$$I(X; RC|B) + I(Z; RB|C) \geqslant 0.$$

Indeed, suppose that ρ^{ABC} is a pure product state. Then $I(X; RC|B)$ is equal to zero because R and C have no correlations with the measurement output X on A, and $I(Z; RB|C) = 0$ for a similar reason.

6. Non-feedback measurement compression with quantum side information

Our final contribution concerns measurement compression with quantum side information, in the case where the sender is not required to obtain the outcome of the simulation, that is, a non-feedback simulation. We construct a protocol for this task by simply combining elements of other protocols described earlier in the review. Moreover, we show that the protocol is optimal by proving a single-letter converse for the associated rate region. We omit the detailed definition of the information processing task here because it is the obvious non-feedback relaxation along the lines of section 3 for the definition of the MC–QSI task from section 5.

Theorem 14 (Non-feedback MC–QSI)**.** *Let ρ^{AB} be a source state and $\mathcal{N}$ a quantum instrument to simulate on this state:*

$$(\mathcal{N}^{A\to AX} \otimes I^B)(\rho^{AB}) = \sum_x \left(\mathcal{N}_x^A \otimes I^B\right)(\rho^{AB}) \otimes |x\rangle\langle x|^X.$$

There exists a protocol for faithful non-feedback simulation of the quantum instrument with classical communication rate R and common randomness rate S if and only if R and S are in the union of the following regions:

$$\begin{aligned} R &\geqslant I(W;R|B), \\ R+S &\geqslant I(W;XR|B), \end{aligned} \tag{85}$$

where the entropies are with respect to a state of the following form:

$$\sum_{x,w} p_{X|W}(x|w)|w\rangle\langle w|^W \otimes |x\rangle\langle x|^X \otimes Tr_A\left\{\left(I^R \otimes \mathcal{M}_w^A \otimes I^B\right)\left(\phi_\rho^{RAB}\right)\right\}, \tag{86}$$

ϕ_ρ^{RAB} is some purification of the state ρ^{AB}, and the union is with respect to all decompositions of the original instrument $\mathcal{N}$ of the form:

$$(\mathcal{N}^{A\to AX} \otimes I^B)(\rho^{AB}) = \sum_{x,w} p_{X|W}(x|w)\left(\mathcal{M}_w^A \otimes I^B\right)(\rho^{AB}) \otimes |x\rangle\langle x|^X. \tag{87}$$

While demonstrating achievability of the quoted rates will consist of the routine combination of elements from other parts of the review, the converse is more subtle. In particular, a general protocol for non-feedback MC–QSI will have Bob perform an instrument on the B^n system. Arguing that it is sufficient to restrict to states of the form (86) will involve comparing that protocol to a related simulation in which the instrument is implemented approximately by Alice. While the modified protocol would generally require more communication than the original, for the purposes of the converse, it need not significantly increase the relevant mutual informations.

Proof Sketch of Achievability. The protocol for achievability naturally combines elements of protocols that we have considered in section 3 for non-feedback measurement compression and in section 5 for measurement compression with quantum side information. The protocol begins with Alice and Bob sharing many copies of a state ρ^{AB}. They would like to simulate an instrument $\mathcal{N}^{A\to AX}$, composed of the completely positive, trace non-increasing maps $\{\mathcal{N}_x^A\}$, so that they end up with many copies of a state of the following form:

$$(\mathcal{N}^{A\to AX} \otimes I^B)(\rho^{AB}) = \sum_x \left(\mathcal{N}_x^A \otimes I^B\right)(\rho^{AB}) \otimes |x\rangle\langle x|^X.$$

We omit the details of the proof of the achievability part because it follows readily from the methods detailed in sections 3 and 5. Instead, we state the achievability part as the following resource inequality:

$$\langle \rho^{AB} \rangle + I(W;R|B)[c\to c] + I(W;X|RB)[cc] \geqslant \langle \mathcal{N}^{A\to AX}(\rho^{AB}) \rangle. \tag{88}$$

where the information quantities are with respect to a state of the following form:

$$\sum_{x,w} p_{X|W}(x|w)|w\rangle\langle w|^W \otimes |x\rangle\langle x|^X \otimes \mathrm{Tr}_A\left\{\left(I^R \otimes \mathcal{M}_w^A \otimes I^B\right)\left(\phi_\rho^{RAB}\right)\right\}. \tag{89}$$

In the above, ϕ_ρ^{RAB} is a purification of the state ρ^{AB} and the maps $\left\{\mathcal{M}_w^A\right\}$ arise from a decomposition of the original instrument into the following form:

$$\sum_x \mathcal{N}_x^A(\sigma) \otimes |x\rangle\langle x|^X = \sum_{x,w} p_{X|W}(x|w)\mathcal{M}_w^A(\sigma) \otimes |x\rangle\langle x|^X,$$

when acting on some arbitrary state σ. In particular, the protocol operates by Alice and Bob performing a simulation of $\mathcal{M}_w^A$, though Alice hashes the outcome of the simulated measurement. She sends the hash along to Bob using noiseless classical bits channels, and he then performs sequential decoding to search among all of the post-measurement states that are consistent with the hash and his share of the common randomness. This causes a negligible disturbance to the shared state in the asymptotic limit as long as the communication rates are as in (88). Finally, he simulates the classical post-processing channel $p_{X|W}(x|w)$ locally, leading to a savings in the cost of common randomness consumption. □

With the achievability part in hand, we now move on to the proof of the converse.

Proof of Converse. We now prove this converse part. A modification of figure 7 (without the extra processing of L and M on Alice's side) depicts the most general protocol for a non-feedback simulation of the measurement with QSI. The protocol begins with the reference, Alice, and Bob sharing many copies of the state ϕ_ρ^{RAB} and Alice sharing common randomness M with Bob. She then chooses a quantum instrument $\Upsilon^{(m)}$ based on the common randomness M and performs it on her systems A^n. The measurement returns outcome L, and the overall state is as follows:

$$\theta^{R^nA^nB^nLM} \equiv \sum_{l,m} \frac{1}{|\mathcal{M}|}\left(\Upsilon_l^{(m)}\right)^{A^n}\left(\left(\phi_\rho^{RAB}\right)^{\otimes n}\right) \otimes |l\rangle\langle l|^L \otimes |m\rangle\langle m|^M,$$

where $\Upsilon_l^{(m)}$ is a completely positive, trace non-increasing map. Alice sends the register L to Bob. Based on L and M, he performs some quantum instrument on his systems B^n with trace non-increasing maps $\{\mathcal{F}_s^{(lm)}\}$ followed by the stochastic map $p_{\hat{X}^n|S,L,M}(\hat{x}^n|s,l,m)$ to give his estimate $\hat{x}^n$ of the measurement outcome. The resulting state is as follows:

$$\begin{aligned}\omega^{R^nA^nLMSB^n\hat{X}^n} \equiv \sum_{l,m,s,\hat{x}^n} \frac{1}{|\mathcal{M}|} p_{\hat{X}^n|S,L,M}(\hat{x}^n|s,l,m)\left(\left(\Upsilon_l^{(m)}\right)^{A^n} \otimes \left(\mathcal{F}_s^{(lm)}\right)^{B^n}\right)\left(\left(\phi_\rho^{RAB}\right)^{\otimes n}\right)\\ \otimes |l\rangle\langle l|^L \otimes |m\rangle\langle m|^M \otimes |s\rangle\langle s|^S \otimes |\hat{x}^n\rangle\langle\hat{x}^n|^{\hat{X}^n}.\end{aligned}$$

The following condition should hold for all $\epsilon > 0$ and sufficiently large n for a faithful non-feedback simulation:

$$\left\|\omega^{R^n\hat{X}^nB^n} - \sum_{x^n} \mathrm{Tr}_{A^n}\left\{(I \otimes \mathcal{N}_{x^n})\left(\phi_\rho^{RAB}\right)^{\otimes n}\right\} \otimes |x^n\rangle\langle x^n|^{X^n}\right\|_1 \leqslant \epsilon, \tag{90}$$

where $\hat{X}^n$ is a classical register isomorphic to X^n.

We prove the first bound as follows:

$$\begin{aligned} nR &\geqslant H(L)_\theta \\ &\geqslant I(L; MB^nR^n)_\theta \\ &= I(LMB^n; R^n)_\theta + I(L; MB^n)_\theta - I(R^n; MB^n)_\theta \\ &\geqslant I(LMB^n; R^n)_\theta - I(R^n; B^n)_\theta \end{aligned}$$

$$
\begin{aligned}
&\geqslant I(LMSB^n; R^n)_\omega - I(R^n; B^n)_\omega - n\epsilon' \\
&= H(R^n|B^n)_\omega - H(R^n|LMSB^n)_\omega - n\epsilon' \\
&\geqslant \sum_k [H(R_k|B_k)_\omega - H(R_k|LMSB_k)_\omega] - n2\epsilon' \\
&= \sum_k I(LMS; R_k|B_k)_\omega - n2\epsilon' \\
&= nI(LMS; R|BK)_\sigma - n2\epsilon' \\
&\geqslant nI(LMS; R|BK)_\sigma + nI(R; K|B)_\sigma - n3\epsilon' \\
&= nI(LMSK; R|B)_\sigma - n3\epsilon'.
\end{aligned}
$$

The first two inequalities are similar to what we had before. The first equality is an identity for quantum mutual information. The third inequality follows because there are no correlations between R^nB^n and M so that $I(MB^n; R^n)_\omega = I(B^n; R^n)_\omega$. The fourth inequality follows from quantum data processing of LMB^n to produce $LMSB^n$ and from the fact that this does not change the state too much (we apply the condition in (90) and the Alicki–Fannes' inequality). The second equality is an identity for quantum mutual information. The fifth inequality follows from strong subadditivity of quantum entropy:

$$
H(R^n|LMSB^n)_\omega \leqslant \sum_k H(R_k|LMSB_k)_\omega,
$$

and because the state on R^nB^n is close to a tensor-power state so that by lemma 10, we have

$$
H(R^n|B^n)_\omega \geqslant \sum_k H(R_k|B_k)_\omega - n\epsilon'.
$$

The third equality is another identity. The fourth equality comes about by defining the state σ as follows:

$$
\begin{aligned}
\sigma^{RALMS\hat{X}K} \equiv \sum_{l,m,k,\hat{x},s} &\frac{1}{n|\mathcal{M}|} p_{\hat{X}|LMS}(\hat{x}|lms) \\
&\times \mathrm{Tr}_{(RAB)_1^{k-1}(RAB)_{k+1}^n} \left\{ \left(\left(\Upsilon_l^{(m)} \right)^{A^n} \otimes \left(\mathcal{F}_s^{(lmk)} \right)^{B^n} \right) \left(\left(\phi_\rho^{RAB} \right)^{\otimes n} \right) \right\} \\
&\otimes |l\rangle\langle l|^L \otimes |m\rangle\langle m|^M \otimes |s\rangle\langle s|^S \otimes |\hat{x}\rangle\langle \hat{x}|^{\hat{X}} \otimes |k\rangle\langle k|^K, \qquad (91)
\end{aligned}
$$

where the map $p_{\hat{X}|LMS}(\hat{x}|lms)$ is defined from $p_{\hat{X}^n|LMS}(\hat{x}^n|lms)$ by keeping only the kth symbol from $\hat{x}^n$. It also follows by exploiting the fact that K is a uniform classical random variable, with distribution $1/n$, determining which systems $R_kA_kB_k\hat{X}_k$ to select. From the fact that the measurement simulation is faithful, we can apply the Alicki–Fannes' inequality to conclude that

$$
I(R\hat{X}; K|B)_\sigma = |I(R\hat{X}; K|B)_\sigma - I(RX; K|B)_\tau| \leqslant \epsilon', \qquad (92)
$$

where τ is a state like σ but resulting from the tensor-power state for ideal measurement compression (and due to its IID structure, it has no correlations with any particular system k so that $I(RX; K|B)_\tau = 0$). The same reasoning along with strong subadditivity also implies that

$$
I(R; K|B)_\sigma \leqslant \epsilon'. \qquad (93)
$$

The final equality is an application of the chain rule for quantum mutual information. The state σ for the final information term has the form:

$$
\mathcal{M}^{AB\to ABX}(\phi^{RAB}) = \sum_{x,w} p_{X|W}(x|w)|x\rangle\langle x|^X \otimes \mathcal{M}_w^{AB}(\phi^{RAB}), \qquad (94)
$$

with $LMSK = W$ and the completely positive, trace non-increasing maps $\mathcal{M}_w^{AB}$ defined by

$$\varrho^{AB} \mapsto \frac{1}{n|\mathcal{M}|}\mathrm{Tr}_{(AB)_1^{k-1}(AB)_{k+1}^n}\left\{\left(\left(\Upsilon_l^{(m)}\right)^{A^n} \otimes \left(\mathcal{F}_s^{(lm)}\right)^{B^n}\right)\left(\left(\phi_\rho^{AB}\right)^{\otimes k-1} \otimes \varrho^{AB} \otimes \left(\phi_\rho^{AB}\right)^{\otimes n-k}\right)\right\}.$$

At this point, we have proved the first inequality in (85) for a state of the form in (94) where the map $\mathcal{M}_w^{AB}$ acts on the joint system AB. We now show that it is possible to construct from $\mathcal{M}_w^{AB}$ a map acting only on the system A (as stated in the theorem) causing only a negligible change to the information quantity in (85). The idea behind this is a simple application of Uhlmann's theorem. First, consider that the following inequality holds from the condition in (90) and from monotonicity of trace distance:

$$\left\| \sum_{x,w} p_{X|W}(x|w)\mathrm{Tr}_A\{\mathcal{M}_w^{AB}(\phi^{RAB})\} - \sum_x \mathrm{Tr}_A\{\mathcal{N}_x(\phi^{RAB})\} \right\|_1 \leqslant \epsilon$$

The state ϕ^{RAB} is a purification of $\sum_x \mathrm{Tr}_A\{\mathcal{N}_x(\phi^{RAB})\}$, and the following state:

$$\sum_{w,x,i} M_{w,i}^{AB}|\phi\rangle^{RAB}|w\rangle^W \sqrt{p_{X|W}(x|w)}|x\rangle^X|i\rangle^I, \tag{95}$$

is a purification of $\sum_{x,w} p_{X|W}(x|w)\mathrm{Tr}_A\{\mathcal{M}_w^{AB}(\phi^{RAB})\}$, where we assume that the completely positive maps $\mathcal{M}_w^{AB}$ have the following Kraus representation:

$$\mathcal{M}_w^{AB}(\rho^{AB}) = \sum_i M_{w,i}^{AB}\rho^{AB}\left(M_{w,i}^\dagger\right)^{AB}.$$

By Uhlmann's theorem, there exists an isometry $U^{A\to AWXI}$ such that the trace distance between $U^{A\to AWXI}(\phi^{RAB})$ and the state in (95) is less than $2\sqrt{\epsilon}$. To have the map $\mathcal{M}^{AB\to ABX}$ be implemented solely on Alice's system, we can simply perform the isometry $U^{A\to AWXI}$, discard the register I, and perform von Neumann measurements of the registers W and X. (One could also discard register X, and then process W with $p_{X|W}(x|w)$ to produce X—it is possible to do this since X is classical.) This amounts to an approximate implementation of the following instrument:

$$\sum_{x,w} p_{X|W}(x|w)|x\rangle\langle x|^X \otimes |w\rangle\langle w|^W \otimes \mathcal{M}_w^{AB}(\phi^{RAB}),$$

from which we can discard register W to obtain an approximation of the map $\mathcal{M}^{AB\to ABX}$. Thus, from the map $\mathcal{M}^{AB\to ABX}$, it is possible to construct a nearby map of the form in (87), so that it suffices to optimize over the class of decompositions given in (87).

We now prove the second bound:

$$\begin{aligned}
n(R+S) &\geqslant H(LM)_\theta \\
&\geqslant H(LM|B^n)_\theta \\
&\geqslant I(LM;\hat{X}^nR^n|B^n)_\theta \\
&= I(LMB^n;\hat{X}^nR^n)_\theta - I(B^n;\hat{X}^nR^n)_\theta \\
&\geqslant I(LMSB^n;\hat{X}^nR^n)_\omega - I(B^n;\hat{X}^nR^n)_\omega - n\epsilon' \\
&= H(\hat{X}^nR^n|B^n)_\omega - H(\hat{X}^nR^n|LMSB^n)_\omega - n\epsilon' \\
&\geqslant \sum_k [H(\hat{X}_kR_k|B_k)_\omega - H(\hat{X}_kR_k|LMSB_k)_\omega] - n2\epsilon' \\
&= \sum_k I(LMS;\hat{X}_kR_k|B_k)_\omega - n2\epsilon' \\
&= nI(LMS;\hat{X}R|KB)_\sigma - n2\epsilon' \\
&\geqslant nI(LMS;\hat{X}R|KB)_\sigma + nI(K;\hat{X}R|B)_\sigma - n3\epsilon' \\
&= nI(LMSK;\hat{X}R|B)_\sigma - n3\epsilon'.
\end{aligned}$$

The first three inequalities follow from similar reasons as our previous inequalities. The first equality is an identity. The fourth inequality follows from quantum data processing of LMB^n to produce $LMSB^n$ and from the fact that this does not change the state too much (we apply the condition in (90) and the Alicki-Fannes' inequality). The fifth inequality follows from strong subadditivity of entropy:

$$H(\hat{X}^n R^n|LMSB^n)_\omega \leqslant \sum_k H(\hat{X}_k R_k|LMSB_k)_\omega,$$

and from the fact that the measurement simulation is faithful so that

$$\left| H(\hat{X}^n R^n|B^n)_\omega - \sum_k H(\hat{X}_k R_k|B_k)_\omega \right| \leqslant n\epsilon',$$

where we have applied a variation of lemma 10. The third equality is an identity. The fourth equality follows by considering the state σ as defined in (91). The sixth inequality follows from (92). The final equality is the chain rule for quantum mutual information. We can then consider the same argument as stated before in order to construct a map acting only on A from one acting on AB. Similarly, the resulting state has the form in (87). □

7. Conclusion

This review provided a second look at Winter's measurement compression theorem [65], detailing the information processing task, providing examples for understanding it, reviewing Winter's achievability proof, and detailing a new approach to its single-letter converse theorem. We proved a new theorem characterizing the optimal rates for classical communication and common randomness for a measurement compression protocol where the sender is not required to obtain the outcome of the measurement simulation. We then reviewed the Devetak–Winter theorem on classical data compression with quantum side information, providing new proofs of the achievability and converse parts of this theorem. From there, we presented a new protocol called measurement compression with quantum side information (a protocol first announced in [38]). This protocol has several applications, including its part in the 'classically-assisted state redistribution' protocol, which is the most general protocol on the static side of the quantum information theory tree, and its role in reducing the classical communication cost in local purity distillation [41]. We then outlined a connection between this protocol and recent work in entropic uncertainty relations. Finally, we proved a single-letter theorem for the task of measurement compression with quantum side information when the sender is not required to obtain the outcome of the measurement simulation.

There are several open questions to consider going forward from here. First, are there applications of the MC–QSI protocol to rate distortion, as was the case for the Luo–Devetak protocol in [44]? Are there further applications of the measurement compression protocol in general? Is it possible to formulate a measurement compression protocol that is independent of the state on which it acts (similar to the general reverse Shannon theorem from [5, 10])? The answers to these questions could further illuminate our understanding of quantum measurement and address other important areas of quantum information theory.

Acknowledgments

We thank Ke Li (Carl) for suggesting the possibility of the measurement compression with quantum side information protocol to us. We are grateful to Cedric Beny, Aram Harrow, Daniel Gottesman, Masanao Ozawa, and Andreas Winter for useful discussions. MMW acknowledges

support from the Centre de Recherches Mathématiques at the University of Montreal. He also acknowledges both Nagoya University and the Perimeter Institute for Theoretical Physics, where some of this work was conducted. PH acknowledges support from the Canada Research Chairs program, the Perimeter Institute, CIFAR, FQRNT's INTRIQ, NSERC, and ONR through grant N000140811249. FB acknowledges support from the Program for Improvement of Research Environment for Young Researchers from Special Coordination Funds for Promoting Science and Technology (SCF) commissioned by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. MH received support from the Chancellor's postdoctoral research fellowship, University of Technology Sydney (UTS), and was also partly supported by the National Natural Science Foundation of China (grant no. 61179030) and the Australian Research Council (grant no. DP120103776).

Appendix A. Typical sequences and typical subspaces

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:

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

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 [15]). 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 $p_{X^n}(x^n) \equiv p_X(x_1) \ldots p_X(x_n)$, ϵ is positive number that becomes arbitrarily small as n becomes large, and c is some positive constant. Then the following three properties hold [15]

$$\Pr\left\{ X^n \in T_\delta^{X^n} \right\} \geqslant 1 - \epsilon, \tag{A.1}$$

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

$$\forall x^n \in T_\delta^{X^n} : \quad 2^{-n[H(X)+c\delta]} \leqslant p_{X^n}(x^n) \leqslant 2^{-n[H(X)-c\delta]}. \tag{A.3}$$

We omit using c in the main text and instead subsume it as part of δ.

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

$$\rho \equiv \sum_x p_X(x) |x\rangle\langle x|,$$

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

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

and let $\Pi_{\rho,\delta}^n$ denote the projector onto it. Then properties analogous to (A.1)–(A.3) 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:

$$\mathrm{Tr}\{\Pi^n_{\rho,\delta}\,\rho^{\otimes n}\} \geqslant 1-\epsilon, \tag{A.4}$$

$$\mathrm{Tr}\{\Pi^n_{\rho,\delta}\} \leqslant 2^{n[H(B)+c\delta]}, \tag{A.5}$$

$$2^{-n[H(B)+c\delta]}\,\Pi^n_{\rho,\delta} \leqslant \Pi^n_{\rho,\delta}\,\rho^{\otimes n}\,\Pi^n_{\rho,\delta} \leqslant 2^{-n[H(B)-c\delta]}\,\Pi^n_{\rho,\delta}, \tag{A.6}$$

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 the pruned distribution in (24), 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:

$$\Pi^n_{\rho_{x^n},\delta} \equiv \bigotimes_{x\in\mathcal{X}} \Pi^{I_x}_{\rho_x,\delta},$$

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 $\Pi^{I_x}_{\rho_x,\delta}$ is the typical projector for the state ρ_x. The conditionally typical subspace has the three following properties:

$$\mathrm{Tr}\{\Pi^n_{\rho_{x^n},\delta}\,\rho_{x^n}\} \geqslant 1-\epsilon, \tag{A.7}$$

$$\mathrm{Tr}\{\Pi^n_{\rho_{x^n},\delta}\} \leqslant 2^{n[H(B|X)+c\delta]}, \tag{A.8}$$

$$2^{-n[H(B|X)+c\delta]}\,\Pi^n_{\rho_{x^n},\delta} \leqslant \Pi^n_{\rho_{x^n},\delta}\,\rho_{x^n}\,\Pi^n_{\rho_{x^n},\delta} \leqslant 2^{-n[H(B|X)-c\delta]}\,\Pi^n_{\rho_{x^n},\delta}, \tag{A.9}$$

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 [23, 63, 60]:

$$\forall x^n \in T^{X^n}_\delta : \mathrm{Tr}\{\rho_{x^n}\,\Pi_\rho\} \geqslant 1-\epsilon,$$
$$\sum_{x^n} p'_{X^m}(x)\rho_{x^n} \leqslant [1-\epsilon]^{-1}\rho^{\otimes n}. \tag{A.10}$$

Appendix B. Useful lemmas

Here we collect some useful lemmas.

Lemma 15 (Gentle Operator lemma [63, 47])**.** *Let Λ be a positive operator where $0 \leqslant \Lambda \leqslant I$ (usually Λ is a POVM element), ρ a state, and ϵ a positive number such that the probability of detecting the outcome Λ is high:*

$$\mathrm{Tr}\{\Lambda\rho\} \geqslant 1-\epsilon.$$

Then the measurement causes little disturbance to the state ρ:

$$\|\rho - \sqrt{\Lambda}\rho\sqrt{\Lambda}\|_1 \leqslant 2\sqrt{\epsilon}.$$

Lemma 16 (Gentle Operator lemma for Ensembles [63, 47, 60])**.** *Given an ensemble $\{p_X(x), \rho_x\}$ with expected density operator $\rho \equiv \sum_x p_X(x)\rho_x$, suppose that an operator Λ such that $I \geqslant \Lambda \geqslant 0$ succeeds with high probability on the state ρ:*

$$\mathrm{Tr}\{\Lambda\rho\} \geqslant 1-\epsilon.$$

Then the subnormalized state $\sqrt{\Lambda}\rho_x\sqrt{\Lambda}$ is close in expected trace distance to the original state ρ_x:

$$\mathbb{E}_X\{\|\sqrt{\Lambda}\rho_X\sqrt{\Lambda} - \rho_X\|_1\} \leqslant 2\sqrt{\epsilon}.$$

Lemma 17. *Let ρ and σ be positive operators and Λ a positive operator such that $0 \leqslant \Lambda \leqslant I$. Then the following inequality holds*

$$\mathrm{Tr}\{\Lambda\rho\} \leqslant \mathrm{Tr}\{\Lambda\sigma\} + \|\rho - \sigma\|_1.$$

Lemma 18 (Non-commutative union bound [56])**.** *Let σ be a subnormalized state such that $\sigma \geqslant 0$ and $\mathrm{Tr}\{\sigma\} \leqslant 1$. Let $\Pi_1, \ldots, \Pi_N$ be projectors. Then the following 'non-commutative union bound' holds*

$$\mathrm{Tr}\{\sigma\} - \mathrm{Tr}\{\Pi_N \cdots \Pi_1 \sigma \Pi_1 \cdots \Pi_N\} \leqslant 2\sqrt{\sum_{i=1}^{N} \mathrm{Tr}\{(I - \Pi_i)\sigma\}}.$$

References

[1] Abeyesinghe A, Devetak I, Hayden P and Winter A 2009 The mother of all protocols: restructuring quantum information's family tree *Proc. R. Soc.* A **465** 2537–63 (arXiv:quant-ph/0606225)

[2] Ahlswede R and Winter A J 2002 Strong converse for identification via quantum channels *IEEE Trans. Inform. Theory* **48** 569–79 (arXiv:quant-ph/0012127)

[3] Alicki R and Fannes M 2004 Continuity of quantum conditional information *J. Phys. A: Math. Gen.* **37** L55–7 (arXiv:quant-ph/0312081)

[4] Audenaert K M R and Scheel S 2008 On random unitary channels *New J. Phys.* **10** 023011

[5] Bennett C H, Devetak I, Harrow A W, Shor P W and Winter A 2009 Quantum reverse Shannon theorem arXiv:0912.5537

[6] Bennett C H, Shor P W, Smolin J A and Thapliyal A V 2002 Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem *IEEE Trans. Inform. Theory* **48** 2637 (arXiv:quant-ph/0106052)

[7] Berger T 1971 *Rate Distortion Theory: A Mathematical Basis for Data Compression* (*Information and System Sciences*) (Englewood Cliffs, NJ: Prentice Hall)

[8] Berta M, Brandao F, Christandl M and Wehner S 2011 Entanglement cost of quantum channels arXiv:1108.5357

[9] Berta M, Christandl M, Colbeck R, Renes J M and Renner R 2010 The uncertainty principle in the presence of quantum memory *Nature Phys.* **6** 659–62 (arXiv:0909.0950)

[10] Berta M, Christandl M and Renner R 2011 The quantum reverse Shannon theorem based on one-shot information theory *Commun. Math. Phys.* **306** 579–615 (arXiv:0912.3805)

[11] Buscemi F 2006 On the minimum number of unitaries needed to describe a random-unitary channel *Phys. Lett.* A **360** 256–8

[12] Buscemi F, Hayashi M and Horodecki M 2008 Global information balance in quantum measurements *Phys. Rev. Lett.* **100** 210504

[13] Buscemi F, Keyl M, D'Ariano G M, Perinotti P and Werner R F 2005 Clean positive operator valued measures *J. Math. Phys.* **46** 082109

[14] Coles P J, Colbeck R, Li Y and Zwolak M 2012 Uncertainty relations from simple entropic properties *Phys. Rev. Lett.* **108** 210405 (arXiv:1112.0543)

[15] Cover T M and Thomas J A 1991 *Elements of Information Theory* (New York: Wiley)

[16] Csiszár I and Korner J 1981 *Information Theory: Coding Theorems for Discrete Memoryless Systems* (*Probability and Mathematical Statistics*) (Budapest: Akademiai Kiado)

[17] Cuff P 2008 Communication requirements for generating correlated random variables *Proc. Int. Symp. on Information Theory* (Toronto, Ontario, Canada) pp 1393–7 (arXiv:0805.0065)

[18] D'Ariano G M, Presti P Lo and Perinotti P 2005 Classical randomness in quantum measurements *J. Phys. A: Math. Gen.* **38** 5979–91 (arXiv:quant-ph/0408115)

[19] Datta N, Hsieh M-H and Wilde M M 2011 Quantum rate distortion, reverse Shannon theorems and source–channel separation *IEEE Trans. Inform. Theory* at press (doi:10.1109/TIT.2012.2215575) (arXiv:1108.4940)

[20] Davies E B and Lewis J T 1970 An operational approach to quantum probability *Commun. Math. Phys.* **17** 239–60

[21] Davies E B 1976 *Quantum Theory of Open Systems* (New York: Academic)

[22] Devetak I 2005 Distillation of local purity from quantum states *Phys. Rev.* A **71** 062303 (arXiv:quant-ph/0406234)

[23] Devetak I 2005 The private classical capacity and quantum capacity of a quantum channel *IEEE Trans. Inform. Theory* **51** 44–55 (arXiv:quant-ph/0304127)

[24] Devetak I 2006 Triangle of dualities between quantum communication protocols *Phys. Rev. Lett.* **97** 140503 (arXiv:quant-ph/0505138)
[25] Devetak I, Harrow A W and Winter A 2004 A family of quantum protocols *Phys. Rev. Lett.* **93** 230504 (arXiv:quant-ph/0308044)
[26] Devetak I, Harrow A W and Winter A 2008 A resource framework for quantum Shannon theory *IEEE Trans. Inform. Theory* **54** 4587–618 (arXiv:quant-ph/0512015)
[27] Devetak I and Winter A 2003 Classical data compression with quantum side information *Phys. Rev.* A **68** 042301 (arXiv:quant-ph/0209029)
[28] Devetak I and Winter A J 2004 Distilling common randomness from bipartite quantum states *IEEE Trans. Inform. Theory* **50** 3183–96 (arXiv:quant-ph/0304196)
[29] Devetak I and Yard J 2008 Exact cost of redistributing multipartite quantum states *Phys. Rev. Lett.* **100** 230501
[30] Gamal A El and Kim Y-H 2010 Lecture notes on network information theory arXiv:1001.3404v4
[31] Frank R L and Lieb E H 2012 Extended quantum conditional entropy and quantum uncertainty inequalities arXiv:1204.0825
[32] Giovannetti V, Lloyd S and Maccone L 2012 Achieving the Holevo bound via sequential measurements *Phys. Rev.* A **85** 012302 (arXiv:1012.0386)
[33] Groenewold H J 1971 A problem of information gain by quantal measurements *Int. J. Theor. Phys.* **4** 327–38
[34] Holevo A S 1982 *Probabilistic and Statistical Aspects of Quantum Theory* (Amsterdam: North-Holland)
[35] Holevo A S 1998 The capacity of the quantum channel with general signal states *IEEE Trans. Inform. Theory* **44** 269–73
[36] Horodecki M, Horodecki K, Horodecki P, Horodecki R, Oppenheim J, Sen(De) A and Sen U 2003 Local information as a resource in distributed quantum systems *Phys. Rev. Lett.* **90** 100402 (arXiv:quant-ph/0207168)
[37] Horodecki M, Horodecki P, Horodecki R, Oppenheim J, Sen(De) A, Sen U and Synak-Radtke B 2005 Local versus nonlocal information in quantum-information theory: formalism and phenomena *Phys. Rev.* A **71** 062307 (arXiv:quant-ph/0410090)
[38] Hsieh M-H and Wilde M M 2010 Trading classical communication, quantum communication and entanglement in quantum Shannon theory *IEEE Trans. Inform. Theory* **9** 4705–30 (arXiv:0901.3038)
[39] Hughston L P, Jozsa R and Wootters W K 1993 A complete classification of quantum ensembles having a given density matrix *Phys. Lett.* A **183** 14–18
[40] Kraus K 1983 *States, Effects and Operations: Fundamental Notions of Quantum Theory* (Berlin: Springer)
[41] Krovi H and Devetak I 2007 Local purity distillation with bounded classical communication *Phys. Rev.* A **76** 012321 (arXiv:0705.4089)
[42] Lindblad G 1972 An entropy inequality for quantum measurements *Commun. Math. Phys.* **28** 245–9
[43] Luo Z 2009 Topics in quantum cryptography, quantum error correction and channel simulation *PhD Thesis* University of Southern California, CA, USA
[44] Luo Z and Devetak I 2009 Channel simulation with quantum side information *IEEE Trans. Inform. Theory* **55** 1331–42 (arXiv:quant-ph/0611008)
[45] Martens H and de Muynck W M 1990 Nonideal quantum measurements *Found. Phys.* **20** 255–81
[46] Massar S and Popescu S 2000 Amount of information obtained by a quantum measurement *Phys. Rev.* A **61** 062303
[47] Ogawa T and Nagaoka H 2007 Making good codes for classical-quantum channel coding via quantum hypothesis testing *IEEE Trans. Inform. Theory* **53** 2261–6
[48] Ozawa M 1984 Quantum measuring processes of continuous observables *J. Math. Phys.* **25** 79–87
[49] Ozawa M 1986 On information gain by quantum measurements of continuous observables *J. Math. Phys.* **27** 759–63
[50] Renes J M 2011 Duality of privacy amplification against quantum adversaries and data compression with quantum side information *Proc. R. Soc.* A **467** 1604–23 (arXiv:1003.0703)
[51] Renes J M and Boileau J-C 2008 Physical underpinnings of privacy *Phys. Rev.* A **78** 032335 (arXiv:0803.3096)
[52] Renes J M and Boileau J-C 2009 Conjectured strong complementary information tradeoff *Phys. Rev. Lett.* **103** 020402 (arXiv:0806.3984)
[53] Renes J M and Renner R 2011 Noisy channel coding via privacy amplification and information reconciliation *IEEE Trans. Inform. Theory* **57** 7377–85 (arXiv:1012.4814)
[54] Renes J M and Renner R 2012 One-shot classical data compression with quantum side information and the distillation of common randomness or secret keys *IEEE Trans. Inform. Theory* **58** 1985–91 (arXiv:1008.0452)
[55] Schumacher B and Westmoreland M D 1997 Sending classical information via noisy quantum channels *Phys. Rev.* A **56** 131–8
[56] Sen P 2011 Achieving the Han–Kobayashi inner bound for the quantum interference channel by sequential decoding arXiv:1109.0802

[57] Shannon C E 1948 A mathematical theory of communication *Bell Syst. Tech. J.* **27** 379–423
[58] Slepian D and Wolf J K 1973 Noiseless coding of correlated information sources *IEEE Trans. Inform. Theory* **19** 471–80
[59] Tomamichel M and Renner R 2011 Uncertainty relation for smooth entropies *Phys. Rev. Lett.* **106** 110506 (arXiv:1009.2015)
[60] Wilde M M 2011 From classical to quantum Shannon theory arXiv:1106.1445
[61] Wilde M M, Guha S, Tan S-H and Lloyd S 2012 Explicit capacity-achieving receivers for optical communication and quantum reading *Proc. Int. Symp. on Information Theory (1–6 July 2012)* pp 551–5 (arXiv:1202.0518)
[62] Wilde M M and Renes J M 2012 Quantum polar codes for arbitrary channels *Proc. Int. Symp. on Information Theory (1–6 July 2012)* pp 334–8 (arXiv:1201.2906)
[63] Winter A 1999 Coding theorem and strong converse for quantum channels *IEEE Trans. Inform. Theory* **45** 2481–5
[64] Winter A J 2002 Compression of sources of probability distributions and density operators arXiv:quant-ph/0208131
[65] Winter A J 2004 'Extrinsic' and 'intrinsic' data in quantum measurements: asymptotic convex decomposition of positive operator valued measures *Commun. Math. Phys.* **244** 157–85 (arXiv:quant-ph/0109050)
[66] Winter A J and Massar S 2001 Compression of quantum-measurement operations *Phys. Rev.* A **64** 012311 (arXiv:quant-ph/0012128)
[67] Wyner A 1975 The common information of two dependent random variables *IEEE Trans. Inform. Theory* **21** 163–79
[68] Yard J and Devetak I 2009 Optimal quantum source coding with quantum side information at the encoder and decoder *IEEE Trans. Inform. Theory* **55** 5339–51 (arXiv:0706.2907)