MIT Open Access Articles

The Quantum Reverse Shannon Theorem and Resource Tradeoffs for Simulating Quantum Channels

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation	Bennett, Charles H., Igor Devetak, Aram W. Harrow, Peter W. Shor, and Andreas Winter. "The Quantum Reverse Shannon Theorem and Resource Tradeoffs for Simulating Quantum Channels." IEEE Trans. Inform. Theory 60, no. 5 (May 2014): 2926–2959.
As Published	http://dx.doi.org/10.1109/tit.2014.2309968
Publisher	Institute of Electrical and Electronics Engineers (IEEE)
Version	Author's final manuscript
Citable link	http://hdl.handle.net/1721.1/93188
Terms of Use	Creative Commons Attribution-Noncommercial-Share Alike
Detailed Terms	http://creativecommons.org/licenses/by-nc-sa/4.0/

Massachusetts Institute of Technology

DSpace@MIT

The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels

Charles H. Bennett, Igor Devetak, Aram W. Harrow, Peter W. Shor and Andreas Winter

arXiv:0912.5537v5 [quant-ph] 4 Mar 2014

***Abstract*—Dual to the usual noisy channel coding problem, where a noisy (classical or quantum) channel is used to simulate a noiseless one, reverse Shannon theorems concern the use of noiseless channels to simulate noisy ones, and more generally the use of one noisy channel to simulate another. For channels of nonzero capacity, this simulation is always possible, but for it to be efficient, auxiliary resources of the proper kind and amount are generally required. In the classical case, shared randomness between sender and receiver is a sufficient auxiliary resource, regardless of the nature of the source, but in the quantum case the requisite auxiliary resources for efficient simulation depend on both the channel being simulated, and the source from which the channel inputs are coming. For tensor power sources (the quantum generalization of classical IID sources), entanglement in the form of standard ebits (maximally entangled pairs of qubits) is sufficient, but for general sources, which may be arbitrarily correlated or entangled across channel inputs, additional resources, such as entanglement-embezzling states or backward communication, are generally needed. Combining existing and new results, we establish the amounts of communication and auxiliary resources needed in both the classical and quantum cases, the tradeoffs among them, and the loss of simulation efficiency when auxiliary resources are absent or insufficient. In particular we find a new single-letter expression for the excess forward communication cost of coherent feedback simulations of quantum channels (i.e. simulations in which the sender retains what would escape into the environment in an ordinary simulation), on non-tensor-power sources in the presence of unlimited ebits but no other auxiliary resource. Our results on tensor power sources establish a strong converse to the entanglement-assisted capacity theorem.**

Charles H. Bennett is with the IBM T.J. Watson Research Center, Yorktown Heights, NY 10598 (USA). This work was funded in part by ARDA contract DAAD19-01-0056 and DARPA QUEST contract HR0011-09-C0047. Email: chdbennett@gmail.com

Igor Devetak is with IMC Financial Markets, Poststrasse 20, 6300 Zug, Switzerland. This work was performed while he was at IBM T.J. Watson Research Center and the Department of Electrical Engineering at USC. Email: igor.devetak@gmail.com

Aram W. Harrow is with the Department of Physics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. This work was also performed while he was at University of Bristol and the University of Washington. He was funded by NSF grants CCF-0916400 and CCF-1111382 and ARO contract W911NF-12-1-0486. Email: aram@mit.edu

Peter W. Shor is with the Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA and was supported in part by NSF grants CCF-0431787 ("Quantum Channel Capacities and Quantum Complexity") and CCF-0829421 ("Physics Based Approaches to Quantum Algorithms"), as well as the NSF STC on Science of Information. Email: shor@math.mit.edu

Andreas Winter is with ICREA and Física Teòrica: Informació i Fenomens Quàntics, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain. During preparation of this paper he was also affiliated with the Department of Mathematics, University of Bristol and the Centre for Quantum Technologies, National University of Singapore. He acknowledges support by the U.K. EPRSC grant "QIP IRC", the Royal Society, the Philip Leverhulme Trust, EC integrated project QAP (contract IST-2005-15848), as well as STREPs QICS and QCS, and finally the ERC Advanced Grant "IRQUAT". Email: der.winter@gmail.com.

CONTENTS

I. Introduction

A. Motivation

In classical information theory, Shannon's celebrated noisy channel coding theorem [71] establishes the ability of any noisy memoryless channel N to simulate an ideal noiseless binary channel, and shows that its asymptotic efficiency or capacity for doing so is given by a simple expression

$$\begin{aligned} C(N) &= \max_p I(X;Y) \\ &= \max_p \big\{ H(X) + H(Y) - H(XY) \big\}, \end{aligned} \tag{1}$$

where H is the entropy, X the input random variable and $Y = N(X)$ the induced output variable. The capacity, in other words, is equal to the maximum, over input distributions p, of the input-output mutual information for a single use of the channel. Somewhat more recently, a dual theorem, the classical "reverse Shannon theorem" was proved [14], which states that for any channel N of capacity C, if the sender and receiver share an unlimited supply of random bits, an expected $Cn + o(n)$ uses of a noiseless binary channel are sufficient to exactly simulate n uses of the channel. In [84] a version of this construction is given which achieves asymptotically perfect simulation, works on a uniform blocksize $Cn + o(n)$, and uses an amount of shared randomness increasing linearly with n, in contrast to the exponential amount used in [14]. These simulations do not depend on the nature of the source, and work for arbitrarily varying as well as IID sources.

Together with the original Shannon theorem, these theorems show that in the presence of shared randomness, the asymptotic properties of a classical channel can be characterized by a single parameter, its capacity; with all channels of equal capacity being able to simulate one another with unit asymptotic efficiency in the presence of shared randomness. In [14] a quantum analog of the reverse Shannon theorem was conjectured, according to which quantum channels should be characterizable by a single parameter in the presence of unlimited shared entanglement between sender and receiver.

A (discrete memoryless) quantum channel can be viewed physically as a process wherein a quantum system, originating with a sender Alice, is split into a component for a receiver Bob and another for an inaccessible environment (commonly referred to as Eve). Mathematically it can be viewed as an isometric embedding $\mathcal{N}^{A\to BE}$ of Alice's Hilbert space (A) into the joint Hilbert space of Bob (B) and Eve (E). Tracing out Eve yields a completely positive, trace-preserving linear map on density operators from A to B, which we denote $\mathcal{N}^{A\to B}$. Operationally, the two pictures are equivalent, but we will sometimes find it convenient mathematically to work with one or the other.

The theory of quantum channels is richer and less well understood than that of classical channels. Unlike classical channels, quantum channels have multiple inequivalent capacities, depending on what one is trying to use them for, and what additional resources are brought into play. These include

- The ordinary classical capacity C, defined as the maximum asymptotic rate at which classical bits can be transmitted reliably through the channel, with the help of a quantum encoder and decoder.
- The ordinary quantum capacity Q, which is the maximum asymptotic rate at which qubits can be transmitted under similar circumstances.
- The private classical capacity P, which is the maximum rate at which classical bits can be transmitted to Bob while remaining private from Eve, who is assumed to hold the channel's environment E.
- The classically assisted quantum capacity Q_2, which is the maximum asymptotic rate of reliable qubit transmission with the help of unlimited use of a 2-way classical side channel between sender and receiver.
- The entanglement-assisted classical capacity C_E [13], [14], which is the maximum asymptotic rate of reliable bit transmission with the help of unlimited pure state entanglement shared between the sender and receiver.
- Similarly, one can define the entanglement-assisted quantum capacity Q_E [13], [14], which is simply $\frac{1}{2}C_E$, by teleportation [9] and super-dense coding [15].[1]

Somewhat unexpectedly, the entanglement assisted capacities are the simplest to calculate, being given by an expression analogous to Eq. (1). In [14] (see also [54]) it was shown that

$$C_E(\mathcal{N}) = \max_\rho \big\{ H(\rho) + H(\mathcal{N}(\rho)) - H(I \otimes \mathcal{N}(\Phi_\rho)) \big\}, \tag{2}$$

where the optimization is over all density matrices ρ on A and Φ_ρ^{RA} is a purification of ρ by a reference system R (meaning that Φ_ρ^{RA} is a pure state and $\mathrm{Tr}_R\, \Phi_\rho^{RA} = \rho^A$). The entanglement-assisted capacity formula Eq. (2) is formally identical to Eq. (1), but with Shannon entropies replaced by von Neumann entropies. It shares the desirable property with Eq. (1) of being a concave function of ρ, making it easy to compute [2]. We can alternately write the RHS of Eq. (2) as

$$\max_\rho I(R;B)_\rho, \tag{3}$$

using the definitions

$$\begin{aligned} |\Psi\rangle &= (I^R \otimes \mathcal{N}^{A\to BE}) \big| \Phi_\rho^{RA} \big\rangle \\ I(R;B)_\rho &= I(R;B)_\Psi = H(R)_\Psi + H(B)_\Psi - H(RB)_\Psi \\ &= H(\Psi^R) + H(\Psi^B) - H(\Psi^{RB}). \end{aligned}$$

We will use $I(R;B)_\rho$ and $I(R;B)_\Psi$ interchangeably, since the mutual information and other entropic properties of Ψ are uniquely determined by ρ.

Aside from the constraints $Q \le P \le C \le C_E$, and $Q \le Q_2$, which are obvious consequences of the definitions, and $Q_2 \le Q_E = \frac{1}{2}C_E$, which follows from [74], the five capacities appear to vary rather independently (see for example [10] and [72]). Except in special cases, it is not possible, without knowing the parameters of a channel, to infer any one of these capacities from the other four.

[1] Another powerful assistive resource, unlimited noiseless quantum back-communication from receiver to sender, turns out to be equivalent to unlimited shared entanglement [18]. Thus the capacity of a channel assisted by such back-communication is C_E for classical messages and Q_E for quantum messages.

This complex situation naturally raises the question of how many independent parameters are needed to characterize the important asymptotic, capacity-like properties of a general quantum channel. A full understanding of quantum channels would enable us to calculate not only their capacities, but more generally, for any two channels $\mathcal{M}$ and $\mathcal{N}$, the asymptotic efficiency (possibly zero) with which $\mathcal{M}$ can simulate $\mathcal{N}$, both alone and in the presence of auxiliary resources such as classical communication or shared entanglement.

One motivation for studying communication in the presence of auxiliary resources is that it can simplify the classification of channels' capacities to simulate one another. This is so because if a simulation is possible without the auxiliary resource, then the simulation remains possible with it, though not necessarily vice versa. For example, Q and C represent a channel's asymptotic efficiencies of simulating, respectively, a noiseless qubit channel and a noiseless classical bit channel. In the absence of auxiliary resources these two capacities can vary independently, subject to the constraint $Q \leq C$, but in the presence of unlimited prior entanglement, the relation between them becomes fixed: $C_E = 2Q_E$, because entanglement allows a noiseless 2-bit classical channel to simulate a noiseless 1-qubit channel and vice versa (via teleportation [9] and superdense coding [15]). Similarly the auxiliary resource of shared randomness simplifies the theory of classical channels by allowing channels to simulate one another efficiently according to the classical reverse Shannon theorem.

B. Terminology

The various capacities of a quantum channel $\mathcal{N}$ may be defined within a framework where asymptotic communication resources and conversions between them are treated abstractly [33]. Many independent uses of a noisy channel $\mathcal{N}$, i.e. $\mathcal{N}^{\otimes n}$, corresponds to an asymptotic resource $\langle \mathcal{N} \rangle$, while standard resources such as ebits (maximally-entangled pairs of qubits, also known as EPR pairs), or instances of a noiseless qubit channel from Alice to Bob are denoted $[qq]$ and $[q \to q]$ respectively. Their classical analogues are $[cc]$ and $[c \to c]$, which stand for bits of shared randomness (rbits), and uses of noiseless classical bit channels (cbits). Communication from Bob to Alice is denoted by $[q \leftarrow q]$ and $[c \leftarrow c]$. Within this framework, coding theorems can be thought of as transformations from one communication resource to another, analogous to reductions in complexity theory, but involving resources that are quantitative rather than qualitative, the rate (if other than 1) being indicated by a coefficient preceding the resource expression. We consider two kinds of asymptotic *resource reducibility* or *resource inequality* [33]: viz. asymptotic reducibility via local operations $\leq_{\mathrm{L}}$, usually abbreviated $\leq$, and asymptotic reducibility via clean local operations $\leq_{\mathrm{CL}}$. A resource β is said to be locally asymptotically reducible to α if there is an asymptotically faithful transformation from α to β via local operations: that is, for any $\epsilon, \delta > 0$ and for all sufficiently large n, $n(1+\delta)$ copies of α can be transformed into n copies of β with overall error $< \epsilon = o(1)$. Here, and throughout the paper, we use $o(1)$ to mean a quantity that approaches zero as $n \to \infty$. We use "error" to refer to the trace distance in the context of states, which is defined as

$$\frac{1}{2}\|\rho - \sigma\|_1 = \frac{1}{2}\operatorname{Tr}|\rho - \sigma|.$$

For channels, "error" refers to the diamond norm [60] (see also [69], [63]). The example most studied in this paper is when the target resource $\beta = \langle \mathcal{N} \rangle$ with a channel $\mathcal{N}$. The initial resource α is transformed, via a protocol involving local operations, into a channel $\mathcal{N}'^{(n)}$, with diamond-norm error

$$\|\mathcal{N}^{\otimes n} - \mathcal{N}'^{(n)}\|_\diamond = \max_\rho \left\| \left(\mathrm{id}_R \otimes (\mathcal{N}^{\otimes n} - \mathcal{N}'^{(n)})\right)(\Phi_\rho) \right\|_1,$$

where the maximization is over states ρ on A^n and Φ_ρ is an arbitrary purification of it.

The clean version of this reducibility, $\leq_{\mathrm{CL}}$, which is important when we wish to coherently superpose protocols, adds the restriction that any quantum subsystem discarded during the transformation be in the $|0\rangle$ state up to an error that vanishes in the limit of large n. When $\alpha \leq \beta$ and $\beta \leq \alpha$ we have a resource equivalence, designated $=_{\mathrm{L}}$, or $=$, or for the clean version $=_{\mathrm{CL}}$. Resource reducibilities and equivalences will often be referred to as resource relations or RRs.

For example, the coding theorem for entanglement-assisted classical communication can be stated as

$$\langle \mathcal{N} \rangle + \infty[qq] \geq C_E(\mathcal{N})\ [c \to c]. \tag{4}$$

where $C_E(\mathcal{N})$ is defined as in Eq. (2).

In this language, to simulate (resp. cleanly simulate) a channel $\mathcal{N}$ is to find standard resources α (made up of qubits, ebits, cbits and so on) such that $\langle \mathcal{N} \rangle \leq \alpha$ (resp. $\leq_{\mathrm{CL}}$). For example, the simplest form of the classical reverse Shannon theorem can be stated as $\forall_N \langle N \rangle \leq C(N)[c \to c] + \infty[cc]$, with $C(N)$ defined in Eq. (1).

We will also introduce notation for two refinements of the problem. First, we (still following [33]) define the *relative resource* $\langle \mathcal{N} : \rho \rangle$ as many uses of a channel $\mathcal{N}$ whose asymptotic accuracy is guaranteed or required only when n uses of $\mathcal{N}$ are fed an input of the form $\rho^{\otimes n}$. This means that the error is evaluated with respect to $\Phi_\rho^{\otimes n}$ rather than the worst case entangled input state:

$$\|\mathcal{N}^{\otimes n} - \mathcal{N}'^{(n)}\|_{\rho^{\otimes n}} = \left\| \left(\mathrm{id}_R \otimes (\mathcal{N}^{\otimes n} - \mathcal{N}'^{(n)})\right)(\Phi_\rho^{\otimes n}) \right\|_1.$$

Most coding theorems still apply to relative resources, once we drop the maximization over input distributions. So for a classical channel $\langle N : p \rangle \geq I(X;Y)_p[c \to c]$ and for a quantum channel $\langle \mathcal{N} : \rho \rangle + \infty[qq] \geq I(R;B)_\rho[c \to c]$ (notation following Eq. (2)).

Second, we will consider simulating channels with *passive feedback*. The classical version of a passive feedback channel has Alice obtain a copy of Bob's output $Y = N(X)$. We denote this form of channel by N_F if the original channel is N. For a quantum channel, we cannot give Alice a copy of Bob's output because of the no-cloning theorem [88], but instead define a *coherent feedback* version of the channel as an isometry in which the part of the output that does not go to Bob is retained by Alice, rather than escaping to the environment [87]. We denote this $\mathcal{N}_F^{A \to BE}$, where the

subscript F indicates that E is retained by Alice. When it is clear from the context, we will henceforth use "feedback" to mean conventional passive feedback for a classical channel and coherent feedback for a quantum channel.[2]

Coherent feedback is an example of quantum state redistribution [57], [35], [90] in which the same global pure state Ψ is redistributed among a set of parties. The redistribution corresponding to a feedback channel $\mathcal{N}_F^{A\to BE}$ involves Alice, Bob, and a purifying reference system R. Alice's share A of the initial state $\Psi^{A:R}$, is split into two parts, E and B, with E remaining with her party, while B passes to Bob, who initially held nothing, leading to a final state $\Psi^{E:B:R}$.

Classical and coherent feedback are thus rather different notions, indeed one might say opposite notions, since in coherent feedback Alice gets to keep *everything but* what Bob receives, and as a result coherent feedback is sometimes a stronger resource than free classical back-communication. Despite these differences, there are close parallels in how feedback affects the tradeoff between static resources (rbits, ebits) and dynamic resources (cbits, qubits) required for channel simulation. In both cases, when the static resource is restricted, simulating a non-feedback version of the channel requires less of the dynamic resource than simulating a feedback version, because the non-feedback simulation can be economically split into two sequential stages. For a feedback simulation, no such splitting is possible.

Other notational conventions we adopt are as follows. If $|\psi\rangle$ is a pure state then $\psi := |\psi\rangle\langle\psi|$ and ψ^X refers to the state of the X subsystem of ψ. For a subsystem X, we define $|X|$ to be the cardinality of X if X is classical or $\dim X$ when X is quantum. We take $\log$ and $\exp$ to be base 2. The fidelity [77] between ρ and σ is $\|\sqrt{\rho}\sqrt{\sigma}\|_1$ and the trace distance is $\frac{1}{2}\|\rho-\sigma\|_1$. For a channel $\mathcal{N}^{A\to B}$ we observe that $\mathcal{N} = \mathrm{Tr}_E \circ \mathcal{N}_F$ and we define the *complementary channel* $\hat{\mathcal{N}}^{A\to E} := \mathrm{Tr}_B \circ \mathcal{N}_F$. Since isometric extensions of channels are unique only up to an overall isometry on E, the same is true for the complementary channel [56], and our results will not be affected by this ambiguity.

Additional definitions related to entanglement spread will be introduced in Sec. II-C.

C. Overview of results

In this paper we consider what resources are required to simulate a quantum channel. In particular, one might hope to show, by analogy with the classical reverse Shannon theorem, that $Q_E(\mathcal{N})$ qubits of forward quantum communication, together with a supply of shared ebits, suffice to efficiently simulate any quantum channel $\mathcal{N}$ on any input. This turns out not to be true in general (see below), but it is true in some important special cases:

- When the input is of tensor power form $\rho^{\otimes n}$, for some ρ. In this case, we are simulating the relative resource $\langle\mathcal{N}:\rho\rangle$.
- When the channel $\mathcal{N}$ has the property that its output entropy $H(\mathcal{N}(\rho))$ is uniquely determined by the state of the environment. Such channels include those with classical inputs or outputs.

However, for general channels on general (i.e. non-tensor-power) inputs, we show that efficient simulation requires additional resources beyond ordinary entanglement. Any of the following resources will suffice:

- more general forms of entanglement, such as an entanglement-embezzling state [78], in place of the supply of ordinary ebits, or
- additional communication from Alice to Bob, or
- backward classical or quantum communication, from Bob to Alice.

The quantum reverse Shannon theorem is thus more fastidious than its classical counterpart. While classical shared random bits (rbits) suffice to make all classical channels equivalent and cross-simulable, standard ebits cannot do so for quantum channels. The reason is that quantum channels may require different numbers of ebits to simulate on different inputs. Therefore, to maintain coherence of the simulation across a superposition of inputs, the simulation protocol must avoid leaking to the environment these differences in numbers of ebits used. Fortunately, if the input is of tensor power form $\rho^{\otimes n}$, the entanglement "spread" required is rather small ($O(\sqrt{n})$), so it can be obtained at negligible additional cost by having Alice initially share with Bob a slightly generous number of ebits, then at the end of the protocol return the unused portion for him to destroy. On non-tensor-power inputs the spread may be $O(n)$, so other approaches are needed if one is to avoid bloating the forward communication cost. If the channel itself already leaks complete information about the output entropy to the environment, there is nothing more for the simulation to leak, so the problem becomes moot. Otherwise, there are several ways of coping with a large entanglement spread without excessive forward communication, including: 1) using a more powerful entanglement resource in place of standard ebits, namely a so-called entanglement-embezzling state [78],

$$|\varphi_N\rangle = \frac{1}{\sqrt{\sum_{j=1}^N \frac{1}{j}}} \sum_{j=1}^N \frac{1}{\sqrt{j}} |j\rangle|j\rangle \tag{5}$$

from which (in the limit of large N) a variable amount of entanglement can be siphoned off without leaving evidence of how much was taken, or 2) using a generous supply of standard ebits but supplementing the protocol by additional backward classical communication to coherently "burn off" the unused ebits. We discuss the role of entanglement spread in the quantum reverse Shannon theorem in Sec. II-C. There we will precisely define the resource $[\Subset\Subset]$, which informally

[2]The term "feedback" has been used in multiple ways. Bowen[19] compares several kinds of feedback, both quantum and classical. In his terminology, both the classical and coherent feedbacks we consider here are *passive*, meaning that they do not grant the sender and receiver any additional resource but require them to perform an additional task (e.g. giving the sender a copy of the output) beyond what would have been required in an ordinary execution or simulation of the channel. For this reason passive feedback capacities are never greater than the corresponding plain capacities. *Active* feedback, by contrast, involves granting the sender and receiver an additional resource (e.g. unlimited quantum back-communication, as in [18]), to perform the *same* task as in a plain execution or simulation of the channel. Accordingly, active feedback capacities are never *less* than the corresponding plain capacities. We do not discuss active feedback further in this paper.

can be thought of as an embezzling state $|\varphi_N\rangle$ with N allowed to be arbitrarily large.

When simulating quantum feedback channels, we are sometimes able to establish resource equivalences rather than reducibilities, for example (as we will see in part (a) of Theorem 3)

$$\langle \mathcal{N}_F : \rho \rangle = \frac{1}{2} I(R;B)[q \to q] + \frac{1}{2} I(E;B)[qq]. \qquad (6)$$

This both indicates the numbers of qubits and ebits asymptotically necessary and sufficient to perform the redistribution $\Psi^{A:R} \to \Psi^{E:B:R}$ on tensor powers of a source with density matrix ρ^A, and expresses the fact that any combination of resources asymptotically able to perform the feedback simulation of $\mathcal{N}$ on ρ can be converted into the indicated quantities of qubits and ebits. These results reflect the fact that the state redistribution performed by a quantum feedback channel is asymptotically reversible. One interesting special case is when $\mathcal{N}$ is a noiseless classical channel, in which case Eq. (6) reduces to the "cobit" resource equality [40]. This observation, and our derivation of Eq. (6), are due to [32].

Applications: Our results also have implications for proving rate-distortion theorems and strong converses for the entanglement-assisted capacities. The rate-distortion problem is a variant of the reverse Shannon theorem which differs in that instead of simulating a specific channel with high blockwise fidelity the goal is to minimize an average distortion condition. This is a less stringent condition than demanded by the reverse Shannon theorem, so our simulations imply rate-distortion theorems at the rate one would expect: the least capacity of any channel satisfying the distortion bound. This connection was observed for classical channels in [84] (see also [73]) and for quantum channels in [31]. The second application of our result is to derive a strong converse theorem, meaning that attempting to send classical bits through a quantum channel at rates above C_E results in an exponentially small success probability. We discuss this application further in Sec. IV-E.

Coordination capacity: Another interpretation of reverse Shannon theorems is in terms of "coordination capacities", defined as the minimum rate of communication required to achieve certain correlated probability distributions subject to constraints on some of the variables [29]. For example, the classical reverse Shannon theorem corresponds to the goal of reproducing the input-output distribution of a channel given one party's knowledge of the input. However, the framework of coordination capacity also encompasses many network-coding generalizations of this task.

II. Statement of results

Figure 1 shows the parties, and corresponding random variables or quantum subsystems, involved in the operation of a discrete memoryless classical channel (top left) and a discrete memoryless quantum channel (top right). Dashed arrows indicate additional data flows characterizing a feedback channel. The bottom of the figure gives flow diagrams for simulating such channels using, respectively, a classical encoder and decoder (bottom left) or a quantum encoder and decoder (bottom right). Shared random bits (rbits) and forward classical communication (cbits) are used to simulate the classical channel; shared entanglement (ebits) and forward quantum communication (qubits) are used to simulate the quantum channel. As usual in Shannon theory, the encoder and decoder typically must operate in parallel on multiple inputs in order to simulate multiple channel uses with high efficiency and fidelity.

Where it is clear from context we will often use upper case letters X, B, etc. to denote not only a classical random variable (or quantum subsystem) but also its marginal probability distribution (or density matrix) at the relevant stage of a protocol, for example writing $H(B)$ instead of $H(\rho^B)$. Similarly we write $I(E;B)$ for the quantum mutual information between outputs E and B in the upper right side of Figure 1. However, it is not meaningful to write $I(A;B)$, because subsystems A and B do not exist at the same time. Thus the conventional classical notation $I(X;Y)$ for the input-output mutual information may be considered to refer, in the quantum way of thinking, to the mutual information between Y and a *copy* of X, which could always have been made in the classical setting.

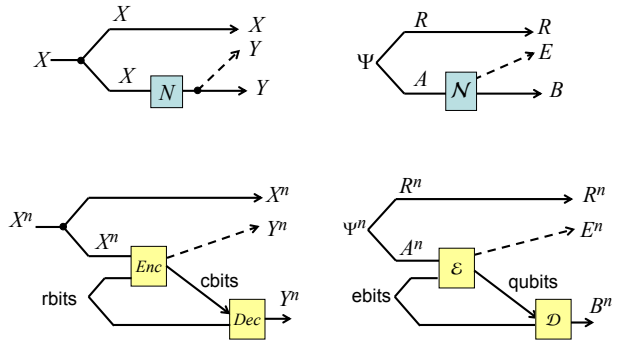

Fig. 1. Parties and subsystems associated with classical and quantum channels (top left and right, resp.) and with their simulation using standard resources (bottom left and right respectively). The dashed lines represent systems that are sent to Alice only in the case of feedback simulations.

Figure 2 shows some of the known results on communications resources required to simulate classical and quantum channels under various conditions.

A. Classical Reverse Shannon Theorem

Most of these results are not new; we collect them here for completeness, and give alternate proofs that will help prepare for the analogous quantum results. The high-shared-randomness and feedback cases below (a,b,e) were proved in [13], [14], [84]. The low- and zero-shared-randomness cases (c,d,f) were demonstrated by Cuff [28] building on Wyner's classic common randomness formula [89]. The connection to rate distortion was first developed in the 1996 Steinberg-Verdú

<table>
<tr><th colspan="2" rowspan="2">Kind of Channel
Kind of Simulation</th><th colspan="2">Classical</th><th colspan="2">Quantum</th></tr>
<tr><th>Classical Feedback</th><th>Non-feedback</th><th>Coherent Feedback</th><th>Non-feedback</th></tr>
<tr><td rowspan="2">Excess shared ebits or rbits</td><td>Tensor-power source</td><td colspan="2">$c = I(X;Y)$
when $r \geq H(Y|X)$</td><td colspan="2">$q = I(R;B)/2$
when $e \geq I(E;B)/2$</td></tr>
<tr><td>General source</td><td colspan="2">$c = C(N) = \max_p I(X;Y)$</td><td colspan="2">$q = Q_E(\mathcal{N}) = \max_\rho I(R;B)/2$
Ordinary ebits insufficient</td></tr>
<tr><td rowspan="2">Limited shared ebits or rbits</td><td>Tensor-power or IID source</td><td>$c(r) = \max\{I(X;Y), H(Y) - r\}$</td><td>$c(r) = \min\{\max($
$I(X;W),$
$I(XY;W) - r)$
$: W$s.t. $I(X;Y\|W) = 0\}$.</td><td>$q(e) = \max\{\frac{1}{2}I(R;B), H(B) - e\}$</td><td>$q(e) = \lim_{n\to\infty} \max$
$\{\frac{1}{2n}I(R;E_B B^n),$
$\frac{1}{n}H(E_B B^n) - e\}$</td></tr>
<tr><td>General source</td><td>$c(r) = \max_X$
$\max\{I(X;Y),$
$H(Y) - r\}$</td><td>$c(r) = \max_X \min_W$
$\{\max(I(X;W),$
$I(XY;W) - r) :$
$I(X;Y\|W) = 0\}$.</td><td colspan="2">Various tradeoffs possible (see text)
Ordinary ebits insufficient</td></tr>
<tr><td rowspan="2">No shared ebits or rbits</td><td>Tensor-power or IID source</td><td>$c = H(Y)$</td><td>$c = \min\{\ I(XY;W)$
$: W$s.t. $I(X;Y\|W) = 0\}$.</td><td>$q = H(B)$
$= H(\mathcal{N}(\rho))$</td><td>$q = \lim_{n\to\infty} \min$
$\{\frac{1}{n}H(\omega) : \exists\omega, \mathcal{N}_1, \mathcal{N}_2$
s.t. $\mathcal{N}_1(\rho^{\otimes n}) = \omega$ &
$\mathcal{N}_2(\omega) = \mathcal{N}(\rho)^{\otimes n}\}$</td></tr>
<tr><td>General source</td><td>$c = \max_X$
$H(Y)$</td><td>$c =$
$\max_X \min_W \{I(XY;W)$
$: I(X;Y\|W) = 0\}$.</td><td>$q = \max_\rho H(B)$
$=$
$\max_\rho H(\mathcal{N}(\rho))$</td><td>$q = \max_\rho \lim_{n\to\infty} \min$
$\{\frac{1}{n}H(\omega) : \exists\omega, \mathcal{N}_1, \mathcal{N}_2$
s.t. $\mathcal{N}_1(\rho^{\otimes n}) = \omega$ &
$\mathcal{N}_2(\omega) = \mathcal{N}(\rho)^{\otimes n}\}$</td></tr>
</table>

Fig. 2. Resource costs of simulating classical and quantum channels: Some known results on the forward communication cost (c=cbits or q=qubits) for simulating classical and quantum channels are tabulated as a function of the kind of source (tensor power or arbitrary), the kind of simulation (feedback or non-feedback), and the quantity of shared random bits (r) or ebits (e) available to assist simulation. For non tensor power quantum sources (green shaded cells), efficient entanglement-assisted simulation is not possible in general using ordinary ebits, because of the problem of entanglement spread. To obtain an efficient simulation in such cases requires additional communication (wlog backward classical communication), or a stronger form of entanglement resource than ordinary ebits, such as an entanglement-embezzling state.

paper [73], which also proved a variant of the high-randomness case.

Theorem 1 (Classical Reverse Shannon Theorem (CRST))**.** *Let N be a discrete memoryless classical channel with input X (a random variable) and induced output $Y = N(X)$. We will use $I(X;Y)$ to indicate the mutual information between input and output. Let N_F denote the feedback version of N, which gives Alice a copy of Bob's output $Y = N(X)$. Trivially $N \leq N_F$ and $\langle N : p\rangle \leq \langle N_F : p\rangle$ for all input distributions p.*

(a) Feedback simulation on known sources with sufficient shared randomness to minimize communication cost:

$$\langle N_F : p\rangle \leq I(X;Y)[c \to c] + H(Y|X)[cc]. \quad (7)$$

In fact this is tight up to the trivial reduction $[cc] \leq [c \to c]$. In other words, for c and r nonnegative,

$$\langle N_F : p\rangle \leq c[c \to c] + r[cc] \quad (8)$$

iff $c \geq I(X;Y)$ and $c + r \geq H(Y)$.

(b) Feedback simulation on general sources with sufficient shared randomness to minimize communication cost:

$$\langle N_F\rangle \leq C(N)[c \to c] + (\max_p H(Y) - C(N))[cc]. \quad (9)$$

(c) Non-feedback simulation on known sources, with limited shared randomness: *When shared randomness is present in abundance, feedback simulation requires no more communication than ordinary non-feedback simulation, but when only limited shared randomness is available, the communication cost of non-feedback simulation can be less.*

$$\langle N : X\rangle \leq c[c \to c] + r[cc] \quad (10)$$

if and only if there exists a random variable W with $I(X;Y|W) = 0$, such that $c \geq I(X;W)$ and $c + r \geq I(XY;W)$.

(d) Non-feedback simulation on known sources with no shared randomness: *A special case of case (c) is the fact that*

$$\langle N : p\rangle \leq c[c \to c] \quad (11)$$

if and only if there exists W such that $I(X;Y|W) = 0$ and $c \geq I(XY;W)$.

(e) Feedback simulation on arbitrary sources, with arbitrary shared randomness: *For non-negative r and c,*

$$\langle N_F \rangle \leq c[c \to c] + r[cc] \tag{12}$$

iff $c \geq C(N) = \max_p I(X;Y)$ and $r \geq \max_p H(Y) - \max_p I(X;Y)$. Because the two maxima may be achieved for different p the last condition is not simply $r \geq H(Y|X)$.

(f) Without feedback *we have, for non-negative r and c,*

$$\langle N \rangle \leq c[c \to c] + r[cc] \tag{13}$$

if and only if for all X there exists W with $I(X;Y|W) = 0$, such that $c \geq I(X;W)$ and $c + r \geq I(XY;W)$.

Parts (b,e,f) of the theorem reflect the fact that the cost of a channel simulation depends only on the empirical distribution or type class of the input[3], which can be communicated in at asymptotically negligible cost ($O(\log n)$ bits), and that an i.i.d. source p is very likely to output a type p' with $\|p - p'\|_1 \sim 1/\sqrt{n}$. Also note that in general the resource reducibility Eq. (12) is not a resource equivalence because $H(Y)$ and $I(X;Y)$ may achieve their maxima on different X.

Part (c), and the low-randomness simulations in general, are based on the possibility of splitting the simulation into two stages with the second performed by Bob, and part of the first stage's randomness being recycled or derandomized[4]. Since Alice does not get to see the output of the second stage, this is a non-feedback simulation. Indeed, part (a) implies that non-trivial cbit-rbit tradeoffs are only possible for non-feedback simulations.

Fig. 3 and Fig. 4 schematically illustrate the form of the cbit-rbit tradeoffs. For feedback simulation on a fixed source, the tradeoff between communication and shared randomness is trivial: Beginning at the point $c = I(X;Y), r = H(Y|X)$ on the right, r can only be decreased by the same amount as c is increased, so that $c = H(Y)$ when $r = 0$. By contrast, if the simulation is not required to provide feedback to the sender, a generally nontrivial tradeoff results, for which the amount of communication at $r = 0$ is given by Wyner's common information expression $\min\{I(XY;W) : I(X;Y|W) = 0\}$. This is evident in Fig. 5 showing the tradeoff for non-feedback simulation of the classical binary erasure channel for several values of the erasure probability t. This figure also shows that for some channels (in particular for erasure channels with $t > 0.5$), even the non-feedback tradeoff begins with a -45 degree straight line section at low r values.

The converse to (a) follows from Shannon's original noisy channel coding theorem, which states that $\langle N : p \rangle \geq I(X;Y)_p[c \to c]$. A slight refinement [3], [4] implies that $\langle N_F : p \rangle \geq I(X;Y)_p[c \to c] + H(Y|X)_p[cc]$.

Thus we have the following resource equivalences.

[3]Types are defined and reviewed in Sec. III-C.

[4]Here, "recycled" means that using a sublinear amount of additional randomness, privacy amplification can be used to make the shared randomness approximately independent of the output of the first stage. Our proof (in Sec. III) will instead use the somewhat simpler "derandomization" approach in which we argue that some of the random bits in the X–W stage can be set in a way that works for all input strings x^n simultaneously.

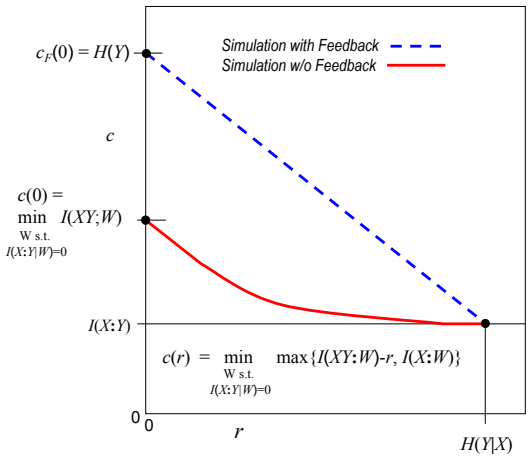

Fig. 3. Classical communication c versus shared randomness r tradeoff for feedback and non-feedback simulations of a classical channel on a specified source p (Theorem 1).

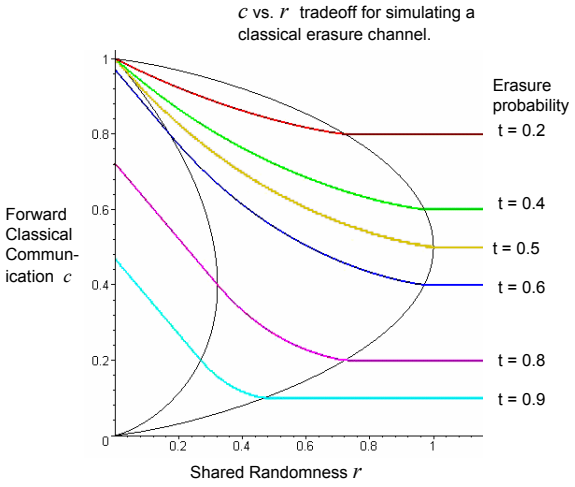

Fig. 5. Classical communication c vs shared randomness r tradeoff for non-feedback simulation of binary erasure channels with erasure probabilities $t = 0.2,\ 0.4,\ 0.5,\ 0.6,\ 0.8$ and 0.9 (colored graphs). One can show that in Theorem 1 part (f) it is enough to consider W such that both legs $X \to W$ and $W \to Y$ are erasure channels. The two black curves mark the boundaries of the region where the tradeoff has slope -1, viz. $r \leq H_2(c/2) - c$, and where it is horizontal, $r \geq H_2(c)$. Note that for $t \leq \frac{1}{2}$, Wyner's quantity $c(0) = 1$, and that for these channels the tradeoff graphs have no section of slope -1. These tradeoff curves were first given in [28].

Corollary 2.

$$\langle N_F : p \rangle = I(X;Y)[c \to c] + H(Y|X)[cc] \tag{14}$$

$$\begin{aligned} \langle N_F : p \rangle + \infty[cc] &= \langle N : p \rangle + \infty[cc] \\ &= I(X;Y)[c \to c] + \infty[cc] \end{aligned} \tag{15}$$

$$\begin{aligned} \langle N_F \rangle + \infty[cc] &= \langle N \rangle + \infty[cc] \\ &= (\max_p I(X;Y))[c \to c] + \infty[cc] \end{aligned} \tag{16}$$

Remark: The task considered in case (d) above, of simulating a channel on a known source by forward communication alone without shared randomness, is a variant of the problem originally considered by Wyner [89], who sought the minimum

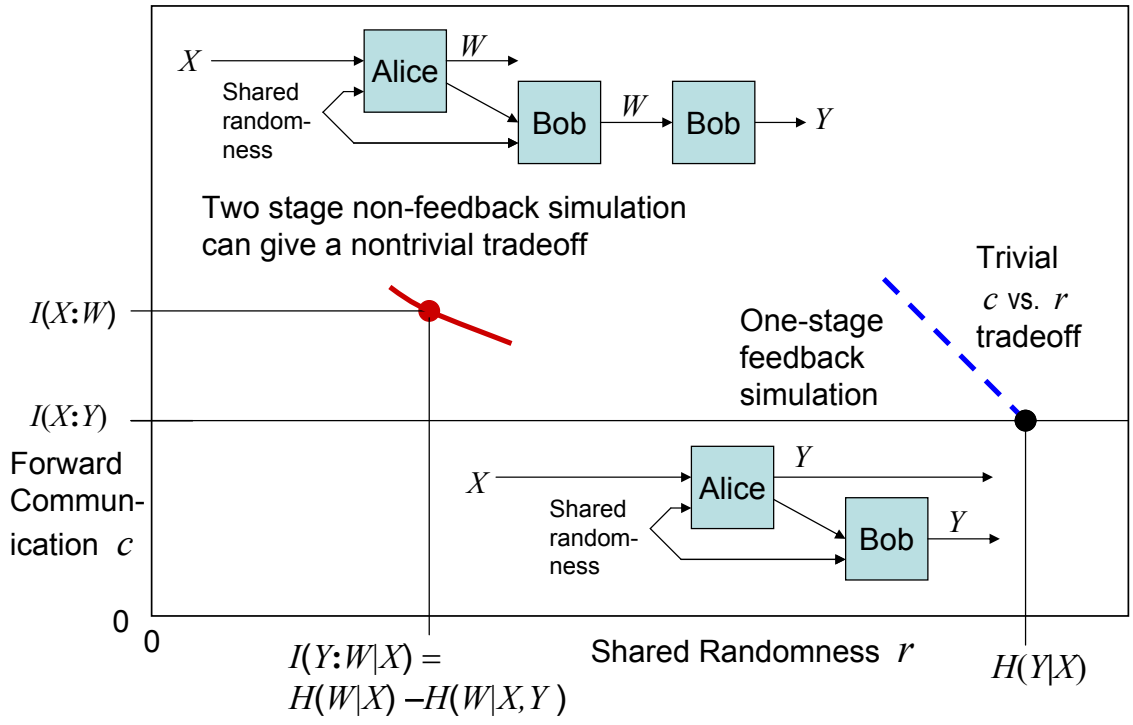

Fig. 4. Two-stage non-feedback simulation of a classical channel, via a Markov chain $X \to W \to Y$ allows a nontrivial tradeoff between forward communication c and shared randomness r. A typical point on the optimal tradeoff curve is shown with $c = I(X:W)$ and $r = I(Y:W|X)$, and with a segment of the optimal tradeoff curve depicted. The second term, $H(W|XY)$, in the expression for r represents the portion of the shared randomness in the first stage simulation of $X \to W$ that can be recycled or derandomized. On the right side is also depicted the "full randomness" solution consisting of a one-stage feedback simulation that uses communication $I(X:Y)$ and randomness $H(Y|X)$. Since cbits can always be traded for rbits, this yields the upper bound depicted by the 45-degree dashed line coming out of this point.

rate of a source allowing two correlated random variables X and Y to be generated from it by separate decoders. He called this the common information between X and Y, and showed it was given by $\min\{I(XY;W) : I(X;Y|W) = 0\}$.

B. Quantum Reverse Shannon Theorem (QRST)

Theorem 3 (Quantum Reverse Shannon Theorem)**.** *Let $\mathcal{N}$ be a quantum channel from $A \to B$ or equivalently an isometry from $A \to BE$ and $\mathcal{N}_F$ the feedback channel that results from giving system E to Alice. If we are given an input density matrix ρ^A then entropic quantities such as $I(R;B)$ or $I(R;B)_\rho$ refer to the state $\Psi^{RBE} = (I^R \otimes \mathcal{N}^{A\to BE})(\Phi_\rho^{RA})$, where Φ_ρ is any state satisfying $\Phi_\rho^A = \rho$.*

(a) Feedback simulation on known tensor power input, with sufficient ebits of entanglement to minimize the forward qubit communication cost:

$$\forall_\rho \ \langle \mathcal{N}_F : \rho \rangle = \ \tfrac{1}{2} I(R;B)_\rho[q \to q] + \tfrac{1}{2} I(E;B)_\rho[qq]. \tag{17}$$

In view of the trivial tradeoff between ebits and qubits for simulating a feedback channel, this implies that the qubit communication rate necessary and sufficient for feedback simulation of a channel on a tensor power source using ordinary entanglement at the rate e ebits per channel use is

$$q_F(e) = \max\{\tfrac{1}{2} I(R;B), H(B) - e\}. \tag{18}$$

(b) Known tensor power input, non-feedback simulation, entanglement possibly insufficient to minimize the forward communication cost:

$$\langle \mathcal{N} : \rho \rangle \leq \ q[q \to q] + e[qq], \tag{19}$$

if and only if for all $\delta > 0$ there exists an $n > 0$ and an isometry $V : E^n \to E_A E_B$ such that

$$q \geq \frac{1}{n} \cdot \tfrac{1}{2} I(R^n; B^n E_B)_\Psi - \delta \text{ and} \tag{20}$$

$$q + e \geq \frac{1}{n} H(B^n E_B)_\Psi - \delta \text{ where} \tag{21}$$

$$|\Psi\rangle^{R^n B^n E_A E_B} := V^{E^n \to E_A E_B} \mathcal{N}_F^{\otimes n} |\Phi_\rho\rangle^{\otimes n}. \tag{22}$$

Thus the communication cost for non-feedback simulation on a tensor power source, as a function of e, is given by

$$q(e) = \liminf_{n\to\infty, \exists V: E^n \to E_A, E_B} \max\{\tfrac{1}{2} I(R^n; B^n E_B)/n, H(B^n E_B)/n - e\}. \quad (23)$$

(c) Known tensor power input, non-feedback, no entanglement: *This is obtained from setting $e = 0$ in case (b) above. In this case, Eq. (20) is always dominated by Eq. (21) and we have that*

$$\langle \mathcal{N} : \rho \rangle \leq q[q \to q], \quad (24)$$

iff $q \geq \lim_{n\to\infty} \frac{1}{n} \min_V H(B^n E_B)$, where the minimum is over isometries $V : E^n \to E_A E_B$. The latter is a well-known quantity: it is the regularized entanglement of purification (EoP) [75] $E_P^\infty(\Psi^{RB}) = \lim_{n\to\infty} \frac{1}{n} E_P((\Psi^{RB})^{\otimes n})$ of the channel's Choi-Jamiołkowski state Ψ.

(d) Arbitrary input, feedback simulation: *For a communication resource α in the sense of [33] comprising any combination of ebits, embezzling states $[\in\in]$, backward cbits $[c \leftarrow c]$, and/or forward or backward quantum communication,*

$$\alpha \geq \langle \mathcal{N}_F \rangle \quad (25)$$

iff there exists a resource β such that for all ρ,

$$\alpha \geq_{CL} \langle \mathcal{N}_F : \rho \rangle + \beta. \quad (26)$$

Specifically, using embezzling states we have

$$\langle \mathcal{N}_F \rangle \leq Q_E(\mathcal{N})[q \to q] + [\in\in] \quad (27)$$

and when considering back communication

$$\langle \mathcal{N}_F \rangle \leq Q_E(\mathcal{N})[q \to q] + C[c \leftarrow c] + (\max_\rho H(B)_\rho - Q_E(\mathcal{N}))[qq] \quad (28)$$

iff $C \geq \max_\rho H(B)_\rho - \min_\rho H(B|R)_\rho - C_E(\mathcal{N})$. Other examples are discussed in Sec. II-C.

(e) Arbitrary input, no feedback: *This case combines elements of cases (b) and (d), although we now consider only fully coherent input resources. If α is a combination of ebits, embezzling states $[\in\in]$ and forward and/or backward qubits, then $\alpha \geq \langle \mathcal{N} \rangle$ iff for all $\delta > 0$ there exists an $n > 0$, a resource β_n and an isometry $V_n : E^n \to E_A E_B$ such that*

$$\alpha \geq_{CL} \frac{1}{n} \langle V_n \circ \mathcal{N}_F^{\otimes n} \rangle + \beta_n. \quad (29)$$

Part (a) of Theorem 3 can equivalently be stated as

$$\langle \mathcal{N}_F : \rho \rangle = I(R; B)_\rho [q \to qq] + H(B|R)_\rho [qq], \quad (30)$$

where $[q \to qq]$ denotes a co-bit [40], [33], which is equivalent to $([q \to q] + [qq])/2$. The formulation in Eq. (30) is parallel to the classical version in Eq. (14) if we replace quantum feedback with classical feedback, co-bits with cbits and ebits with rbits.

A weaker version of (a) was proven in a long unpublished and now obsolete version of the present paper. The idea there was to simulate the channel using a noisy form of teleportation, and then to use measurement compression [85][5]. The full statement of (a) has since been proved by Devetak [32] using his triangle of dualities among protocols in the "family tree" – see also [33]; by Horodecki *et al.* [57] as the inverse of the "mother" protocol, a coherent version of state merging; and by Abeyesinghe *et al.* [1] in the context of a direct derivation of the "mother" protocol. We will present another proof of (a) in Sec. IV, partly in order to prepare for the proof of the rest of Theorem 3.

To prove (b), we argue that any protocol using only qubits and ebits for a non-feedback simulation of $\mathcal{N}^{\otimes n}$ is equivalent to one that performs a feedback simulation of $V^{E^n \to E_A E_B} \circ \mathcal{N}_F^{\otimes n}$. The argument is that the resources used (qubits and ebits) leak nothing to the environment, so the only non-unitary elements are those that are deliberately introduced by Alice and Bob. Thus, we can replace any non-unitary operation by an isometry that instead sends the system to be discarded to a local "environment", labeled E_A for Alice and E_B for Bob. By Uhlmann's theorem and the fact that any two purifications are related by an isometry, it follows that if our original simulation had fidelity $1 - \epsilon$ with the action of $\mathcal{N}^{\otimes n}$, then this modified simulation has fidelity $1 - \epsilon$ with $V^{E^n \to E_A E_B} \circ \mathcal{N}_F^{\otimes n}$ for some isometry V. This is an equivalence, since this procedure turns any simulation of $\mathcal{N}^{\otimes n}$ into a method of simulating $V^{E^n \to E_A E_B} \circ \mathcal{N}_F^{\otimes n}$ for an isometry V, and the reverse direction is achieved simply by discarding the E_A, E_B systems.

Part (c) is simply a special case of (b), and was proven in the case when $\mathcal{N}$ is a CQ channel (that is, has classical inputs) by Hayashi [46]. It corresponds to the regularized entanglement of purification [75] of $|\Psi\rangle$. In both cases, the additivity problem (i.e. the question of whether regularization is necessary) is open, although recent evidence suggests strongly that the entanglement of purification is *not* additive [20] and thus that it is not a single-letter formula for the simulation cost.

Proving, and indeed understanding, parts (d) and (e) will require the concept of entanglement spread, which we will introduce in Sec. II-C. At first glance, the statements of the theorem may appear unsatisfying in that they reduce the question of whether $\langle \mathcal{N} \rangle \leq \alpha$ or $\langle \mathcal{N}_F \rangle \leq \alpha$ to the question of whether certain other clean resource reductions hold. However, according to part (a) of Theorem 3, the corresponding clean resource reductions involve the standard resources of qubits and ebits. As we will explain further in Sec. II-C, this will allow us to quickly derive statements such as Eq. (27) and

[5]More concretely, suppose that Alice uses the "Homer Simpson protocol," which means applying $\mathcal{N}$ to her input and then teleporting the output to Bob, using a classical message of size $2 \log d_A$. Alice's entire part of the protocol can be viewed as a measurement that she performs on her input state and on half of a maximally entangled state. The mutual information between her classical message and Bob's residual quantum state is given by $I(R; B)$. Therefore [85] can be used to simulate n applications of this measurement by a block measurement with $\approx \exp(nI(R; B))$ outcomes. Finally, it is necessary to observe that the error analysis in [85] shows that the simulated measurement not only has the correct output statistics, but essentially has the correct Kraus operators. Thus the compressed measurement gives a high-fidelity simulation of the Homer Simpson protocol, and thus of the original channel. However, the measurement compression step relies on knowledge of the input density matrix ρ, and so new ideas are necessary for the non-tensor-power case.

Eq. (28). An alternate proof[6] of the QRST for general sources using embezzling states as the entanglement resource, Eq. (27), was given by Berta, Christandl, and Renner [16].

The situation in part (d) when embezzling states are not present (i.e. general input, unlimited ebits, and some combination of forward quantum communication and backwards quantum and classical communication) is somewhat surprising in that the simulation requires an asymptotically greater rate of communication than the communication capacity of the channel. To capture this gap, we introduce the following definition.

Definition 4. *The* spread deficit *of a channel $\mathcal{N}$ is defined as*

$$\Delta_{\text{sim}}(\mathcal{N}) := \max_{\rho} H(B)_\rho - \min_{\sigma} H(B|R)_\sigma - C_E(\mathcal{N}). \quad (31)$$

Thus, we could equivalently say that the resource inequality in Eq. (28) holds iff $C \geq \Delta_{\text{sim}}(\mathcal{N})$.

It is important to note that the maximization of $H(B)$ and the minimization of $H(B|R)$ on the RHS of Eq. (31) are taken separately. Indeed, $C_E(\mathcal{N})$ is simply the maximization of $H(B)_\rho - H(B|R)_\rho$ over all ρ, so Eq. (31) expresses how much larger this expression can be by breaking up the optimization of those two terms.

Fortunately, each term in the RHS of Eq. (31) is additive, so there is no need to take the limit over many channel uses. The additivity of $H(B)_\rho$ follows immediately from the subadditivity of the von Neumann entropy, or equivalently the nonnegativity of the quantum mutual information. The other two terms have already been proven to be additive in previous work: [34] showed that $\min H(B|R) = -\max H(B|E)$ is additive and [2] showed that C_E is additive. Thus, we again obtain a single-letter formula in the case of unlimited ebits.

The fact that $\Delta_{\text{sim}}(\mathcal{N})$ provides a single-letter characterization involving convex optimizations makes it possible to explicitly and efficiently evaluate it. For the important class of so-called covariant channels, such as the depolarizing and erasure channels, entropic quantities are invariant under unitary rotation of the inputs. In this case, $H(B)$, $H(R) - H(E)$ and $I(R;B)$ are all simultaneously maximized for the maximally-mixed input and $\Delta_{\text{sim}}(\mathcal{N}) = 0$. However, channels that lack this symmetry will generally have nonzero Δ_{sim}. As an example, we plot the entanglement-assisted capacity $C_E(\mathcal{N})$ against the ebit-assisted simulation cost $C_E(\mathcal{N}) + \Delta_{\text{sim}}(\mathcal{N})$ for the amplitude-damping channel in Fig. 6.

For the amplitude damping channel, the spread deficit is comparable to the other costs of the simulation. But there exist other channels for which the spread deficit can dominate the cost of the channel simulation. Here is an example: for any d, define the "variable-entropy" channel $\mathcal{M}_d$ mapping 2 dimensions to $d+1$ dimensions as follows: it measures the input, and upon outcome 0, outputs $|0\rangle\langle 0|$, and upon outcome 1, outputs $\frac{1}{d}\sum_{i=1}^{d} |i\rangle\langle i|$. For any d, $C_E(\mathcal{M}_d) = 1$, but $\Delta_{\text{sim}}(\mathcal{M}_d) = \log(d+1) - 1$, which is asymptotically larger as d grows[7]. Thus, when performing a feedback simulation of $\mathcal{M}_d$ using cbits and ebits, nearly all of the communication cost comes from the need to create entanglement spread (discussed further in Sec. II-C).

What about non-feedback simulations? In this case, it turns out that the variable-entropy and amplitude-damping channels can both be simulated at the communication rate given by C_E. We will discuss this in more detail in Sec. II-C, but the intuitive reason for this is that non-feedback simulations allow us to damage the environment of the channel and in particular to measure it. This can result in collapsing superpositions between different amounts of entanglement, thus reducing the contribution of entanglement spread. However, there remain channels whose simulation cost with ebits is higher than with embezzling states or back communication even for non-feedback simulation; we describe an example (the "Clueless Eve" channel) in Sec. IV-E3.

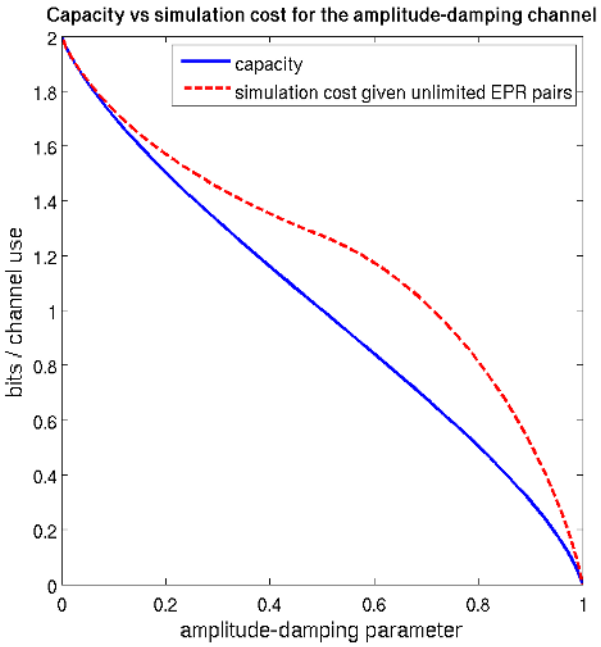

Fig. 6. The amplitude damping channel with parameter γ has Kraus operators $|0\rangle\langle 0| + \sqrt{1-\gamma}|1\rangle\langle 1|$ and $\sqrt{\gamma}|0\rangle\langle 1|$. The lower, solid, curve is the entanglement-assisted classical capacity of the amplitude-damping channel, or equivalently the (w.l.o.g. feedback) simulation cost in cbits when back communication or embezzling states are given, or when the source is a tensor power. The upper, dashed, curve is the feedback simulation cost in cbits (calculated using Eq. (31)) when instead unlimited ebits are given. The gap between the two curves is the spread deficit from Definition 4, and illustrates the extra communication cost of producing entanglement spread.

The proofs of parts (d) and (e) will be given in Sec. IV. To prove them, we restrict attention to the case when α

[6]This proof was developed in parallel with ours (cf discussion in [22]) and differs primarily by describing merging in terms of one-shot entropies (compared with our applying merging only to "flat" spectra) and by reducing to the tensor-power case using the post-selection principle of [22] (compared with our use of Schur duality to reduce to the flat case). Note that the post-selection principle can also be thought of in terms of the Schur basis as the statement that tensor-power states are "almost flat" in a certain sense (cf. [47]).

[7]Proof: $\mathcal{M}_d$ can be perfectly simulated with a single bit of forward classical communication, which proves that $C_E(\mathcal{M}_d) \leq 1$, while the obvious protocol for sending one classical bit through the channel proves that $C_E(\mathcal{M}_d) \geq 1$. To evaluate $\Delta_{\text{sim}}(\mathcal{M}_d)$, observe that $H(B)$ achieves its maximal value of $\log(d+1)$ upon input $\frac{1}{d+1}|0\rangle\langle 0| + \frac{d}{d+1}|1\rangle\langle 1|$, while $H(B|R)$ can be zero if the input $|0\rangle\langle 0|$ is given.

is a combination of entanglement-embezzling states and/or "standard" resources (qubits, cbits and ebits). However, for part (e), we need to further restrict our claim to exclude cbits, for reasons related to the fact that we do not know the tradeoff curve between quantum and classical communication when simulating classical channels.

Remark: Analogously to the low-shared randomness regime in classical channel simulation (Figure 4 and cases (c) and (d) of the CRST), simulating a non-feedback channel permits a nontrivial tradeoff between ebits and qubits, in contrast to the trivial tradeoff for feedback simulation. While the cbit-rbit tradeoff curve for simulating classical channels is additive and given by a single-letter formula [89], [28], no such formula or additivity result is known for the qubit cost in the zero- and low-entanglement regime.

Remark: Interestingly, quantum communication or entanglement can sometimes improve simulations of even classical channels. In [86] an example of a classical channel is given with d-dimensional inputs which requires $\Omega(\log d)$ classical bits to simulate, but can be simulated quantumly using $O(d^{-1/3})$ qubits of communication, asymptotically. Curiously, the classical reverse Shannon theorem (Theorem 1) is only a special case of the quantum reverse Shannon theorem (Theorem 3) when in the unlimited shared entanglement regime; one of the problems left open by this work is to understand how entanglement can be more efficient than shared randomness in creating correlated classical probability distributions. More generally, which values of c, q, r, e are consistent with the reducibility $\langle\mathcal{N}\rangle \leq c[c \to c] + q[q \to q] + r[cc] + e[qq]$? We know how to convert this problem to the equivalent relative resource problem with $\langle\mathcal{N}\rangle$ replaced with $\langle\mathcal{N} : \rho\rangle$, but this in turn we do not have an answer for.

Remark: Our results imply unbounded gaps (for growing dimension) between the costs of simulating channels when (a) no entanglement is given, (b) a linear or unlimited rate of ebits are given, and (c) stronger forms of entanglement, such as embezzling states, are given. An example of a large gap between (a) and (b) is given by the Werner-Holevo channel [79], defined on d-dimensional inputs to be $\mathcal{N}(\rho) = ((\operatorname{Tr}\rho)I - \rho^T)/(d-1)$. This channel has $C_E(\mathcal{N}) \lessapprox 1$, but when acting on half of a maximally entangled state produces a state with entanglement of purification equal to $\log d$ [24]. Thus, the gap between the ebit-assisted simulation cost and the unassisted simulation cost grows with dimension. For an asymptotically growing gap between (b) and (c), we give an example in Sec. IV-E3.

C. Entanglement spread

To understand parts (d) and (e) of Theorem 3, we need to introduce the idea of entanglement spread. This concept is further explored in [42], [52], but we review some of the key ideas here.

If Alice's input is known to be of i.i.d. form $\rho^{\otimes n}$ then we know that the channel simulation can be done using $\frac{1}{2}I(R;B)[q \to q] + \frac{1}{2}I(B;E)[qq]$. To see the complications that arise from a general input, it suffices to consider the case when Alice's input is of the form $(\rho_1^{\otimes n} + \rho_2^{\otimes n})/2$. We omit explicitly describing the reference system, but assume that Alice's input is always purified by some reference and that the fidelity of any simulation is with respect to this purification.

Assume that $\rho_1^{\otimes n}$ and $\rho_2^{\otimes n}$ are nearly perfectly distinguishable and that the channel simulation should not break the coherence between these two states. Naively, we might imagine that Alice could first determine whether she holds $\rho_1^{\otimes n}$ or $\rho_2^{\otimes n}$ and coherently store this in a register $i \in \{1, 2\}$. Next she could conditionally perform the protocol for i.i.d. inputs that uses $\frac{1}{2}I(R;B)_{\rho_i}[q \to q] + \frac{1}{2}I(B;E)_{\rho_i}[qq]$. To use a variable amount of communication, it suffices to be given the resource $\max_i \frac{1}{2}I(A;B)_{\rho_i}[q \to q]$, and to send $|0\rangle$ states when we have excess channel uses. But unwanted entanglement cannot in general be thrown away so easily. Suppose that $I(B;E)_{\rho_1} > I(B;E)_{\rho_2}$, so that simulating the channel on $\rho_1^{\otimes n}$ requires a higher rate of entanglement consumption than $\rho_2^{\otimes n}$. Then it is not possible to start with $\frac{1}{2}nI(B;E)_{\rho_1}$ (or indeed any number) of ebits and perform local operations to obtain a superposition of $\frac{1}{2}nI(B;E)_{\rho_1}$ ebits and $\frac{1}{2}nI(B;E)_{\rho_2}$ pairs.

The general task we need to accomplish is to coherently create a superposition of different amounts of entanglement. Often it is convenient to think about such superpositions as containing a small "control" register that describe how many ebits are in the rest of the state. For example, consider the state

$$|\psi\rangle = \sum_{i=1}^{m} \sqrt{p_i}|i\rangle^A |i\rangle^B |\Phi\rangle^{\otimes n_i} |00\rangle^{\otimes N - n_i}, \tag{32}$$

where $0 \leq n_i \leq N$ for each i. Crudely speaking[8], we say that $\max_i n_i - \min_i n_i$ is the amount of entanglement spread in the state $|\psi\rangle$, where the $\max$ and $\min$ are taken over values of i for which p_i is nonnegligible.

A more precise and general way to define entanglement spread for any bipartite state $|\psi\rangle$ is (following [52]) as $\Delta(\psi^A) = H_0(\psi^A) - H_\infty(\psi^A)$, where $H_0(\rho) = \log \operatorname{rank}\rho$ and $H_\infty(\rho) = -\log\|\rho\|_\infty$. (The quantities H_0 and H_∞ are also known as $H_{\max}$ and $H_{\min}$ respectively. Alternatively, they can be interpreted as Rényi entropies.) Ref. [52] also defined an ϵ-smoothed version of entanglement spread by

$$\Delta_\epsilon(\rho) = \min\{\Delta(\sigma) : 0 \leq \sigma \leq \rho, \operatorname{Tr}\sigma \geq 1 - \epsilon\}$$

that reflects the communication cost of approximately preparing $|\psi\rangle$. More precisely, we have

Theorem 5 (Theorem 8 of [52])**.** *If $|\psi\rangle$ can be created from ebits using C cbits of communication and error $\leq \epsilon = \delta^8/4$, then*

$$C \geq \Delta_\delta(\psi^A) + 3\log(1-\delta) \tag{33}$$

The factor of 3 in Eq. (33) is because the definition of Δ_ϵ we have used is actually the alternate version used in Remark 4 of [52]. We can similarly define $H_{0,\epsilon}(\rho) := \log\min\{\operatorname{rank}\sigma : 0 \leq \sigma \leq \rho, \operatorname{Tr}\sigma \geq 1-\epsilon\}$ and $H_{\infty,\epsilon}(\rho) := -\log\min\{\|\sigma\|_\infty : 0 \leq \sigma \leq \rho, \operatorname{Tr}\sigma \geq 1-\epsilon\}$. Our definition of $H_{0,\epsilon}$ is the same as the one used in [52], but our definition of $H_{\infty,\epsilon}$ may be as

[8] This neglects the entanglement in the $|ii\rangle$ register. However, in typical applications, this will be logarithmic in the total amount of entanglement.

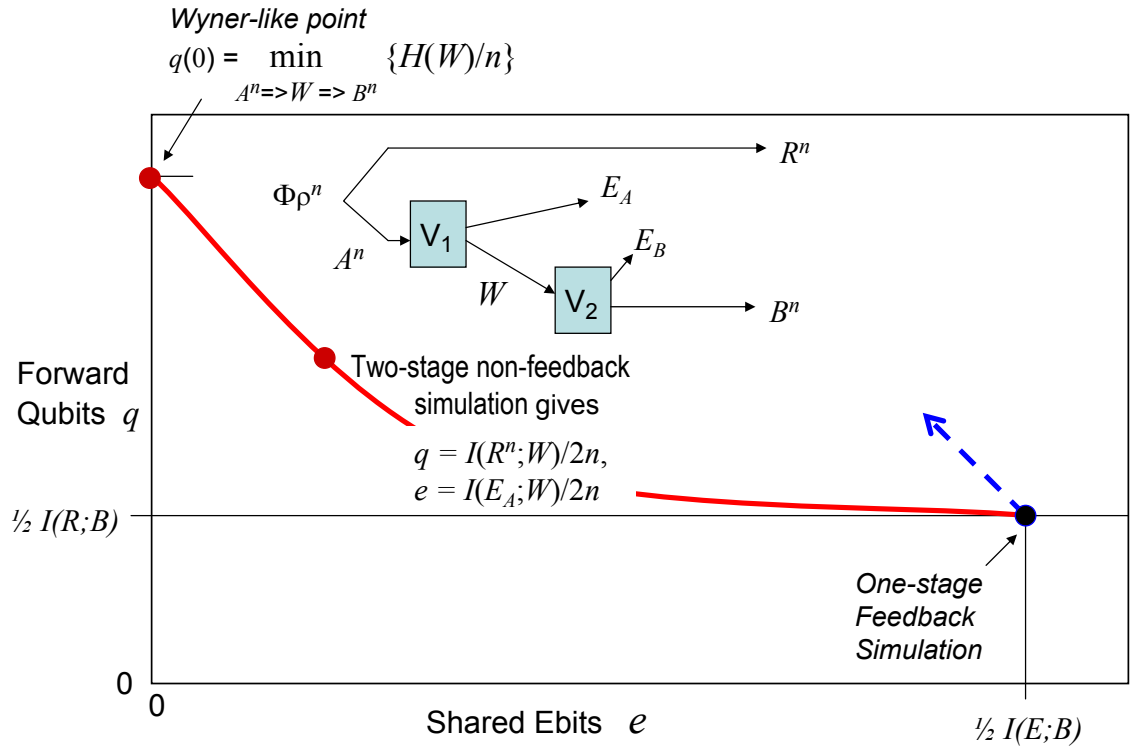

Fig. 7. Two-stage non-feedback simulation of a quantum channel (solid red curve) on a specified input ρ, via an intermediate state W, makes possible a nontrivial tradeoff between forward communication q and shared entanglement e. By contrast, for a feedback simulation (right, dashed blue curve) only a trivial tradeoff is possible, where any deficit in ebits below the $\frac{1}{2}I(E;B)$ needed for optimal simulation must be compensated by an equal increase in the number of qubits used.

much as $-\log(1-\epsilon)$ smaller. As with the definitions in [52], our quantities trivially satisfy

$$\Delta_\epsilon(\rho) \geq H_{0,\epsilon}(\rho) - H_{\infty,\epsilon}(\rho). \tag{34}$$

Our quantities can also be expressed as

$$H_{0,\epsilon}(\rho) = \min_M H_0(\sqrt{M}\rho\sqrt{M}) \tag{35a}$$

$$H_{\infty,\epsilon}(\rho) = \max_M H_\infty(\sqrt{M}\rho\sqrt{M}), \tag{35b}$$

where in each case M must satisfy $0 \leq M \leq I$ and $\operatorname{Tr} M\rho \geq 1-\epsilon$. In fact, we can WLOG assume that M commutes with ρ and $\operatorname{Tr} M\rho = 1-\epsilon$. Similarly, in the definitions that optimized over $\sigma \leq \rho$, we can assume that ρ and σ commute.

One advantage of this version of $\Delta_\epsilon(\rho)$ is that it has the following natural interpretation as a minimization over nearby *normalized* states.

Lemma 6.

$$\max(0, \Delta_\epsilon(\rho)) = \min\{\Delta_0(\sigma) : \frac{1}{2}\|\rho - \sigma\|_1 \leq \epsilon, 0 \leq \sigma, \operatorname{Tr}\sigma = 1\} \tag{36}$$

The lemma is proved in the appendix. It improves upon Lemma 5 of [52], and could thus be used to tighten Theorem 5, although we do not carry out that exercise here.

There are a few different ways of producing entanglement spread, which are summarized in [42]. For example, one cbit can be used to coherently eliminate one ebit, or to do nothing; and since both of these tasks can be run in superposition, this can also be used to create entanglement spread. Likewise one qubit can coherently either create or disentangle one ebit. To put this on a formal footing, we use the *clean resource reducibility* $\leq_{\text{CL}}$ (called $\overset{\text{clean}}{\leq}$ in [42]). A resource β is said to be "cleanly LO-reducible" to α iff there is an asymptotically faithful clean transformation from α to β via local operations: that is, for any $\epsilon, \delta > 0$ and for all sufficiently large n, $n(1+\delta)$ copies of α can be transformed by local operations into n copies of β with overall diamond-norm error $\leq \epsilon$, and moreover, any quantum subsystem discarded during the transformation is in a standard $|0\rangle$ state, up to an error vanishing in the limit of large n. In particular, entangled states cannot be discarded. This restriction on discarding states means that clean protocols can be safely run in superposition.

Finally, we can define the clean entanglement capacity of a

resource α to be the set $E_{\text{clean}}(\alpha) = \{E : \alpha \geq_{\text{CL}} E[qq]\} \subseteq \mathbb{R}$. Negative values of E correspond to the ability to coherently eliminate entanglement. By time-sharing, we see that $E_{\text{clean}}(\alpha)$ is a convex set. However, it will typically be bounded both from above and below, reflecting the fact that coherently undoing entanglement is a nonlocal task. The clean entanglement capacities of the basic resources are

$$E_{\text{clean}}([q \to q]) = E_{\text{clean}}([q \leftarrow q]) = [-1, 1] \tag{37a}$$

$$E_{\text{clean}}([c \to c]) = E_{\text{clean}}([c \leftarrow c]) = [-1, 0] \tag{37b}$$

$$E_{\text{clean}}([qq]) = \{1\} \tag{37c}$$

To understand Eq. (37a), observe that transmitting one qubit can map $|\Phi_2\rangle^{AB}$ to or from $|00\rangle^{AB}$, in each case without changing anything in the environment. The reasoning behind Eq. (37b) is less obvious since sending any classical message leaks the message to the environment by definition. However, the protocol can be made clean by always sending a uniformly random bit through the channel. If Alice generates this bit locally (or more simply sends a $|+\rangle$ state) and Bob discards it, then this does not change their amount of shared entanglement. Alternatively, if Alice sends her half of an ebit through the classical channel and Bob performs a CNOT with the transmitted bit as control and his half of the ebit as target, then this will eliminate the entanglement while presenting the same information to the environment.

These resources can be combined in various ways to create entanglement spread. For example, to create a superposition of 5 and 9 ebits, we might start with 8 ebits, use two cbits to create a superposition of 6 and 8 ebits and then use one qubit to create the desired superposition of 5 and 9 ebits. Implicit in these sort of protocols is a pair of control registers A_C, B_C that specify how many ebits Alice and Bob would like to end up with. In this example, they would like to map the state $(c_0|00\rangle + c_1|11\rangle)^{A_C B_C} \otimes |\Phi_2\rangle^{\otimes 8}$ (for arbitrary coefficients c_0, c_1) to

$$c_0|00\rangle^{A_C B_C}|\Phi_2\rangle^{\otimes 5}|00\rangle^{\otimes 4} + c_1|11\rangle^{A_C B_C}|\Phi_2\rangle^{\otimes 9}. \tag{38}$$

To achieve this transformation, Alice and Bob perform the following sequence of actions conditioned on their control qubits:

- If Alice's control qubit is zero, she sends two of her entangled qubits through the classical channel. If it is one, she sends two $|+\rangle$ states through the channel. Either way, the environment observes two random bits sent through the channel.
- If Bob's control qubit is zero, he uses the two bits he has received as controls for CNOTs that are applied to his halves of the ebits that Alice sent, thus mapping them to $|00\rangle$. Then he discards the bits he received. If Bob's control bit is one, he simply discards the bit he received. Either way the environment sees another copy of the same random bit being discarded by Bob. Moreover, this bit is now independent of the residual quantum state held by Alice and Bob. Alice and Bob now share

$$c_0|00\rangle|\Phi_2\rangle^{\otimes 6}|00\rangle^{\otimes 2} + c_1|11\rangle \otimes |\Phi_2\rangle^{\otimes 8}.$$

- If Alice's control qubit is zero, she sends half of one of her ebits through the qubit channel and locally creates a $|0\rangle$ state. If her control qubit is one, she locally creates a $|\Phi_2\rangle$ and sends half through the channel.
- If Bob's control qubit is zero, he now holds both halves of one of the $|\Phi_2\rangle$ states. He rotates this to a $|00\rangle$ state and discards one of the $|0\rangle$ qubits. If his control qubit is one, he keeps the transmitted qubit, but also creates and discards a $|0\rangle$ qubit. Alice and Bob are now left with the state in Eq. (38).

Observe that in this example the classical and quantum communication could have been sent in either direction. Thus, while some parts of the simulation protocol can only use forward communication, the spread requirements can be met with communication in either direction.

While this framework gives us a fairly clear understanding of the communication resources required to create entanglement spread, it also shows how unlimited ebits are not a good model of unlimited entanglement. Instead of maximally entangled states, we will use the so-called entanglement-embezzling [78] states $|\varphi_N\rangle^{AB}$, which are parameterized by their Schmidt rank N, and can be used catalytically to produce or destroy any Schmidt rank k state up to an error of $\frac{\log k}{\log N}$ in the trace norm. See [78] for a definition of $|\varphi_N\rangle$ and a proof of their entanglement-embezzling abilities. We let the resource $[€€]$ denote access to an embezzling state of arbitrary size: formally, $[€€] = \bigcup_{N \geq 1} \langle \varphi_N \rangle$ and so we have

$$E_{\text{clean}}([€€]) = (-\infty, \infty).$$

By the above discussion, this is strictly stronger than the resource $\infty[qq]$.

We remark that these sorts of entanglement transformations were studied by Nielsen [67] who gave conditions for when an entangled state could be prepared using unlimited classical communication. In this context, the term "maximally entangled" makes sense for ebits, since together with unlimited classical communication they can be used to prepare any other state with the same or smaller Schmidt rank. The low-communication case was also considered by Daftuar and Hayden [30].

We now return to parts (d) and (e) of Theorem 3. In (d), we need to run the simulation protocol for $\langle \mathcal{N}_F : \rho \rangle$ for all possible ρ in superposition.[9] We can discard a resource β at the end of the protocol, but β must be either independent of ρ for a feedback simulation or can depend only on $\hat{\mathcal{N}}(\rho)$ for a non-feedback simulation. By the equality in Eq. (17), this reduces to producing coherent superpositions of varying amounts of qubits and ebits.

The simplest case is when $\alpha = Q(\mathcal{N})[q \to q] + [€€]$. In this case, $\alpha \geq_{\text{CL}} Q(\mathcal{N})[q \to q] + E[qq] + [€€]$ for any E. Thus we can take $\beta = [€€]$ and so we have $\alpha \geq_{\text{CL}} \langle \mathcal{N} : \rho \rangle + \beta$ for all ρ. This establishes Eq. (27).

The most general case without embezzling states is when

$$\alpha = Q_1[q \to q] + Q_2[q \leftarrow q] + C_2[c \leftarrow c] + E[qq]. \tag{39}$$

[9] For technical reasons, our coding theorem will adopt a slightly different approach. But for the converse and for the present discussion, we can consider general inputs to be mixtures of tensor power states.

In this case, we always have the constraint

$$Q_1 \geq Q(\mathcal{N}) = \max_{\rho} \tfrac{1}{2} I(R;B)_\rho, \tag{40}$$

since $Q_1[q \to q]$ is the only source of forward communication. Suppose that $\beta = (E-e)[qq]$, for some $0 \leq e \leq E$, i.e. we will use all of the communication, but discard $E-e$ ebits of entanglement. Now, for each ρ, being able to simulate the channel on input $\rho^{\otimes n}$ requires creating at least $I(R;B)_\rho$ mutual information and $I(B;E)+\beta$ entanglement which is only possible if

$$\alpha \geq_{\text{CL}} \tfrac{1}{2} I(R;B)_\rho [q \to q] + (\tfrac{1}{2} I(E;B)_\rho + E - e)[qq].$$

Equivalently

$$\begin{aligned}(Q_1 - \tfrac{1}{2} I(R;B)_\rho)[q \to q] + Q_2[q \leftarrow q] + C_2[c \leftarrow c] \\ \geq_{\text{CL}} (\tfrac{1}{2} I(E;B)_\rho - e)[qq].\end{aligned} \tag{41}$$

We can calculate when Eq. (41) holds by using the spread capacity expressions in Eq. (37). First, if $\frac{1}{2} I(E;B)_\rho - e \geq 0$ then the $C_2[c \leftarrow c]$ is not helpful and we simply have $Q_1 - \frac{1}{2} I(R;B)_\rho + Q_2 \geq \frac{1}{2} I(E;B)_\rho - e$, or equivalently

$$Q_1 + Q_2 \geq H(B)_\rho - e.$$

Alternatively, if $\frac{1}{2} I(E;B)_\rho - e \leq 0$ then we have the inequality $Q_1 - \frac{1}{2} I(R;B)_\rho + Q_2 + C_2 \geq e - \frac{1}{2} I(E;B)_\rho$, which is equivalent to

$$Q_1 + Q_2 + C_2 \geq e - H(B|R)_\rho.$$

We will consider the case when E is sufficiently large so that it does not impose any constraints on the other parameters. This results in the bound

$$2(Q_1 + Q_2) + C_2 \geq \max_{\rho} H(B)_\rho - \min_{\rho} H(B|R)_\rho - C_E(\mathcal{N}).,$$

whose RHS is precisely $\Delta_{\text{sim}}(\mathcal{N})$ from Definition 4. (42)

The role of communication can be thought of as both creating mutual information between R and B and in creating entanglement spread. Both are necessary for channel simulation, but only forward communication can create mutual information, while backwards or forward communication (or even other resources, such as embezzling states) can be used to create spread.

The non-feedback case (e) of Theorem 3 adds one additional subtlety: since the simulation gives part of the input to Eve, it does not have to preserve superpositions between as many different input density matrices. In particular, if the input density matrix is $\rho^{\otimes n}$, then Eve learns $\hat{\mathcal{N}}(\rho)$. Thus, we need to run our protocol in an incoherent superposition over different values of $\hat{\mathcal{N}}(\rho)$ and then in a coherent superposition within each $\hat{\mathcal{N}}^{-1}(\omega)$. Intuitively we can think of the input as a superposition over purifications of different tensor powers $\rho^{\otimes n}$. This picture can be made rigorous by the post-selection principle [22] and gentle tomography [50], [11], but we will not explore this approach in detail. In this picture Eve learns $\omega = \hat{\mathcal{N}}(\rho)$ up to accuracy $O(1/\sqrt{n})$, and this collapses the superposition to inputs $\rho^{\otimes n}$ with $\rho \in \hat{\mathcal{N}}^{-1}(\omega)$. Thus we need only consider entanglement spread over the sets $\hat{\mathcal{N}}^{-1}(\omega)$.

Unfortunately, even in the case of a fixed input ρ, the additivity question is open. Until it is resolved, we cannot avoid regularized formulas. However, conceptually part (e) adds to part (d) only the issues of regularization and optimization over ways of splitting E^n into parts for Alice and Bob.

At this point it is natural to ask whether spread is only helpful for *feedback* simulations. The amplitude damping channel of Fig. 6 and the variable-entropy channel both have efficient non-feedback simulations on general inputs, using C_E bits of forward communication, even when entanglement is supplied as ordinary ebits. One way to see why is to observe that in each case $H(\mathcal{N}(\rho))$ is uniquely determined by $\hat{\mathcal{N}}(\rho)$, so that measuring the average density matrix of Eve will leave no room for spread. An optimal simulation can gently measure the average density matrix of Eve, transmit this information classically to Bob, and then use the appropriate number of ebits.

A second way to see that spread is not needed to simulate these two channels is to give explicit choices of the isometry in Eq. (29). This is easier to do for the variable-entropy channel, for which

$$\mathcal{M}_{d,F} = |00\rangle^{BE}\langle 0|^A + \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |ii\rangle^{BE}\langle 1|^A.$$

Define an isometry $V_1 : E \to E_A E_B$ by

$$V_1 = \sum_{i=1}^{d} \left(\frac{1}{\sqrt{d}} |0i\rangle^{E_A} |i\rangle^{E_B} \langle 0|^E + |1i\rangle^{E_A} |i\rangle^{E_B} \langle 1|^E \right).$$

Then $V_1 \circ \mathcal{M}_{d,F}$ is equivalent to transmitting a classical bit from Alice to Bob and creating a d-dimensional maximally entangled state between Bob and Eve. This can be simulated using one cbit by having Bob locally create a d-dimensional maximally mixed state.

However, the above reasoning does not extend to more complicated situations. In Sec. IV-E3 we exhibit a channel whose efficient simulation requires spread-generating resources such as embezzling states or back communication even in the non-feedback setting.

D. Relation to other communication protocols

Special cases of Theorem 3 include remote state preparation [12] (and the qubit-using variant, super-dense coding of quantum states [43]) for CQ-channels $\mathcal{N}(\rho) = \sum_j \langle j|\rho|j\rangle \sigma_j$; the co-bit equality $[q \to qq] = ([q \to q] + [qq])/2$ [40]; measurement compression [85] (building on [65], [66]) for qc-channels $\mathcal{N}(\rho) = \sum_j \operatorname{Tr}(\rho M_j)|j\rangle\langle j|$ where (M_j) is a POVM; entanglement dilution [8] for a constant channel $\mathcal{N}(\rho) = \sigma_0$; and entanglement of purification (EoP) [75] – it was shown by Hayashi [46] that optimal visible compression of mixed state sources is given by the regularized EoP.

The Wyner protocol for producing a classical correlated distribution [89] is a static analogue of the cbit-rbit tradeoff. Similarly, the entanglement of purification is a static version of the qubits-but-no-ebits version of the QRST.

For feedback channels, [32] showed that the QRST can be combined with the so-called "feedback father" to obtain the resource equivalence Eq. (17). On the other hand, [32] also

showed that running the QRST backwards on a fixed i.i.d. source yields state merging [57], a.k.a. fully-quantum Slepian-Wolf. This implies that merging can be used to provide an alternate construction of the QRST on a known i.i.d. source [1]. More recently, [90] has introduced *state redistribution* which simultaneously generalizes state merging and splitting, by determining the optimal rate at which a system can be sent from one party to another when both parties hold ancilla systems that are in some way entangled with the system being sent.

As remarked earlier, the version of the classical reverse Shannon theorem proved here, Theorem 1, differs from the version originally proved in [14] (which also first conjectured Theorem 3). In the earlier version, the simulation was exactly faithful even for finite block size, and asymptotically efficient in the amount of communication used, but exponentially inefficient in the amount of shared randomness. The version proved here is only asymptotically faithful, but importantly stronger in being asymptotically efficient in its use both of classical communication and shared randomness. None of our simulations, nor other results in this area (apart from [27]), achieve the zero-error performance of [14]. We believe that zero-error simulation of classical channels using optimal rates of communication requires exponential amounts of shared randomness, and that for quantum channels, zero-error simulations do not exist in general. Apart from some easy special cases (e.g. quantum feedback channels), we do not know how to prove these conjectures.

III. Simulation of classical channels

A. Overview

This section is devoted to the proof of Theorem 1 (the classical reverse Shannon theorem). Previously the high-randomness cases of Theorem 1 were proved in [14], [84] and its converse was proved in [84]. Here we will review their proof and show how it can be extended to cover the low-randomness case (parts (c,d,e) of Theorem 1). Similar results have been obtained independently in [28].

The intuition behind the reverse Shannon theorem can be seen by considering a toy version of the problem in which all probabilities are uniform. Consider a regular bipartite graph with vertices divided into (X, Y) and with edges $E \subset X \times Y$. Since the graph is regular, every vertex in X has degree $|E|/|X|$ and every vertex in Y has degree $|E|/|Y|$. For $x \in X$, let $\Gamma(x) \subset Y$ be its set of neighbors. We can use this to define a channel from X to Y: define $N(y|x)$ to be $1/|\Gamma(x)| = |X|/|E|$ if $y \in \Gamma(x)$ and 0 if not. In other words, N maps x to a random one of its neighbors. We call these channels "unweighted" since their transition probabilities correspond to an unweighted graph.

In this case, it is possible to simulate the channel N using a message of size $\approx \log(|X| \cdot |Y|/|E|)$ and using $\approx \log(|E|/|X|)$ bits of shared randomness. This can be thought of as a special case of part (a) of Theorem 1 in which N is an unweighted channel and we are only simulating a single use of N. This is achieved by approximately decomposing N into a probabilistic mixture of channels and using the shared randomness to select which one to use. We will choose these channels such that their ranges are disjoint subsets of Y, and in fact, will construct them by starting with a partition of Y and working backwards. The resulting protocol is analyzed in the following lemma.

Lemma 7. *Consider a channel $N : X \to Y$ with $N(y|x) = 1_{\Gamma(x)}(y)/|\Gamma(x)|$, where 1_S denotes the indicator function for a set S. Choose positive integers r, m such that $rm = |Y|$ and let $\gamma = m|E|/|X|\,|Y|$. Choose a random partition of Y into subsets $Y_1, \ldots, Y_r$, each of size m, and for $y \in Y$ define $i(y)$ to be the index of the block containing y. Define*

$$\tilde{N}(y|x) = \frac{1_{\Gamma(x)}(y)}{r \cdot |\Gamma(x) \cap Y_{i(y)}|}$$

to be the channel that results from the following protocol:

1) *Let $i \in [r]$ be a uniformly chosen random number shared by Alice and Bob.*
2) *Given input x, Alice chooses a random element of $\Gamma(x) \cap Y_i$ (assuming that one exists) and transmits its index $j \in [m]$ to Bob.*
3) *Bob outputs the j^{th} element of Y_i.*

Then it holds with probability $\geq 1 - 2re^{-\gamma\epsilon^2}$ that $\|N(\cdot|x) - \tilde{N}(\cdot|x)\|_1 \leq \epsilon$ for all x.

If we choose $\gamma = 2(\ln 4|E|)/\epsilon^2$ then there is a nonzero probability of a good partition existing. In this case we can derandomize the construction and simply say that a partition of Y exists such that the above protocol achieves low error on all inputs.

The idea behind Lemma 7 is that for each x and i, the random variable $|\Gamma(x) \cap Y_i|$ has expectation close to

$$|\Gamma(x)| \cdot |Y_i|/|Y| = \frac{|E|}{|X|} \cdot \frac{m}{|Y|} = \gamma,$$

with typical fluctuations on the order of $\sqrt{\gamma}$. If γ is large then these fluctuations are relatively small, and the channel simulation is faithful. Similar "covering lemmas" appeared in Refs. [84], [28], and were anticipated by Ref. [39] and Thm 6.3 of [89]. The details of the proof are described in Sec. III-B.

The difference between Lemma 7 and the classical reverse Shannon theorem (i.e. part (a) of Theorem 1) is that in the latter we are interested in an asymptotically growing number of channel uses n and in simulating general channels N, instead of unweighted channels. It turns out that when n is large, N^n looks mostly like an unweighted channel, in a sense that we will make precise in Sec. III-C. We will see that Alice need communicate only $O(\log(n))$ bits to reduce the problem of simulating N^n to the problem of simulating an unweighted channel. This will complete the proof of the direct part of part (a) of Theorem 1.

One feature of the protocol in Lemma 7 is that Bob uses only shared randomness (i) and the message from Alice (j) in order to produce his output y. As a result, the protocol effectively simulates the feedback channel N_F in which Alice also gets a copy of y. Conversely, in order to simulate a feedback channel, Bob cannot use local randomness in any significant way.

On the other hand, if Alice does not need to learn y, then we can consider protocols in which some of the random bits used are shared and some are local to Alice or Bob. This will allow us to reduce the use of shared randomness at the cost of some extra communication. The resulting trade-off between the resources is given in part (c) of Theorem 1. In order to prove it, we will again first consider the unweighted case.

The idea will be to decompose the channel $N(y|x)$ as the composition of channels $N_1(w|x)$ and $N_2(y|w)$; i.e. $N(y|x) = (N_2 \circ N_1)(x) = \sum_{w\in W} N_1(w|x)N_2(y|w)$. In this case Alice can simulate the channel N on input x, by simulating N_1 to produce intermediate output w on which Bob locally applies N_2 to produce y. Since w is generally more correlated with x than y, this will require more communication than simply simulating N directly as in Lemma 7. However, since Bob simulates N_2 using local randomness, the protocol may require less shared randomness, and more importantly, the total amount of communication plus shared randomness may be lower.

We will assume that the channels N, N_1 and N_2 are all unweighted channels. Let the corresponding bipartite graphs for N, N_1, N_2 have edges $E_{XY} \subset X\times Y$, $E_{XW} \subset X\times W$ and $E_{YW} \subset W \times Y$, respectively. We use $\Gamma_{XY}(x)$ to denote the neighbors of x in Y; that is, $\Gamma_{XY}(x) = \{y : (x, y) \in E_{XY}\}$. Similarly, we can define $\Gamma_{YX}(y)$ to be the neighbors of y in X, $\Gamma_{XW}(x)$ to be the neighbors of x in W and so on. We assume that the graphs are regular, so that $|\Gamma_{XW}(x)| = |E_{XW}|/|X|$ for all x, $|\Gamma_{WY}(w)| = |E_{WY}|/|W|$ for all w, and so on. Combined with the fact that $N = N_2 \circ N_1$, we find that

$$\begin{aligned}\frac{1_{\Gamma_{XY}(x)}(y)}{|E_{XY}|/|X|} = N(y|x) &= \sum_{w\in W} N_2(y|w)N_1(w|x)\\ &= \sum_{w\in W} \frac{1_{\Gamma_{XW}(x)}(w)}{|E_{XW}|/|X|}\cdot\frac{1_{\Gamma_{WY}(w)}(y)}{|E_{WY}|/|W|}\\ &= \frac{|\Gamma_{XW}(x)\cap\Gamma_{YW}(y)|}{|E_{XW}|\cdot|E_{WY}|/|X|\,|W|}, \quad (43)\end{aligned}$$

Rearranging terms yields the identity

$$|\Gamma_{XW}(x)\cap\Gamma_{YW}(y)| = 1_{\Gamma_{XY}(x)}(y)\frac{|E_{XW}|\,|E_{WY}|}{|E_{XY}|\,|W|}. \quad (44)$$

The protocol is now defined in a way similar to the one in Lemma 7.

Lemma 8. *Choose positive integers r, m such that $m = \gamma|X|\,|W|/|E_{XW}|$ and $r = |E_{XY}|\,|W|/|E_{WY}|\,|X|$. Choose disjoint sets $W_1, \ldots, W_r \subset W$ at random, each of size m. Let $W_i = \{w_{i,1}, \ldots, w_{i,m}\}$. Let $\tilde{N}(y|x)$ be the channel resulting from the following protocol:*

1) *Let $i \in [r]$ be a uniformly chosen random number shared by Alice and Bob.*
2) *Given input x, Alice chooses a random $w_{i,j} \in \Gamma(x)\cap W_i$ (assuming that one exists) and transmits its index $j \in [m]$ to Bob.*
3) *Bob outputs y with probability $N_2(y|w_{i,j})$.*

Then it holds with probability $\geq 1 - 2|E_{XY}|e^{-\gamma\epsilon^2/32}$ that $\|N(\cdot|x) - \tilde{N}(\cdot|x)\|_1 \leq \epsilon$ for all x.

We can take $\gamma = 32(\ln 2|E_{XY}|)/\epsilon^2$ and derandomize the statement of the Lemma.

Note that in general we will have $rm < |W|$, so that this protocol does not use all of W. This should not be surprising, since faithfully simulating the channel N_1 should in general be more expensive than simulating N. The trick is to modify the simulation of N_1 that would be implied by Lemma 7 to use less randomness, since we can rely on Bob's application of N_2 to add in randomness at the next stage.

B. Proof of unweighted classical reverse Shannon theorem

In this section we prove Lemma 7 and Lemma 8. The main tool in both proofs is the Hoeffding bound for the hypergeometric distribution [53]. The version we will need is

Lemma 9 (Hoeffding [53])**.** *For integers $0 < a \leq b < n$, choose A and B to be random subsets of $[n]$ satisfying $|A| = a$ and $|B| = b$. Then $\mu := \mathbb{E}[|A\cap B|] = ab/n$ and*

$$\Pr\left[|A\cap B| \geq (1+\epsilon)\mu\right] \leq e^{-\frac{\mu\epsilon^2}{2}} \quad (45)$$

$$\Pr\left[|A\cap B| \leq (1-\epsilon)\mu\right] \leq e^{-\frac{\mu\epsilon^2}{2}} \quad (46)$$

$$\Pr\left[|\,|A\cap B| - \mu| \geq \epsilon\mu\right] \leq 2e^{-\frac{\mu\epsilon^2}{2}} \quad (47)$$

Now we turn to Lemma 7. We can calculate

$$\begin{aligned}\|N(\cdot|x) - \tilde{N}(\cdot|x)\|_1 &= \sum_{y\in\Gamma(x)} |N(y|x) - \tilde{N}(y|x)|\\ &= \sum_{y\in\Gamma(x)}\left|\frac{1}{|\Gamma(x)|} - \frac{1}{r\cdot|\Gamma(x)\cap Y_{i(y)}|}\right|\\ &= \sum_{i=1}^{r}\sum_{y\in\Gamma(x)\cap Y_i}\left|\frac{1}{|\Gamma(x)|} - \frac{1}{r\cdot|\Gamma(x)\cap Y_i|}\right|\\ &= \sum_{i=1}^{r}\left|\frac{|\Gamma(x)\cap Y_i|}{|\Gamma(x)|} - \frac{1}{r}\right| \quad (48)\end{aligned}$$

To apply Lemma 9, take $A = \Gamma(x)$ and $B = Y_i$, so that $a = |E|/|X| = r\gamma$, $b = m = |Y|/r$, $n = |Y|$ and $\mu = \gamma$. Then each term in the sum in Eq. (48) is $\leq \epsilon/r$ with probability $\geq 1 - 2e^{-\gamma\epsilon^2/2}$. Taking the union bound over all a and i completes the proof of Lemma 7.

The proof of Lemma 8 is similar. This time

$$\begin{aligned}\tilde{N}(y|x) &= \frac{1}{r}\sum_{i=1}^{r}\sum_{w\in W_i}\Pr\left[\text{Alice sends } w|x,i\right]N_2(y|w)\\ &= \frac{1}{r}\sum_{i=1}^{r}\sum_{w\in W_i}\frac{1_{\Gamma_{XW}(x)}(w)}{|\Gamma_{XW}(x)\cap W_i|}\cdot\frac{1_{\Gamma_{WY}(w)}(y)}{|\Gamma_{WY}(w)|}\\ &= \frac{|W|}{r|E_{WY}|}\sum_{i=1}^{r}\frac{|\Gamma_{XW}(x)\cap\Gamma_{YW}(y)\cap W_i|}{|\Gamma_{XW}(x)\cap W_i|}.\end{aligned}$$

We will use Lemma 9 twice. First, consider $|\Gamma_{XW}(x)\cap W_i|$. This has expectation equal to γ and therefore

$$\Pr\left[|\Gamma_{XW}(x)\cap W_i| \geq (1+\epsilon/4)\gamma\right] \leq e^{-\gamma\epsilon^2/32}.$$

(We will see that the one-sided bound simplifies some of the later calculations.) Taking the union bound over all $|X|r \leq$

$|E_{XY}|$ values of x, i, we find that $|\Gamma_{XW}(x) \cap W_i| \leq (1+\epsilon/4)\gamma$ for all x, i with probability $\geq 1 - |E_{XY}|e^{-\gamma\epsilon^2/32}$. Assuming that this is true, we obtain

$$\begin{aligned}\tilde{N}(y|x) &\geq \frac{|W|}{r|E_{WY}|}\sum_{i=1}^{r}\frac{|\Gamma_{XW}(x)\cap\Gamma_{YW}(y)\cap W_i|}{(1+\epsilon/4)\gamma} \\ &= \frac{|W|}{r|E_{WY}|}\frac{|\Gamma_{XW}(x)\cap\Gamma_{YW}(y)\cap\tilde{W}|}{(1+\epsilon/4)\gamma}, \end{aligned} \tag{49}$$

where we define $\tilde{W} = W_1 \cup \ldots \cup W_r$. Note that $\tilde{W}$ is a random subset of W of size rm. Using Eq. (44) we find that $\mathbb{E}[|\Gamma_{XW}(x)\cap\Gamma_{YW}(y)\cap\tilde{W}|]$ is equal to γ when $(x, y) \in E_{XY}$ and 0 otherwise. Again we use Lemma 9 to bound

$$\Pr\left[|\Gamma_{XW}(x)\cap\Gamma_{YW}(y)\cap\tilde{W}| \leq (1-\epsilon/4)\gamma\right] \leq e^{-\gamma\epsilon^2/32}$$

for all $(x, y) \in E_{XY}$. Now we take the union bound over all pairs $(x, y) \in E_{XY}$ to find that

$$|\Gamma_{XW}(x)\cap\Gamma_{YW}(y)\cap\tilde{W}| \geq (1-\epsilon/4)\gamma \tag{50}$$

with probability $\geq 1 - |E_{XY}|e^{-\gamma\epsilon^2/32}$. When both Eq. (49) and Eq. (50) hold and $(x, y) \in E_{XY}$ it follows that

$$\begin{aligned}\tilde{N}(y|x) &\geq \frac{|W|}{r|E_{WY}|}\frac{1-\epsilon/4}{1+\epsilon/4} > (1-\epsilon/2)\frac{W}{r|E_{WY}|} \\ &= (1-\epsilon/2)\frac{|X|}{|E_{XY}|} = (1-\epsilon/2)N(y|x). \end{aligned} \tag{51}$$

Finally we compare with Eq. (43) to obtain

$$\begin{aligned}\|N(y|x) - \tilde{N}(y|x)\|_1 &= 2\sum_{y\in Y}\max(0, N(y|x) - \tilde{N}(y|x)) \\ &< \epsilon\sum_{y\in Y}N(y|x) = \epsilon. \end{aligned} \tag{52}$$

This concludes the proof of Lemma 8.

C. Classical types

In this section we show how the classical method of types can be used to extend Lemmas 7 and 8 to prove the coding parts of Theorem 1. We begin with a summary of the arguments aimed at readers already familiar with the method of types (a more pedagogical presentation is in [26]). The idea is to for Alice to draw a joint type according to the appropriate distribution and to send this to Bob. This requires $O(\log(n))$ bits of communication and conditioned on this joint type they are left with an unweighted channel and can apply Lemma 7. It is then a counting exercise to show that the communication and randomness costs are as claimed. For the low-randomness case, the protocol is based on a decomposition $N^{X\to Y}$ into $N_2^{W\to Y}\circ N_1^{X\to W}$. Alice draws an appropriate joint type for all three variables (X, W, Y) and transmits this to Bob. Again this involves $O(\log(n))$ bits of communication and leaves them with an unweighted channel, this time of the form that can be simulated with Lemma 8.

To prove these claims, we begin by reviewing the method of types, following [26]. We will use $\mathcal{X}, \mathcal{Y}, \mathcal{W}$ to denote single-letter alphabets, while reserving X, Y, W for block variables. Consider a string $x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n$. Define the type of x^n to be the $|\mathcal{X}|$-tuple of integers $t(x^n) := \sum_{j=1}^n e_{x_j}$, where $e_j \in \mathbb{Z}^{|\mathcal{X}|}$ is the unit vector with a one in the i$^{\text{th}}$ position. Thus $t(x^n)$ counts the frequency of each symbol $x \in \mathcal{X}$ in x^n. Let $\mathcal{T}_{\mathcal{X}}^n$ denote the set of all possible types of strings in $\mathcal{X}^n$. Since an element of $\mathcal{T}_{\mathcal{X}}^n$ can be written as $|\mathcal{X}|$ numbers ranging from $0, \ldots, n$ we obtain the simple bound $|\mathcal{T}_{\mathcal{X}}^n| = \binom{n+|\mathcal{X}|-1}{|\mathcal{X}|-1} \leq (n+1)^{|\mathcal{X}|}$. For a type t, let the normalized probability distribution $\bar{t} := t/n$ denote its empirical distribution.

For a particular type $t \in \mathcal{T}_{\mathcal{X}}^n$, denote the set of all strings in $\mathcal{X}^n$ with type t by $T_t = \{x^n \in \mathcal{X}^n : t(x^n) = t\}$. From [26], we have

$$(n+1)^{-|\mathcal{X}|}\exp(nH(\bar{t})) \leq |T_t| = \binom{n}{t} \leq \exp(nH(\bar{t})), \tag{53}$$

where $\binom{n}{t}$ is defined to be $\frac{n!}{\prod_{x\in\mathcal{X}} t_x!}$. Next, let p be a probability distribution on $\mathcal{X}$ and $p^{\otimes n}$ the probability distribution on $\mathcal{X}^n$ given by n i.i.d. copies of p, i.e. $p^{\otimes n}(x^n) := p(x_1)\cdots p(x_n)$. Then for any $x^n \in T_t$ we have $p^{\otimes n}(x^n) = \prod_{x\in\mathcal{X}} p(x)^{t_x} = \exp(-n(H(\bar{t}) + D(\bar{t}\|p)))$. Combining this with Eq. (53), we find that

$$\frac{\exp\left(-nD(\bar{t}\|p)\right)}{(n+1)^{|\mathcal{X}|}} \leq p^{\otimes n}(T_t) \leq \exp\left(-nD(\bar{t}\|p)\right), \tag{54}$$

Thus, as n grows large, we are likely to observe an empirical distribution $\bar{t}$ that is close to the actual distribution p. To formalize this, define the set of *typical sequences* $T_{p,\delta}^n$ by

$$T_{p,\delta}^n := \bigcup_{\substack{t\in\mathcal{T}_{\mathcal{X}}^n \\ \|\bar{t}-p\|_1\leq\delta}} T_t. \tag{55}$$

To bound $p^{\otimes n}(T_{p,\delta}^n)$, we apply Pinsker's inequality [70]:

$$D(q\|p) \geq \frac{1}{2\ln 2}\|p-q\|_1^2 \tag{56}$$

to show that

$$p^{\otimes n}(T_{q,\delta}^n) \geq 1 - (n+1)^{|\mathcal{X}|}\exp\left(-\frac{n\delta^2}{2\ln 2}\right). \tag{57}$$

We will also need the Fannes-Audenaert inequality [36], [7] which establishes the continuity of the entropy function. Let $\eta(x) = -x\log x - (1-x)\log(1-x)$. Then if p, q are probability distribution on d letters,

$$|H(p)-H(q)| \leq \frac{1}{2}\|p-q\|_1\log(d-1) + \eta\left(\frac{1}{2}\|p-q\|_1\right) \tag{58}$$

If we have a pair of strings $x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n$, then we can define their joint type $t(x^n y^n)$ simply to be the type of the string $(x_1y_1, \ldots, x_ny_n) \in (\mathcal{X}\times\mathcal{Y})^n$. Naturally the bounds in Eq. (53) and Eq. (54) apply equally well to joint types, with $\mathcal{X}$ replaced by $\mathcal{X}\times\mathcal{Y}$. If t is a joint type then we can define its marginals $t^{\mathcal{X}} \in \mathbb{Z}^{|\mathcal{X}|}$ and $t^{\mathcal{Y}} \in \mathbb{Z}^{|\mathcal{Y}|}$ by $t_x^{\mathcal{X}} = \sum_{y\in\mathcal{Y}} t_{x,y}$ and $t_y^{\mathcal{Y}} = \sum_{x\in\mathcal{X}} t_{x,y}$. Let $N(y|x)$ denote a noisy channel from $\mathcal{X} \to \mathcal{Y}$ with $N^n(y^n|x^n) := N(y_1|x_1)\cdots N(y_n|x_n)$. Then $N^n(y^n|x^n)$ depends only on the type $t = t(x^n y^n)$ according to $N^n(y^n|x^n) = \prod_{x,y} N(y|x)^{t_{x,y}}$.

We now have all the tools we need to reduce Theorem 1 to Lemmas 7 and 8. First, consider parts (a,b) of Theorem 1, where we have an ample supply of shared randomness. In either case, the protocol is as follows:

1) Alice receives input x^n. This may be expressed in a type-index representation as (t_A, p_A). Here $t_A = t(x^n)$ is the input type, and $p_A \in [|T_{t_A}|]$ is defined by assigning the integers $1, \ldots, |T_{t_A}|$ arbitrarily to the elements of T_{t_A}.
2) Alice simulates the channel N^n locally to generate a provisional output $\tilde{y}^n$. Let $t_B = t(\tilde{y}^n)$ be the output type and $t_{AB} \in \mathbb{Z}^{|\mathcal{X}\times\mathcal{Y}|}$ the joint type between x^n and $\tilde{y}^n$. Having determined these types, Alice discards $\tilde{y}^n$, as it is no longer needed.
3) Alice sends t_B to Bob using $|\mathcal{X} \times \mathcal{Y}| \log(n+1)$ bits.
4) Alice and Bob use $n(H(Y) - C) + o(n)$ bits of shared randomness to pick a subset S_i from a preagreed partitioning of outputs of type t_B into approximately equal disjoint subsets, each of cardinality approximately 2^{nC}, where C is the Shannon capacity of channel N.
5) Alice finds a string $y^n \in S_i$ having the same joint type with x^n as $\tilde{y}^n$ had. But because y^n lies in the chosen subset S_i, which Bob already knows, Alice can transmit y^n to Bob more efficiently, by a message of size only $nC + o(n)$ bits, using the method of Lemma 7 (Let $X = T_{t_A}$, $Y = T_{t_B}$ and $E = T_{t_{AB}} \subset X \times Y$ define a regular bipartite graph. To simulate the action of N^n on $x^n \in X$, conditioned on $(x^n, y^n) \in E$, we need only to choose a random neighbor y^n of x^n in this graph.)

This protocol is depicted in Fig. 8.

It remains only to analyze the cost of this last step. The communication cost is (taking notation from the statement of Lemma 7)

$$\begin{aligned}
\log(m) &= \log\left(\frac{|X|\,|Y|\gamma}{|E|}\right) \\
&= \log\left(\frac{|T_{t_A}|\,|T_{t_B}|\,(2\ln(2|T_{t_{AB}}|)/\epsilon^2)}{|T_t|}\right) \\
&\leq n(H(\bar{t}_A) + H(\bar{t}_B) - H(\bar{t}_{AB})) + |\mathcal{X}\times\mathcal{Y}|\log(n+1) \\
&\quad + \log(2\ln(2)nH(\bar{t}_{AB})/\epsilon^2) \\
&= nI(\mathcal{X};\mathcal{Y})_{\bar{t}_{AB}} + O(\log(n)) + \log(1/\epsilon^2).
\end{aligned}$$

Since $C(N) \geq I(\mathcal{X};\mathcal{Y})_{\bar{t}}$ for all t, this establishes part (b) of Theorem 1. Continuing, we estimate the randomness cost to be

$$\begin{aligned}
\log(r) &= \log\left(\frac{|E|}{|X|\gamma}\right) \leq \log\left(\frac{|E|}{|X|}\right) \\
&\leq n(H(\bar{t}_{AB}) - H(\bar{t}_A)) + O(\log(n)) \\
&= nH(\mathcal{Y}|\mathcal{X})_{\bar{t}_{AB}} + O(\log(n)).
\end{aligned}$$

To prove part (a), we need to relate entropic quantities defined for $\bar{t}$ to the corresponding quantities for p. This will be done with typical sets (Eq. (57)) and the Fannes-Audenaert inequality (Eq. (58)). If p is a distribution on X then let $q = N_F(p)$ be the joint distribution on X and Y that results from sending X through N and obtaining output Y. Then Eq. (57) implies that following the above protocol results in values of $\bar{t}$ that are very likely to be close to q. In particular, $q^{\otimes n}(T^n_{q,\delta}) \geq 1 - (n+1)^d e^{-n\delta^2/2}$, where $d = |\mathcal{X} \times \mathcal{Y}|$. Next, the Fannes-Audenaert inequality says that if $\bar{t} \in T^n_{q,\delta}$ then $|H(\bar{t}) - H(q)| \leq \delta \log(d/\delta)$. Applying this to each term in $I(\mathcal{X};\mathcal{Y}) = H(\mathcal{X}) + H(\mathcal{Y}) - H(\mathcal{X}\mathcal{Y})$, we obtain that $|I(\mathcal{X};\mathcal{Y})_{\bar{t}} - I(\mathcal{X};\mathcal{Y})_q| \leq 3\delta\log(d/\delta)$ and $|H(\mathcal{Y}|\mathcal{X})_{\bar{t}} - I(\mathcal{Y}|\mathcal{X})_q| \leq 2\delta\log(d/\delta)$. Taking δ to be $n^{-1/4}$, we obtain a sequence of protocols where both error and inefficiency simultaneously vanish as $n \to \infty$.

Similarly, for part (c), we need to consider the joint distribution q of $\mathcal{X}\mathcal{W}\mathcal{Y}$ that results from drawing $\mathcal{X}$ according to p, sending it through N_1 to obtain $\mathcal{W}$ and then sending $\mathcal{W}$ through N_2 to obtain Y. The protocol is as follows:

1) Suppose Alice's input is x^n.
2) Alice simulates $N_1^n(x^n)$ to obtain $\tilde{w}^n$ and then simulates $N_2^n(\tilde{w}^n)$ to obtain $\tilde{y}^n$.
3) Alice sets $t_{AWB} = t(x^n\tilde{w}^n\tilde{y}^n)$. She will not make any further use of $\tilde{w}^n$ or $\tilde{y}^n$.
4) Alice sends t to Bob using $|\mathcal{X}\times\mathcal{W}\times\mathcal{Y}|\log(n+1)$ bits.
5) Define $X = T_{t_A}$, $W = T_{t_\mathcal{W}}$, $Y = T_{t_B}$, $E_{XY} = T_{t_{AB}}$, $E_{XW} = T_{t_{AW}}$ and $E_{WY} = T_{t_{WB}}$. To simulate the action of N^n on $x^n \in X$, conditioned on $(x^n, w^n, y^n) \in T_t$, we need only to choose a random element of $\Gamma_{XY}(x^n)$ in this graph. This is achieved with Lemma 8.

The analysis of this last step is similar to that of the previous protocol. The communication cost is $\log(m) = \log\left(|X|\,|W|\gamma/|E_{XW}|\right) = nI(\mathcal{X};\mathcal{W})_{\bar{t}} + O(\log n)$ and the randomness cost is

$$\begin{aligned}
\log(r) &= \log\left(\frac{|E_{XY}|\cdot|W|}{|E_{WY}|\,|X|}\right) \\
&= n(H(\mathcal{X}\mathcal{Y})_{\bar{t}} + H(\mathcal{W})_{\bar{t}} - H(\mathcal{W}\mathcal{Y})_{\bar{t}} - H(\mathcal{X})_{\bar{t}}) + O(\log n) \\
&\qquad \text{using Eq. (53)} \\
&= n(H(\mathcal{X}\mathcal{Y})_{\bar{t}} + H(\mathcal{X}\mathcal{W})_{\bar{t}} - H(\mathcal{X}\mathcal{W}\mathcal{Y})_{\bar{t}} - H(\mathcal{W})_{\bar{t}}) + O(\log n) \\
&\qquad \text{using the Markov condition } I(\mathcal{X};\mathcal{Y}|\mathcal{W})_{\bar{t}} = 0 \\
&= nI(\mathcal{Y};\mathcal{W}|\mathcal{X})_{\bar{t}} + O(\log n) \\
&= n(I(\mathcal{X}\mathcal{Y};\mathcal{W})_{\bar{t}} - I(\mathcal{X};\mathcal{W})_{\bar{\tau}}) + O(\log n) \qquad (59)
\end{aligned}$$

This concludes the proofs of the existence of channel simulations claimed in Theorem 1.

D. Converses

In this section we discuss why the communication rates for the above protocols cannot be improved. The lower bound for simulating feedback channels was proven in [84] and for non-feedback channels in [28]. We will not repeat the proofs here, but only sketch the intuition behind them.

First, the communication cost must always be at least $C(N)$, or $I(X;Y)_p$ if the input is restricted to be from the distribution p. Otherwise we could combine the simulation with Shannon's [forward] noisy channel coding theorem to turn a small number of noiseless channel uses into a larger number of uses. This is impossible even when shared randomness is allowed.

Next, if N_F (i.e., the channel including noiseless feedback) is to be simulated, then Bob's output (with entropy $H(Y)$) must be entirely determined by the C bits of classical communication sent and the R bits of shared randomness used. Therefore we must have $C + R \geq H(Y)$.

The situation is more delicate when the simulation does not need to provide feedback to Alice. Suppose we have a protocol

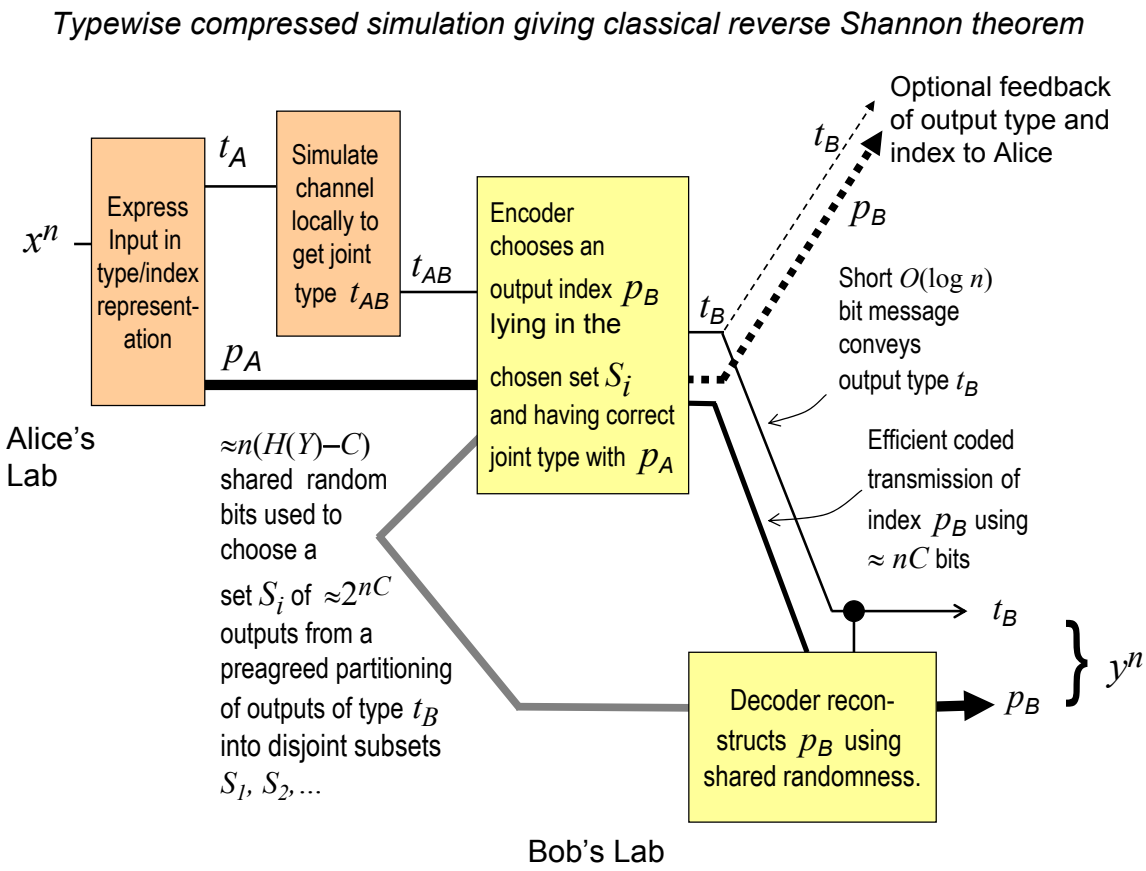

Fig. 8. The protocol for the classical reverse Shannon theorem (Theorem 1).

that uses C cbits and R rbits. Then let $W = (W_1, W_2)$ comprise both the message sent (W_1) and the shared random string (W_2). We immediately obtain $I(XY;W) \leq H(W) \leq C + R$. Additionally, the shared randomness W_2 and the message X are independent even given the message W_1; in other words $I(X;W_2|W_1) = 0$. Thus $I(X;W) = I(X;W_1) \leq H(W_1) \leq C$. Finally we observe that $X - W - Y$ satisfies the Markov chain condition since Bob produces Y only by observing W. This argument is discussed in more detail in [28], where it is also proven that it suffices to consider single-letter optimizations.

These converses are also meaningful, and essentially unchanged, when we consider negative R, corresponding to protocols that output shared randomness.

We observe that some of these converses are obtained from coding theorems and others are obtained from more traditional entropic bounds. In the cases where the converses are obtained from coding theorems then we in fact generally obtain strong converses, meaning that fidelity decreases exponentially when we try to use less communication or randomness than necessary. This is discussed in [84] and we will discuss a quantum analogue of this point in Sec. IV-E.

IV. Simulation of quantum channels on arbitrary inputs

This section is devoted to proving parts (d) and (e) of Theorem 3.

A. The case of flat spectra

By analogy with Sec. III-B, we will first state an unweighted or "flat" version of the quantum reverse Shannon theorem. We will then use a quantum version of type theory (based on Schur-Weyl duality) to extend this to prove the QRST for general inputs.

Definition 10. *An isometry $V^{A\to BE}$ is called* flat *if, when applied to half of a maximally entangled state $|\Phi\rangle^{RA}$, it produces a state $|\psi\rangle^{RBE}$ with ψ^R, ψ^B and ψ^E each maximally mixed.*

We note two features of the definition. First, the requirement that ψ^A be maximally mixed is satisfied automatically, but we include it to emphasize that each marginal of ψ should be maximally mixed. Second, the definition of a flat isometry does not depend on the choice of maximally entangled input $|\Phi\rangle$.

An important special case of flat channels occurs when A, B, E are irreps of some group G and V is a G-invariant map. We will return to this point in Sec. IV-C2.

Lemma 11 ([57], [1]). *Let A, B, E have dimensions D_A, D_B, D_E respectively. Consider furthermore quantum systems K_A, K_B, M with dimensions D_K, D_K, D_M, respectively, such that $D_B = D_K D_M$ and $D_M \geq \frac{256}{\delta\epsilon^4}\sqrt{D_A D_B / D_E}$. If $V^{A\to BE}$ is a flat isometry, it can be simulated up to error ϵ with respect to the maximally mixed input state, by consuming $\log(D_K)$ ebits and sending $\log(D_M)$ qubits from Alice to Bob. More precisely, there exist isometries $S_H^{K_B M\to B}, S_V^{AK_A\to ME}$ such that*

$$V^{A\to BE}|\Phi_{D_A}\rangle^{RA} \approx_\epsilon S_H^{K_B M\to B} S_V^{AK_A\to ME}|\Phi_{D_K}\rangle^{K_A K_B}|\Phi_{D_R}\rangle^{RA}. \quad (60)$$

Furthermore, S_H can be taken to be a Haar random unitary of dimension D_B with S_V chosen deterministically based on S_H and V. Eq. (60) holds with probability $\geq 1-\delta$ over the random choice of S_H.

The protocol in Eq. (60) is depicted in Fig. 9.

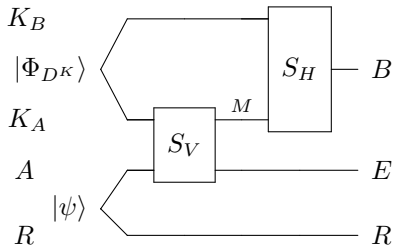

Fig. 9. The simulation of flat channels described in Lemma 11. The entangled state $|\Phi_{D_K}\rangle$ is consumed in order to simulate the action of $V^{A\to BE}$ on the A part of $|\psi\rangle^{RA}$. S_H is chosen from the Haar measure on $\mathcal{U}_{D_B}$, while S_V is chosen to depend on S_H and V, as described by [1]. While any $|\psi\rangle^{RA}$ can be input into the channel, the fidelity guarantee of Eq. (60) only holds when ψ^A is maximally mixed.

In an earlier unpublished version of this work, we proved a version of Lemma 11 using the measurement compression theorem of [85]. This version used classical instead of quantum communication (with correspondingly different rates), but by making the communication coherent in the sense of [40], [33] it is possible to recover Lemma 11.

However, a conceptually simpler proof of Lemma 11 was later given by [32], [57], [1]. This proof is based on reversing "state merging," which we can think of as the task of Bob sending a subsystem B to Alice in a way that preserves its correlation with a subsystem E which Alice already has, as well as with a purifying reference system R. In other words, merging is a state redistribution of the form

$$\Psi^{R:E:B} \to \Psi^{R:EB}. \quad (61)$$

The simplest proof of state merging is given in [1], where it is shown that if Bob splits B randomly into systems M and K_B of the appropriate sizes (i.e. by applying $S_H^\dagger$), and sends M to Alice, then Alice will be able to locally transform E, M into two subsystems A and K_A such that A is completely entangled with the reference system R (and thus can be locally transformed by Alice into E, B, the desired goal of the merging.). On the other hand K_A is nearly completely entangled with the remaining K_B system that Bob kept, so that it represents a byproduct of entanglement between Alice and Bob that has been generated by the protocol. When executed in reverse, the merging becomes splitting, and the $K_A K_B$ entanglement becomes a resource that is consumed, along with the quantum transmission of system M from Alice to Bob, in order to implement the state-splitting redistribution

$$\Psi^{R:A} \to \Psi^{R:EB} \to \Psi^{R:E:B}. \quad (62)$$

B. Tensor power inputs

We next need to reduce the general channel simulation problem to the problem of simulating flat channels. To get the idea of how this works, consider first the problem of simulating $\mathcal{N}^{\otimes n}$ on a tensor power input $\rho^{\otimes n}$. While several solutions to this problem have been previously described in [57], [32], [1] and this section is not strictly necessary for the proof of Theorem 3, we will present a protocol for tensor power inputs here in a way that will help us understand the general case.

Let $|\sigma\rangle^{ABE} = (I\otimes\mathcal{N})|\Phi_\rho\rangle^{AA'}$ and $|\psi\rangle = |\sigma\rangle^{\otimes n}$. Unfortunately, none of ψ^A, ψ^B nor ψ^E are in general maximally mixed. Even restricting to typical subspaces still leaves these states with eigenvalues that vary over a range of $2^{\pm O(\sqrt{n})}$.

On the other hand, these eigenvalues have a large amount of degeneracy. Let $\{|a_1\rangle, \ldots, |a_{d_A}\rangle\}$ be the eigenbasis of $\rho = \sigma^A$. Then the eigenvectors of ψ^A can be taken to be of the form $|a_{i_1}\rangle \otimes \cdots \otimes |a_{i_n}\rangle$, for $i = (i_1, \ldots, i_n) \in [d_A]^n$. Moreover the corresponding eigenvalue is determined entirely by the *type* t_A of i, just as in the classical case. There are $\binom{n+d_A-1}{n}$ such types. For fixed d_A, this number is polynomial in n, and thus the "which type" information can be transmitted using $O(\log n)$ qubits. Conditioned on this information, we are left with a flat spectrum over a space whose dimension depends on the type.

The same decomposition into types can be performed for the B and E systems, and for constant d_B and d_E we will still have at most $\mathrm{poly}(n)$ types t_B and t_E. Furthermore, we can decompose the action of $U_{\mathcal{N}}^{\otimes n}$ into a map from t_A to a superposition of t_B and t_E followed by a flat map within the type classes, which we call $V_{t_A t_B t_E}$. Thus, letting $\cong$ denote a global change of basis, we have

$$U_{\mathcal{N}}^{\otimes n} \cong \sum_{t_A, t_B, t_E} |t_B, t_E\rangle\langle t_A| \otimes V_{t_A, t_B, t_E}. \quad (63)$$

The only remaining question is to determine the communication rate. Here we can use the classical theory of types from Sec. III-C to argue that almost all of the weight of ψ is concentrated in strings with $\bar{t}_A, \bar{t}_B, \bar{t}_E$ close to the spectra of σ^A, σ^B and σ^E respectively. If "close" is defined to be distance δ, then ignoring the atypical types incurs error at most $\exp(-n\delta')$ and we are left with subspaces of dimensions $D_A = \exp(n(H(A)_\sigma \pm \delta''))$, $D_B = \exp(n(H(B)_\sigma \pm \delta''))$ and $D_E = \exp(n(H(E)_\sigma \pm \delta''))$, where δ', δ'' are constants depending on δ.[10] depending on δ), . Applying Lemma 11 we obtain the claimed communication rates of $\frac{H(A)+H(B)-H(E)}{2} =$

[10]These claims are based on standard methods of information theory. By "distance" δ we refer to the trace distance $\|\sigma^X - \bar{t}^X\|_1$. The error bound on ignoring atypical types is obtained from Eq. (57) and the bound on entropy is from the Fannes-Audenaert inequality (Eq. (58)).

$\frac{1}{2}I(A;B)$ qubits and $\frac{H(B)+H(E)-H(A)}{2} = \frac{1}{2}I(B;E)$ ebits per use of $\mathcal{N}$.

Two subtleties arise from combining communication protocols involving different input and output types. The first problem is that we have to be careful about who knows what when: unlike in the classical channel simulation protocol, Alice would like to communicate to Bob only t_B and not t_A or t_E. Indeed, she would like to forget t_A and retain only knowledge of t_E for herself. This is addressed by using the fact that Bob's decoding unitary S_H in Lemma 11 can be chosen to depend only on t_B, since we can choose a single S_H for each t_B, and w.h.p. Eq. (60) holds simultaneously for all t_A, t_E. Denote the resulting decoding map S_{H,t_B} and call Alice's encoding $S_{V_{t_A,t_B,t_E}}$. Then from Eqs. (60) and (63), we have that $U_{\mathcal{N}}^{\otimes n}$ can be approximately simulated by starting with an appropriate entangled state (more on this below) and applying

$$\left(\sum_{t_B} |t_B\rangle\langle t_B| \otimes S_{H,t_B}\right) \times \left(\sum_{t_A,t_B,t_E} |t_B, t_E\rangle\langle t_A| \otimes S_{V_{t_A,t_B,t_E}}\right), \quad (64)$$

where the first line is applied by Bob and the second line is applied by Alice.

The second problem is that when we apply Lemma 11 to the V_{t_A,t_B,t_E}, the dimensions D_A, D_B, D_E (and thus D_K, D_M as well) vary by as much as $\exp(\pm n\delta)$, and yet our protocol needs to act on a single entangled state and send a single message. For the M register we can address this by simply taking D_M to equal $\max\lceil \frac{256}{\epsilon^4}|T_{t_A}||T_{t_B}|/|T_{t_E}|\rceil$, where the maximum is taken over all typical triples of t_A, t_B, t_E. Thus, D_M is independent of any of the registers communicated during the protocol.

However, since $D_R D_M$ must equal $|T_{t_B}|$, we cannot avoid having D_R vary with t_B. (There is a minor technical point related to D_B needing to be an integer, but this can be ignored at the cost of an exponentially small error.) As a result, we need to run the protocol of Lemma 11 in superposition using different numbers of ebits in different branches of the superposition. This cannot be accomplished simply by discarding the unnecessary ebits in the branches of the superposition that need less entanglement; instead we need to use one of the techniques from Sec. II-C. Fortunately, since the number of ebits varies by only $O(n\delta)$ across different values of t_B, we only need to generate $O(n\delta)$ bits of entanglement spread. This can be done with $O(n\delta)$ extra qubits of communication, leading to an asymptotically vanishing increase in the communication rate. And since the amount of entanglement generated depends on only t_B, this does not require leaking any information to Bob that he will not already have. Alice first creates her half of the entanglement at the same time as she is transformating $|t_A\rangle$ into $|t_B, t_E\rangle$. Then she sends her half of the entanglement to Bob (after it has been mixed with the input by $S_{V_{t_A,t_B,t_E}}$) along with her copy of $|t_B\rangle$. This ensures that Alice keeps no record of the amount of entanglement she has created, while Bob is able to perform his part of the entanglement-generation protocol.

Earlier versions of the quantum reverse Shannon theorem did not need to mention this sublinear amount of entanglement spread because the extra sublinear communication cost could be handled automatically by the protocols used. However, when we consider non-tensor power inputs in Sec. IV-D we will need to make a more explicit accounting of the costs of entanglement spread. Thus, the reason our "warm-up" is more complicated than the previous proofs of the i.i.d.-source QRST is that it already contains much of the complexity of the full proof.

C. A quantum theory of types

There is one further difficulty which arises when considering non-tensor power inputs. This problem can already been seen in the case when the input to the channel is of the form $\frac{1}{2}(\rho_1^{\otimes n} + \rho_2^{\otimes n})$. If ρ_1 and ρ_2 do not commute, then we cannot run the protocol of the previous section without first estimating the eigenbases of ρ_1 and ρ_2. Moreover, we need to perform this estimation in a non-destructive way and then be able to uncompute our estimates of the eigenbasis, as well as any intermediate calculations used. Such techniques have been used to perform quantum data compression of $\rho^{\otimes n}$ when ρ is unknown [58], [11]. However, even for that much simpler problem they require delicate analysis. We believe that it is possible to prove the quantum reverse Shannon theorem by carefully using state estimation in this manner, but instead will present a somewhat simpler proof that makes use of representation theory.

The area of representation theory we will use is known as Schur duality (or Schur-Weyl duality). It has also been used for data compression of unknown tensor power states [49], [50], [45] and entanglement concentration from tensor powers of unknown pure entangled states [48], [51]. Some reviews of the role of Schur duality in quantum information can be found in Chapters 5 and 6 of [41] and Chapters 1 and 2 of [21]. A detailed explanation of the mathematics behind Schur duality can also be found in [37]. Our treatment will follow [41]. In Sec. IV-C1, we will explain how Schur duality can serve as a quantum analogue of the classical method of types that we described in Sec. III-C. Then in Sec. IV-C2 we will show this can be applied to channels, allowing us to decompose $\mathcal{N}^{\otimes n}$ into a superposition of flat channels. Finally, in Sec. IV-C3 we will use this to describe quantum analogues of conditional types. We will use this to show that the atypical flat sub-channels involve only an exponentially small amount of amplitude.

In Sec. IV-D, we will use these tools to prove Theorem 3.

1) Schur duality and quantum states: This section will review the basics of Schur duality and will explain how it can serve as a quantum analogue of the classical method of types. Let $\mathcal{S}_n$ denote the permutation group on n objects and let $\mathcal{U}_d$ denote the d-dimensional unitary group. Both groups have a natural action on $(\mathbb{C}^d)^{\otimes n}$. For $u \in \mathcal{U}_d$ define $\mathbf{Q}(u) = u^{\otimes n}$ and for $s \in \mathcal{S}_n$ define $\mathbf{P}(s)$ to permute the n systems according

to s: namely,

$$\mathbf{P}(s) = \sum_{i_1,\ldots,i_n \in [d]} |i_1, \ldots, i_n\rangle\langle i_{s(1)}, \ldots, i_{s(n)}|. \quad (65)$$

This convention is chosen so that $\mathcal{P}(s)$ is a representation[11] These two representations commute, and can be simultaneously decomposed into irreducible representations (a.k.a. irreps). We can also think of $\mathbf{Q}(u)\mathbf{P}(s)$ as a reducible representation of $\mathcal{U}_d \times \mathcal{S}_n$.

Define $\mathcal{I}_{d,n}$ to be the set of partitions of n into d parts: that is $\mathcal{I}_{d,n} = \{\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_d) \in \mathbb{Z}^d : \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d \geq 0 \text{ and } \sum_{i=1}^d \lambda_i = n\}$. Note that $|\mathcal{I}_{d,n}| \leq |\mathcal{T}_{d,n}| \leq (n+1)^d = \mathrm{poly}(n)$. It turns out that $\mathcal{I}_{d,n}$ labels the irreps of both $\mathcal{U}_d$ and $\mathcal{S}_n$ that appear in the decompositions of $\mathbf{Q}$ and $\mathbf{P}$. Define these representation spaces to be $\mathcal{Q}_\lambda$ and $\mathcal{P}_\lambda$ and define the corresponding representation matrices to be $\mathbf{q}_\lambda(u)$ and $\mathbf{p}_\lambda(s)$. Sometimes we write $\mathcal{Q}_\lambda^d$ or $\mathbf{q}_\lambda^d$ to emphasize the d-dependence; no such label is needed for $\mathcal{P}_\lambda$ since λ already determines n.

Schur duality states that $(\mathbb{C}^d)^{\otimes n}$ decomposes under the simultaneous actions of $\mathbf{Q}$ and $\mathbf{P}$ as

$$(\mathbb{C}^d)^{\otimes n} \cong \bigoplus_{\lambda \in \mathcal{I}_{d,n}} \mathcal{Q}_\lambda^d \otimes \mathcal{P}_\lambda \quad (66)$$

This means that we can decompose $(\mathbb{C}^d)^{\otimes n}$ into three registers: an irrep label λ which determines the actions of $\mathcal{U}_d$ and $\mathcal{S}_n$, a $\mathcal{U}_d$-irrep $\mathcal{Q}_\lambda$ and an $\mathcal{S}_n$-irrep $\mathcal{P}_\lambda$. Since the dimension of $\mathcal{Q}_\lambda$ and $\mathcal{P}_\lambda$ depends on λ, the registers are not in a strict tensor product. However, by padding the $\mathcal{Q}_\lambda$ and $\mathcal{P}_\lambda$ registers we can treat the λ, $\mathcal{Q}_\lambda$ and $\mathcal{P}_\lambda$ registers as being in a tensor product.

The isomorphism in Eq. (66) implies the existence of a unitary transform U_{Sch} that maps $(\mathbb{C}^d)^{\otimes n}$ to $\bigoplus_{\lambda \in \mathcal{I}_{d,n}} \mathcal{Q}_\lambda^d \otimes \mathcal{P}_\lambda$ in a way that commutes with the action of $\mathcal{U}_d$ and $\mathcal{S}_n$. Specifically we have that for any $u \in \mathcal{U}_d$ and any $s \in \mathcal{S}_n$,

$$U_{\text{Sch}} \mathbf{Q}(U)\mathbf{P}(s) U_{\text{Sch}}^\dagger = \sum_{\lambda \in \mathcal{I}_{d,n}} |\lambda\rangle\langle\lambda| \otimes \mathbf{q}_\lambda^d(U) \otimes \mathbf{p}_\lambda(s). \quad (67)$$

While we have described Schur duality in terms of the representation theory of $\mathcal{S}_n$ and the Lie group $\mathcal{U}_d$, there exists a similar relation between $\mathcal{S}_n$ and the general linear group GL_d. Indeed, $\mathbf{q}_\lambda(U)$ is a polynomial function of the entries of U (of degree $\sum_i \lambda_i$), and so can be extended to non-unitary and even non-invertible arguments. After doing so, one can show an analogue of Eq. (67) for tensor power states (taking $U = \rho$ and $s = \mathrm{id}$)

$$U_{\text{Sch}} \rho^{\otimes n} U_{\text{Sch}}^\dagger = \sum_{\lambda \in \mathcal{I}_{d,n}} |\lambda\rangle\langle\lambda| \otimes \mathbf{q}_\lambda^d(\rho) \otimes I_{\mathcal{P}_\lambda}. \quad (68)$$

So far we have not had to describe in detail the structure of the irreps of $\mathcal{U}_d$ and $\mathcal{S}_n$. In fact, we will mostly not need to do this in order to develop quantum analogues of the classical results from Sec. III-C. Here, the correct analogue of a classical type is in fact λ together with $\mathcal{Q}_\lambda$. Classically, we might imagine dividing a type $(t_1, \ldots, t_d)$ into a sorted list $t_1^\downarrow \geq \cdots \geq t_d^\downarrow$ (analogous to λ) and the $\mathcal{S}_d$ permutation that maps $t^\downarrow$ into t (analogous to the $\mathcal{Q}_\lambda$ register). Quantumly, we will see that for states of the form $\rho^{\otimes n}$, the λ register carries information about the eigenvalues of ρ and the $\mathcal{Q}_\lambda$ register is determined by the eigenbasis of ρ.

The main thing we will need to know about $\mathcal{Q}_\lambda$ and $\mathcal{P}_\lambda$ is their dimension. Roughly speaking, if d is constant then $|\mathcal{I}_{d,n}| \leq \binom{n+d-1}{n} \leq \mathrm{poly}(n)$, $\dim \mathcal{Q}_\lambda \leq \mathrm{poly}(n)$ and $\dim \mathcal{P}_\lambda \approx \exp(nH(\bar{\lambda}))$. For completeness, we also state exact formulas for the dimensions of $\mathcal{Q}_\lambda$ and $\mathcal{P}_\lambda$, although we will not need to use them. For $\lambda \in \mathcal{I}_{d,n}$, define $\tilde{\lambda} := \lambda + (d-1, d-2, \ldots, 1, 0)$. Then the dimensions of $\mathcal{Q}_\lambda^d$ and $\mathcal{P}_\lambda$ are given by [37]

$$\dim \mathcal{Q}_\lambda^d = \frac{\prod_{1 \leq i < j \leq d} (\tilde{\lambda}_i - \tilde{\lambda}_j)}{\prod_{m=1}^d m!} \quad (69)$$

$$\dim \mathcal{P}_\lambda = \frac{n!}{\tilde{\lambda}_1! \tilde{\lambda}_2! \cdots \tilde{\lambda}_d!} \prod_{1 \leq i < j \leq d} (\tilde{\lambda}_i - \tilde{\lambda}_j) \quad (70)$$

It is straightforward to bound these by [45], [23]

$$\dim \mathcal{Q}_\lambda^d \leq (n+d)^{d(d-1)/2} \quad (71)$$

$$\binom{n}{\lambda} (n+d)^{-d(d-1)/2} \leq \dim \mathcal{P}_\lambda \leq \binom{n}{\lambda}. \quad (72)$$

Applying Eq. (53) to Eq. (72) yields the more useful

$$\frac{\exp\left(nH(\bar{\lambda})\right)}{(n+d)^{d(d+1)/2}} \leq \dim \mathcal{P}_\lambda \leq \exp\left(nH(\bar{\lambda})\right). \quad (73)$$

To relate this to quantum states, let Π_λ denote the projector onto $\mathcal{Q}_\lambda^d \otimes \mathcal{P}_\lambda \subset (\mathbb{C}^d)^{\otimes n}$. Explicitly Π_λ is given by

$$\Pi_\lambda = U_{\text{Sch}}^\dagger \left(|\lambda\rangle\langle\lambda| \otimes I_{\mathcal{Q}_\lambda^d} \otimes I_{\mathcal{P}_\lambda} \right) U_{\text{Sch}}. \quad (74)$$

From the bounds on $\dim \mathcal{Q}_\lambda^d$ and $\dim \mathcal{P}_\lambda$ in Eqs. (71) and (73), we obtain

$$\frac{\exp\left(nH(\bar{\lambda})\right)}{(n+d)^{d(d+1)/2}} \leq \mathrm{Tr}\, \Pi_\lambda \leq \exp\left(nH(\bar{\lambda})\right) (n+d)^{d(d-1)/2} \quad (75)$$

As in the classical case, i.i.d. states have a sharply peaked distribution of λ values. Let $r = (r_1, \ldots, r_d)$ be the eigenvalues of a state ρ, arranged such that $r_1 \geq r_2 \geq \ldots$. For $\mu \in \mathbb{Z}^d$, define $r^\mu = r_1^{\mu_1} \cdots r_d^{\mu_d}$. As explained in Section 6.2 of [41], one can bound $\mathrm{Tr}\, \Pi_\lambda \rho^{\otimes n} = \mathrm{Tr}\, \mathbf{q}_\lambda^d(\rho) \cdot \dim \mathcal{P}_\lambda$ by

$$\begin{aligned} \exp\left(-nD(\bar{\lambda}\|r)\right) (n+d)^{-d(d+1)/2} &\leq \mathrm{Tr}\, \Pi_\lambda \rho^{\otimes n} \\ &\leq \exp\left(-nD(\bar{\lambda}\|r)\right) (n+d)^{d(d-1)/2} \end{aligned} \quad (76)$$

Similarly, we have $\Pi_\lambda \rho^{\otimes n} = \rho^{\otimes n} \Pi_\lambda = \Pi_\lambda \rho^{\otimes n} \Pi_\lambda$ and

$$\Pi_\lambda \rho^{\otimes n} \Pi_\lambda \leq r^\lambda \Pi_\lambda = \exp[-n(H(\bar{\lambda}) + D(\bar{\lambda}\|r))] \Pi_\lambda. \quad (77)$$

For some values of μ, r^μ can be much smaller, so we cannot express any useful lower bound on the eigenvalues of $\Pi_\lambda \rho^{\otimes n} \Pi_\lambda$, like we can with classical types. Of course, tracing out $\mathcal{Q}_\lambda^d$ gives us a maximally mixed state in $\mathcal{P}_\lambda$, and this is the quantum analogue of the fact that $p^{\otimes n}(\cdot|t)$ is uniformly distributed over T_t.

[11] The product of permutations s_1, s_2 is defined by $(s_1 \cdot s_2)(i) = s_1(s_2(i))$. Our definition in Eq. (65) is chosen so that $\mathbf{P}(s_1 \cdot s_2) = \mathbf{P}(s_1)\mathbf{P}(s_2)$.

We can also define the typical projector

$$\Pi^n_{r,\delta} = \sum_{\lambda:\|\bar{\lambda}-r\|_1\leq\delta} \Pi_\lambda \tag{78}$$

Using Pinsker's inequality, we find that

$$\operatorname{Tr}\Pi^n_{r,\delta}\rho^{\otimes n} \geq 1 - \exp\left(-\frac{n\delta^2}{2}\right)(n+d)^{d(d+1)/2}, \tag{79}$$

similar to the classical case. The typical subspace is defined to be the support of the typical projector. Its dimension can be bounded (using Eqs. (58) and (76)) by

$$\begin{aligned}\operatorname{Tr}\Pi^n_{r,\delta} &\leq |\mathcal{I}_{d,n}| \max_{\lambda:\|\bar{\lambda}-r\|_1\leq\delta} \operatorname{Tr}\Pi_\lambda \\ &\leq (n+d)^{d(d+1)/2}\exp(nH(r)+\eta(\delta)+n\delta\log d).\end{aligned} \tag{80}$$

2) Decomposition of memoryless quantum channels: The point of introducing the Schur formalism is to decompose $\mathcal{N}^{\otimes n}$ (or more accurately, its isometric extension $U_{\mathcal{N}}^{\otimes n}$) into a superposition of flat sub-channels. This is accomplished by splitting A^n, B^n and E^n each into $\lambda, \mathcal{Q}_\lambda$ and $\mathcal{P}_\lambda$ subsystems labelled $\lambda_A, \mathcal{Q}_{\lambda_A}, \mathcal{P}_{\lambda_A}, \lambda_B$, etc. Then the map from $\mathcal{P}_{\lambda_A} \to \mathcal{P}_{\lambda_B}\otimes\mathcal{P}_{\lambda_E}$ commutes with the action of $\mathcal{S}_n$ and as a result has the desired property of being flat.

To prove this more rigorously, a general isometry from $A^n \to B^nE^n$ can be written as a sum of terms of the form $|\lambda_B,\lambda_E\rangle\langle\lambda_A|\otimes|q_B,q_E\rangle\langle q_A|\otimes P_{\lambda_A,\lambda_B,\lambda_E}$, where $|q_A\rangle, |q_B\rangle, |q_E\rangle$ are basis states for the respective $\mathcal{Q}_\lambda$ registers and $P_{\lambda_A,\lambda_B,\lambda_E}$ is a map from $\mathcal{P}_{\lambda_A}\to\mathcal{P}_{\lambda_B}\otimes\mathcal{P}_{\lambda_E}$.

Since $U_{\mathcal{N}}^{\otimes n}$ commutes with the action of $\mathcal{S}_n$, it follows that each $P_{\lambda_A,\lambda_B,\lambda_E}$ must also commute with the action of $\mathcal{S}_n$. Specifically, for any $\lambda_A,\lambda_B,\lambda_E \in \mathcal{I}_{d,n}$ (with $d = \max(d_A,d_B,d_E)$) and any $s\in\mathcal{S}_n$, we have

$$\begin{aligned}&(\mathbf{p}_{\lambda_B}(s)\otimes\mathbf{p}_{\lambda_E}(s))\,P_{\lambda_A,\lambda_B,\lambda_E}\,\mathbf{p}_{\lambda_A}(s) = P_{\lambda_A,\lambda_B,\lambda_E}\\ &\mathbf{p}_{\lambda_A}(s)^\dagger P^\dagger_{\lambda_A,\lambda_B,\lambda_E}P_{\lambda_A,\lambda_B,\lambda_E}\mathbf{p}_{\lambda_A}(s) = P^\dagger_{\lambda_A,\lambda_B,\lambda_E}P_{\lambda_A,\lambda_B,\lambda_E}\end{aligned}$$

By Schur's Lemma $P^\dagger_{\lambda_A,\lambda_B,\lambda_E}P_{\lambda_A,\lambda_B,\lambda_E}$ is proportional to the identity on $\mathcal{P}_{\lambda_A}$. Therefore $P_{\lambda_A,\lambda_B,\lambda_E}$ is proportional to an isometry. Furthermore, $P_{\lambda_A,\lambda_B,\lambda_E}$ maps the maximally mixed state on $\mathcal{P}_{\lambda_A}$ to a state proportional to $P_{\lambda_A,\lambda_B,\lambda_E}P^\dagger_{\lambda_A,\lambda_B,\lambda_E}$. This state commutes with $\mathbf{p}_{\lambda_B}(s)\otimes\mathbf{p}_{\lambda_E}(s)$ for all $s\in\mathcal{S}_n$, and so, if we again use Schur's Lemma, we find that the reduced states on $\mathcal{P}_{\lambda_B}$ and $\mathcal{P}_{\lambda_E}$ are both maximally mixed. Therefore $P_{\lambda_A,\lambda_B,\lambda_E}$ is proportional to a flat isometry.

This is an example of a broader phenomenon. For vector spaces V_1, V_2, define $\operatorname{Hom}(V_1,V_2)$ to be the space of linear maps from V_1 to V_2. Note that $\operatorname{Hom}(V_1,V_2)\cong V_1^*\otimes V_2$, and if $(\mathbf{r}_1,V_1),(\mathbf{r}_2,V_2)$ are representations of a group G, then there is a representation $\mathbf{r}$ of G on $\operatorname{Hom}(V_1,V_2)$ given by $\mathbf{r}(g)T = \mathbf{r}_2(g)T\mathbf{r}_1(g^{-1})$. For a representation $(\mathbf{r},V)$ the G-invariant subspace V^G is defined by

$$V^G := \{|\psi\rangle\in V : \mathbf{r}(g)|\psi\rangle = |\psi\rangle\ \forall g\in G\}.$$

The space $\operatorname{Hom}(V_1,V_2)^G$ is precisely the set of linear operators from V_1 to V_2 that commute with the action of G. Using this notation, Schur's Lemma is equivalent to the statement that if V_1, V_2 are irreducible then $\operatorname{Hom}(V_1,V_2)^G$ is equal to $\{0\}$ if V_1, V_2 are inequivalent and is one-dimensional if V_1, V_2 are equivalent.

Using this language, we can observe that $P_{\lambda_A,\lambda_B,\lambda_E}$ belongs to $\operatorname{Hom}(\mathcal{P}_{\lambda_A},\mathcal{P}_{\lambda_B}\otimes\mathcal{P}_{\lambda_E})^{\mathcal{S}_n}$, i.e. the set of maps from $\mathcal{P}_{\lambda_A}$ to $\mathcal{P}_{\lambda_B}\otimes\mathcal{P}_{\lambda_E}$ that commute with $\mathcal{S}_n$. By the arguments in the paragraph before last, any isometry in $\operatorname{Hom}(\mathcal{P}_{\lambda_A},\mathcal{P}_{\lambda_B}\otimes\mathcal{P}_{\lambda_E})^{\mathcal{S}_n}$ must also be a flat isometry. There is a natural isomorphism from $(\mathcal{P}^*_{\lambda_A}\otimes\mathcal{P}_{\lambda_B}\otimes\mathcal{P}_{\lambda_E})^{\mathcal{S}_n}$ into $\operatorname{Hom}(\mathcal{P}_{\lambda_A},\mathcal{P}_{\lambda_B}\otimes\mathcal{P}_{\lambda_E})^{\mathcal{S}_n}$. We denote this isomorphism by S (making the $\lambda_A,\lambda_B,\lambda_E$-dependence implicit) and normalize S so that if $|\mu\rangle\in(\mathcal{P}^*_{\lambda_A}\otimes\mathcal{P}_{\lambda_B}\otimes\mathcal{P}_{\lambda_E})^{\mathcal{S}_n}$ is a unit vector then $S|\mu\rangle$ is a (flat) isometry. Below we will offer an operational interpretation of the $|\mu\rangle$ register.

To deal with the large numbers of registers, we now introduce some more concise notation.

Definition 12. *Let $P^{\lambda_B,\lambda_E}_{\lambda_A}$ be an orthonormal basis for $(\mathcal{P}^*_{\lambda_A}\otimes\mathcal{P}_{\lambda_B}\otimes\mathcal{P}_{\lambda_E})^{\mathcal{S}_n}$. We also let T_A denote the set of pairs (λ_A,q_A), where $|q_A\rangle$ runs over some fixed orthonormal basis of $\mathcal{Q}_{\lambda_A}$, and similarly we define T_B and T_E.*

Now we can represent $U_{\mathcal{N}}^{\otimes n}$ as

$$U_{\mathcal{N}}^{\otimes n} = \sum_{\substack{\tau_A\in T_A\\ \tau_B\in T_B\\ \tau_E\in T_E}}\ \sum_{\mu\in P^{\lambda_B,\lambda_E}_{\lambda_A}} [V^n_{\mathcal{N}}]^{\tau_A}_{\tau_B,\tau_E,\mu}|\tau_B,\tau_E\rangle\langle\tau_A|\otimes S|\mu\rangle \tag{81}$$

This is depicted as a quantum circuit in Fig. 10.

(We will not need to know anything more about the representation-theoretic structure of $P_{\lambda_A,\lambda_B,\lambda_E}$, but the interested reader can find a more detailed description of this decomposition of $U_{\mathcal{N}}^{\otimes n}$ in Section 6.4 of [41], where $S|\mu\rangle$ is related to the Clebsch-Gordan transform over $\mathcal{S}_n$.)

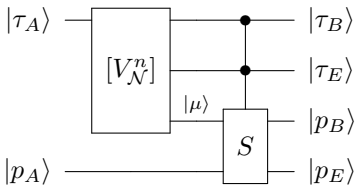

Fig. 10. The quantum channel $U_{\mathcal{N}}^{\otimes n}$ is decomposed in the Schur basis as in Eq. (81). Alice inputs an n qudit state of the form $|\tau_A\rangle|p_A\rangle$ and the channel outputs superpositions of $|\tau_B\rangle|p_B\rangle$ for Bob and $|\tau_E\rangle|p_E\rangle$ for Eve. The intermediate state $|\mu\rangle$ belongs to $(\mathcal{P}^*_{\lambda_A}\otimes\mathcal{P}_{\lambda_B}\otimes\mathcal{P}_{\lambda_E})^{\mathcal{S}_n}$. The figure suppresses the implicit $U_{\mathcal{N}}, n$-dependence of S, and expresses the $\lambda_A,\lambda_B,\lambda_E$-dependence of S by the control wires from the $|\tau_B\rangle$ and $|\tau_E\rangle$ registers.

We now have a situation largely parallel to the classical theory of joint types with τ_A,τ_B,τ_E representing the quantum analogues of types for systems A, B and E. Since τ_B,τ_E,μ together describe the joint type of systems BE, we can think of μ as representing the purely joint part of the type that is not contained in either of the marginal types. Further justifying the analogy with classical types is the fact that all but $\operatorname{poly}(n)$ dimensions are described by the flat isometries $P_{\lambda_A,\lambda_B,\lambda_E}$. Next we need to describe an analogue of jointly typical projectors, so that we can restrict our attention to triples of $(\lambda_A,\lambda_B,\lambda_E)$ that contribute non-negligible amounts of amplitude to $U_{\mathcal{N}}^{\otimes n}$. In the next section, we will argue that $[V^n_{\mathcal{N}}]^{\tau_A}_{\tau_B,\tau_E,\mu}$ is exponentially small unless $(\bar{\lambda}_A,\bar{\lambda}_B,\bar{\lambda}_E)$

correspond to the possible spectra of marginals of some state ψ^{RBE} that is obtained by applying $U_{\mathcal{N}}$ to a pure state on RA.

3) Jointly typical projectors in the Schur basis: In order for Eq. (81) to be useful, we need to control the possible triples (τ_A, τ_B, τ_E) that can have non-negligible weight in the sum. In fact, it will suffice to bound which triples $(\lambda_A, \lambda_B, \lambda_E)$ appear, since these determine the dimensions of the $\mathcal{P}_\lambda$ registers and in turn determine the dominant part of the communication cost. For large values of n, almost all of the weight will be contained in a small set of *typical* triples of $(\lambda_A, \lambda_B, \lambda_E)$. These triples are the quantum analogue of joint types from classical information theory.

Let ρ^A be an arbitrary channel input, and $|\psi\rangle^{ABE} = (I^A \otimes U_{\mathcal{N}}^{A'\to BE})|\Phi_\rho\rangle^{AA'}$ the purified channel output. Now define $R(\mathcal{N})$ to be set of ψ^{ABE} that can be generated in this manner. Further define $\mathcal{T}^*_{\mathcal{N}}$ to be $\{(r_A, r_B, r_E) : \exists \psi^{ABE} \in R(\mathcal{N}) \text{ s.t. } r_A = \operatorname{spec}(\psi^A), r_B = \operatorname{spec}(\psi^B), r_E = \operatorname{spec}(\psi^E)\}$. This set is simply the set of triples of spectra that can arise from one use of the channel. We will argue that it corresponds as well to the set of $(\bar{\lambda}_A, \bar{\lambda}_B, \bar{\lambda}_E)$ onto which a channel's input and output can be projected with little disturbance. Let $T^n_{\mathcal{N},\delta}$ denote the set

$$\{(\lambda_A, \lambda_B, \lambda_E) : \exists (r_A, r_B, r_E) \in \mathcal{T}^*_{\mathcal{N}}, \\ \|\bar{\lambda}_A - r_A\|_1 + \|\bar{\lambda}_B - r_B\|_1 + \|\bar{\lambda}_E - r_E\|_1 \leq \frac{\delta}{\log(d)}\} \quad (82)$$

One difficulty in defining joint types is that applying the projector Π_{λ_A} to the input may not commute with applying $\Pi_{\lambda_B} \otimes \Pi_{\lambda_E}$ to the output. Nevertheless, the following lemma (first proven in Section 6.4.3 of [41]) establishes a version of joint typicality that we can use.

Lemma 13 ([41])**.** *Let $d = \max(d_A, d_B, d_E)$. For any state $|\varphi\rangle^{RA}$ with $|\Psi\rangle = (I \otimes U_{\mathcal{N}})^{\otimes n}|\varphi\rangle^{\otimes n}$,*

$$\left\| |\Psi\rangle - \sum_{(\lambda_A,\lambda_B,\lambda_E)\in T^n_{\mathcal{N},\delta}} I \otimes ((\Pi_{\lambda_B} \otimes \Pi_{\lambda_E}) U_{\mathcal{N}}^{\otimes n} \Pi_{\lambda_A})|\varphi\rangle^{\otimes n} \right\| \\ \leq n^{O(d^2)} \exp\left(-n \frac{\delta^2}{8 \log^2(d)}\right). \quad (83)$$

For completeness, we include a proof in the appendix.

D. Reduction to the flat spectrum case

In this section we prove the coding theorem for the QRST. The outline of the proof is as follows:

1) We show that general inputs can be replaced by $\mathcal{S}_n$-invariant inputs by using a sublinear amount of shared randomness (which can be obtained from any of the other resources used in the protocol).
2) We show that $\mathcal{S}_n$-covariant channels (such as $\mathcal{N}^{\otimes n}$) decompose into a superposition of flat sub-channels. This is based on Sec. IV-C2. The simulation of these flat sub-channels on maximally mixed inputs is described in Sec. IV-A.
3) We show that atypical sub-channels can be ignored with negligible error (using Sec. IV-C3).
4) We paste together simulations of different flat channels using entanglement spread (introduced in Sec. II-C).

We now explain these components in more detail. First, we show how it is possible to assume without loss of generality that our inputs are $\mathcal{S}_n$-symmetric. If we did not mind using a large amount of shared randomness, then using $\log(n!)$ rbits would allow Alice to apply a random permutation $\pi \in \mathcal{S}_n$ to her inputs, and then for Alice to apply π^{-1} to the Eve output and for Bob to apply π^{-1} to his output. In some scenarios, these shared rbits might be a free resource (e.g. when entanglement is unlimited), and their cost could be further reduced by observing that they are incoherently decoupled from the protocol (using the terminology of [33]), and thus can be safely reused.

However, in fact, it is possible for Alice and Bob to safely sample π from a much smaller distribution. The idea is that the protocol has ϵ error on an $\mathcal{S}_n$-invariant input, which means that if the input is randomly permuted, then the average error will be ϵ. On the other hand, the diamond-norm error is never greater than 2. Standard concentration-of-measure arguments can then be used to show that $O(\log(n/\epsilon))$ rbits suffice to reduce the error to $O(\epsilon)$. This is detailed in Lemma 14.

For the rest of this section, we simply assume that Alice is given half of an $\mathcal{S}_n$-invariant input $|\varphi\rangle^{R^n A^n}$. Based on Sec. IV-C2, we can decompose the action of $U_{\mathcal{N}}^{\otimes n}$ into a map from τ_A to τ_B, τ_E, μ followed by a map from p_A, μ to p_B, p_E. The τ_B register has only $\operatorname{poly}(n)$ dimension, and can be transmitted uncompressed to Bob using $O(\log n)$ qubits. On the other hand, the map P_μ is flat, and therefore can be compressed using Lemma 11.

To understand the costs of compressing P_μ, we need to estimate the dimensions of the $\mathcal{P}_{\lambda_A}, \mathcal{P}_{\lambda_B}, \mathcal{P}_{\lambda_E}$ registers. In Sec. IV-C1, we showed that $\dim \mathcal{P}_\lambda \approx \exp(nH(\bar{\lambda}))$ up to $\operatorname{poly}(n)$ factors. So the cost of simulating a flat map from $\mathcal{P}_{\lambda_A}$ to $\mathcal{P}_{\lambda_B} \otimes \mathcal{P}_{\lambda_E}$ is $\frac{1}{2} n[H(\bar{\lambda}_A) + H(\bar{\lambda}_B) - H(\bar{\lambda}_E)] + O(\log n)$ qubits and $\frac{1}{2} n[H(\bar{\lambda}_B) + H(\bar{\lambda}_E) - H(\bar{\lambda}_A)] + O(\log n)$ ebits.

Next, we can relate these costs to entropic quantities. Using Lemma 13 from Sec. IV-C3, it follows that we need only consider the triples $(\bar{\lambda}_A, \bar{\lambda}_B, \bar{\lambda}_E)$ within distance $\delta/\log(d)$ of a spectral triple (r_A, r_B, r_E) corresponding to a possible channel output. Therefore, the problem of simulating $\mathcal{N}_F^{\otimes n}$ can be reduced to producing a superposition of

$$\left(\frac{1}{2} n I(R;B)_\rho + O(n\delta + \log n)\right)[q \to q] \\ + \left(\frac{1}{2} n I(E;B)_\rho + O(n\delta + \log n)\right)[qq] \quad (84)$$

for all possible single-letter ρ (i.e. ρ that are inputs to a single channel use). If we take $\delta \to 0$ as $n \to \infty$ then this corresponds to an asymptotic rate of

$$\frac{1}{2} I(R;B)_\rho [q \to q] + \frac{1}{2} I(E;B)_\rho [qq] \quad (85)$$

per channel use. The resulting protocol is depicted in Fig. 11.

Finally, if our input is not a known tensor power source, then producing Eq. (84) in superposition may require entanglement

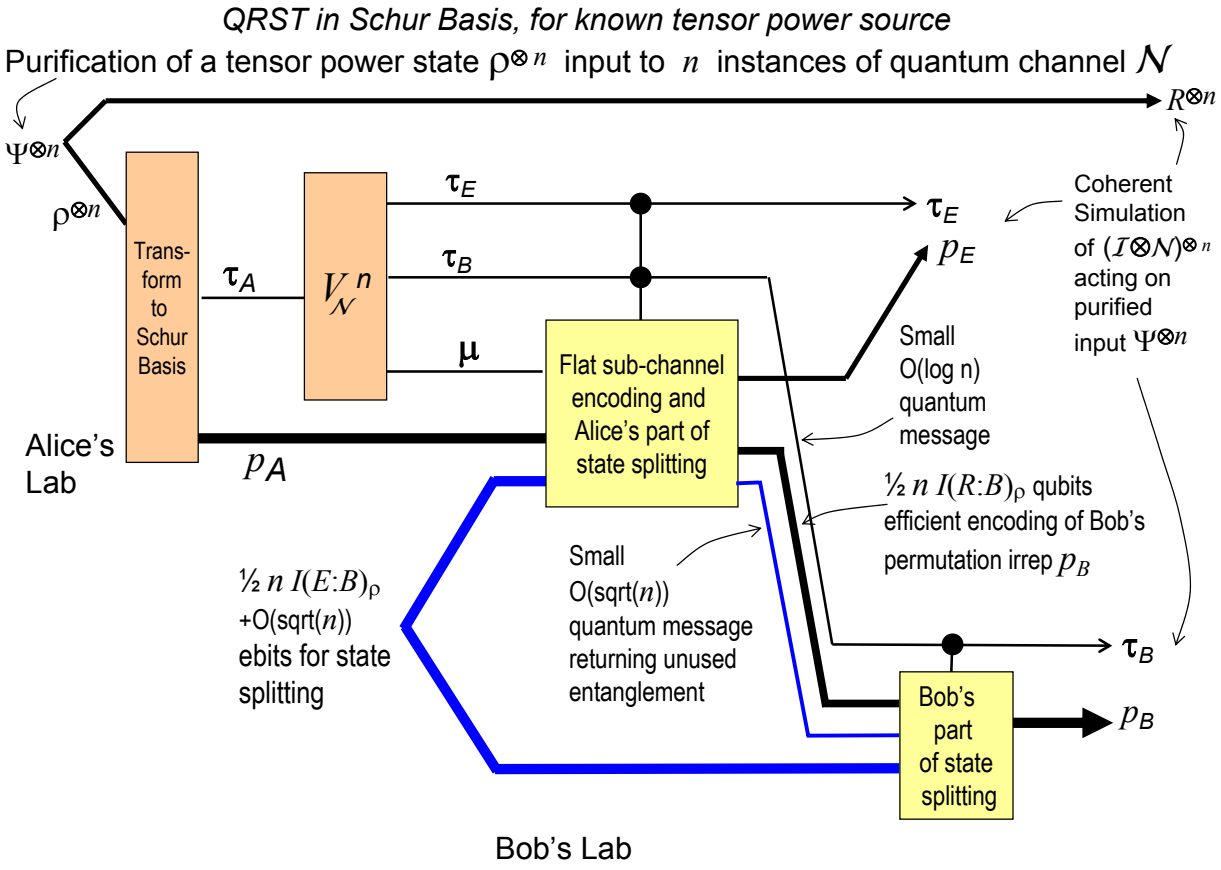

Fig. 11. Quantum protocol for quantum reverse Shannon theorem on a known tensor power source. Alice transforms the tensor power input into the Schur representation, comprising a small τ register containing the quantum type and a large p register containing the permutation irrep. These registers, together with a slight ($O(\sqrt{n})$) excess of halves of ebits shared with Bob, are coherently transformed into about $\frac{1}{2}nI(R;B)$ qubits worth of flat sub-channel codes representing Bob's p register, which Bob decodes with the help of the other halves of the shared ebits and the small τ_B register sent from Alice. Alice also returns the ($O(\sqrt{n})$) unused halves of ebits, allowing them to be coherently destroyed. The remaining registers τ_E and p_E, representing Eve's share of the output, remain with Alice, as required for a quantum feedback simulation of the channel $\mathcal{N}^{\otimes n}$. By discarding them into the environment, one obtains a (not necessarily efficient) non-feedback simulation.

spread. Suppose that $\alpha \geq \beta$ and

$$\beta \geq_{\text{CL}} \frac{1}{2}I(R;B)_\rho[q \to q] + \frac{1}{2}I(E;B)_\rho[qq]$$

for all ρ. Then we can prepare β from α and then use β to produce the resources needed to simulate $\langle \mathcal{N}_F : \rho \rangle$ in superposition across all ρ (or equivalently across all τ_A in the input). This can be done using extra forward communication (in which case the protocol still qualitatively resembles Fig. 11, but the $O(\sqrt{n})$ message with extra entanglement becomes $\Omega(n)$ qubits), using an embezzling state (as depicted in Fig. 12) or using backward communication (as depicted in Fig. 13). The protocol with backwards communication appears to require a temporary shuttling of the small τ_B register from Alice to Bob and back before finally sending it to Bob; otherwise backward communication is used to coherently reduce entanglement the same way that forward communication is.

When we do not need to simulate feedback, the main difference is that we can split the E register into a part for Alice (E_A) and a part for Bob (E_B). Additionally, this splitting is not restricted to be i.i.d., although the corresponding "additivity" question here remains open. That is, for any $n \geq 1$ and any $V : E^n \to E_A E_B$, simulating the action of $\mathcal{N}^{\otimes n}$ can be achieved by simulating $V \circ \mathcal{N}_F^{\otimes n}$. Here Alice gets the output E_A and Bob gets the output $B^n E_B$. Moreover, we are in some cases able to break the superpositions between different τ_A. If feedback is not required, then we can assume without loss of generality that Alice has measured τ_E, estimated $\hat{\mathcal{N}}(\rho)$ to within $O(n^{-1/2})$ accuracy [59] and communicated the resulting estimate to Bob using $o(n)$ communication.

However, in some cases (including an example we will describe in the next section), $\hat{\mathcal{N}}(\rho)$ does not uniquely determine $\mathcal{N}(\rho)$, and thereby determine the rate of entanglement needed. In this case, it will suffice to prepare a superposition of entanglement corresponding to any source in $(\hat{\mathcal{N}}^{\otimes n})^{-1}(\omega)$ for each $\omega \in \text{range}(\hat{\mathcal{N}}^{\otimes n})$. This yields the communication cost claimed in Theorem 3.

We conclude with a rigorous proof that low average-case error can be turned into low worst-case error, allowing the permutation π in Figs. 12 and 13 to be largely derandomized, reducing its shared randomness cost to sublinear in n.

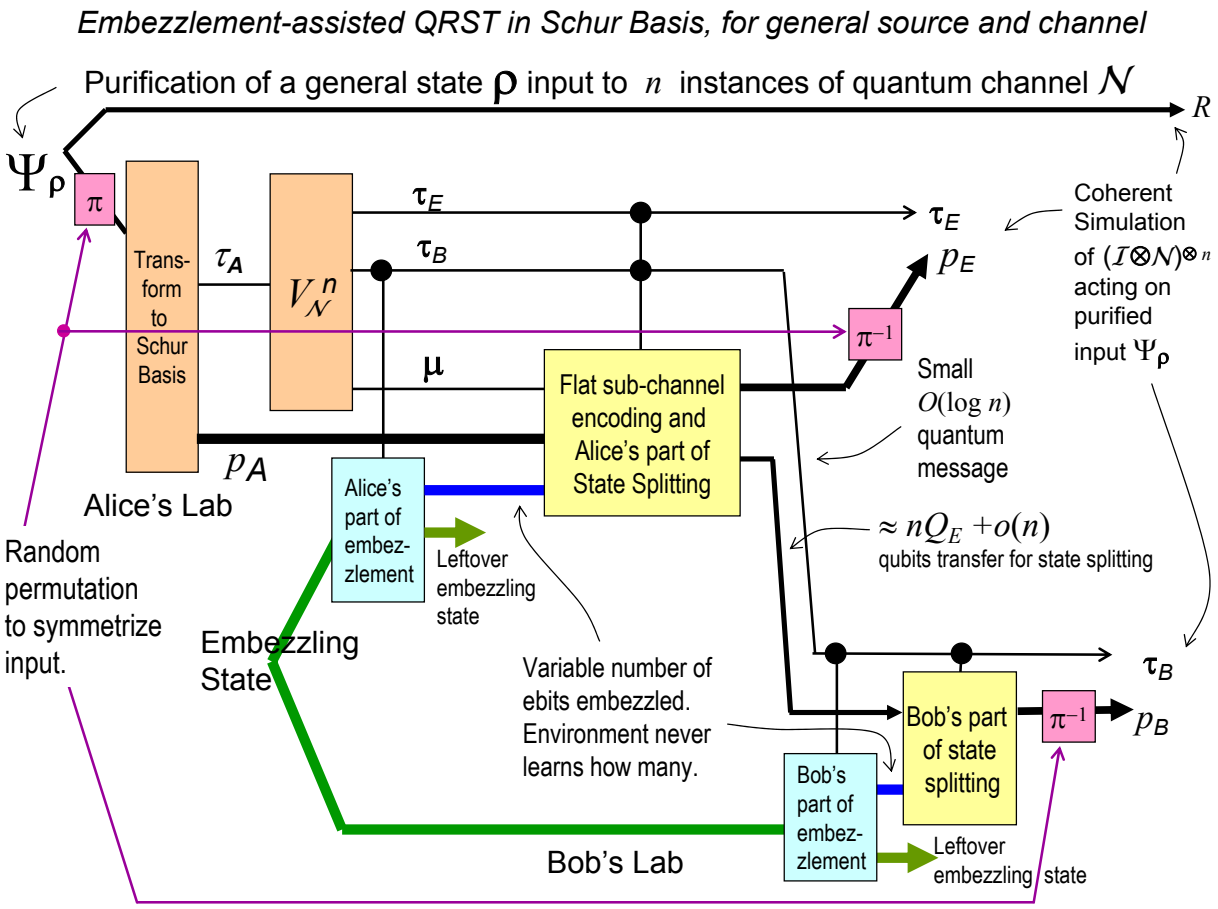

Fig. 12. QRST on a general input using an entanglement-embezzling state (green). Alice first applies a randomizing permutation π to the inputs to n instances of her quantum channel, using information shared with Bob (magenta), thereby rendering the overall input approximately permutation-symmetric. She then uses the τ_B register to embezzle the correct superposition of (possibly very) different amounts of entanglement needed by her sub-channel encoder, leaving a negligibly degraded embezzling state behind. At the receiving end (lower right) Bob performs his half of the embezzlement, coherently decodes the sub-channel codes, and undoes the random permutation. The shared randomness needed for the initial randomizing permutation can also be obtained from the embezzling state, and in any case can be made sublinear in n, as shown in Lemma 14.

Lemma 14. *Let $V^{A\to BE}$ be an isometry that represents an ideal protocol and $\tilde{V}^{A\to BE}$ its approximate realization. Suppose that we have an average-case fidelity guarantee of the form*

$$\langle \Phi_{D_A}|^{RA}(I\otimes \tilde{V}^\dagger V)|\Phi_{D_A}\rangle^{RA} \geq 1-\epsilon. \tag{86}$$

Let μ be a distribution over $\mathcal{U}_{D_A}$ such that $\mathbb{E}_{U\sim\mu} U\rho U^\dagger = I/D_A$ for any density matrix ρ. If $U_1,\ldots,U_m$ are drawn i.i.d. from μ, then with probability $\geq 1-D_A(4/e)^{-m\epsilon/2}$, for any $|\psi\rangle$,

$$\frac{1}{m}\sum_{i=1}^m |\langle\psi|(I\otimes U_i^\dagger \tilde{V}^\dagger V U_i)|\psi\rangle|^2 \geq 1-6\epsilon. \tag{87}$$

Proof: Let $\Delta := I-\tilde{V}^\dagger V$. Observe that $\|\Delta\|_\infty \leq 2$ and that Eq. (86) implies that $\|\Delta\|_1 = \operatorname{Tr}\Delta \leq \epsilon D_A$. Define

$$\Delta_0 := \mathbb{E}_{U\sim\mu}[U\Delta U^\dagger] = \frac{\operatorname{Tr}\Delta}{D_A} I \leq \epsilon I \tag{88}$$

$$\bar{\Delta} := \frac{1}{m}\sum_{i=1}^m U_i \Delta U_i^\dagger, \tag{89}$$

where $U_1,\ldots,U_m$ are drawn i.i.d. from μ. By applying the operator Chernoff bound [5], [76], we find that with probability $\geq 1-D_A(4/e)^{-m\epsilon/2}$, we have $\|\bar{\Delta}-\Delta_0\|_\infty \leq 2\epsilon$. In this case, $\|\bar{\Delta}\|_\infty \leq 3\epsilon$. Eq. (87) follows by Cauchy-Schwarz. ■

Lemma 14 can be applied separately to each Schur subspace, with $D_A := \dim \mathcal{P}_{\lambda_a}$. Thus, a union bound multiplies the probability of a bad choice of permutations by $n^{O(d_A)}$ and we always have $D_A \leq O(n\log d_A)$.

Inefficiencies and errors: Here we briefly tabulate the various sources of inefficiency and error in our simulation protocols for quantum channels. We will consider allowing an inefficiency of $O(n\delta)$ in each step of the protocol and will analyze the resulting errors.

1) In Lemma 11, an extra communication rate of $O(n\delta)$ means that the splitting step incurs an error of $\exp(-n\delta)$.
2) When restricting to typical triples, our definition of $T^n_{\mathcal{N},\delta}$ was chosen so that entropic quantities such as $H(R^n)+H(B^n)-H(E^n)$ would change by $\leq n\delta + o(n)$. According to Lemma 13, this results in error $\exp(-n\delta^2/8\log^2(d))$. This will turn out to be the

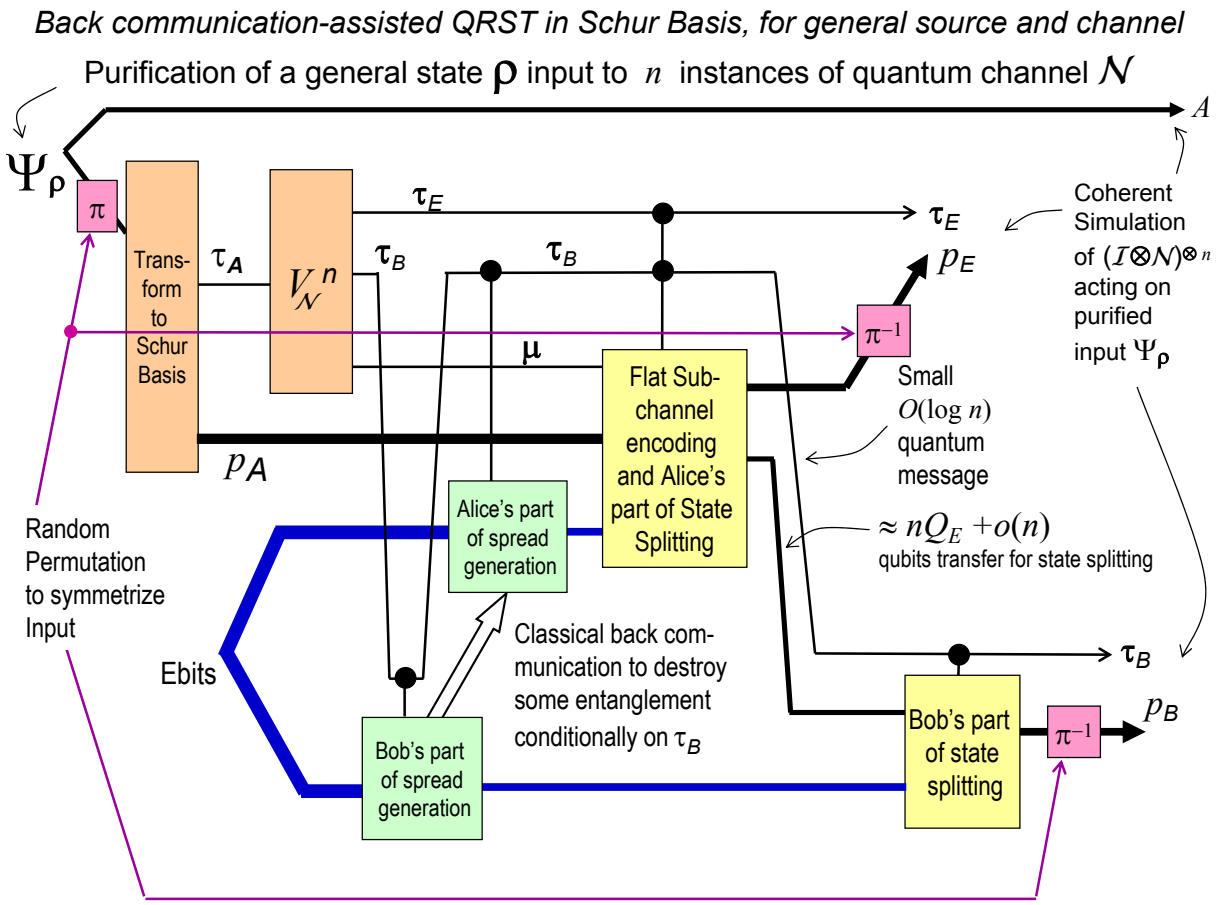

Fig. 13. QRST using classical back communication. Here the requisite spread is generated by starting with a large amount of ordinary (i.e. maximal) entanglement, then using back communication and the τ_B register to coherently burn off some of it. This requires the τ_B register to make a round trip from Alice to Bob then back again, before finally returning to Bob, who needs to be holding it at the end. Other aspects of the protocol are as in the embezzlement-assisted implementation of the preceding figure.

dominant error term; accordingly, we define $\delta' = \delta^2/8\log^2(d)$.

3) Suppose Alice and Bob use $\log(m)$ rbits to permute the channel inputs, and unpermute the channel outputs. Then Lemma 14 implies that the error multiplies by only a constant factor if we take $m = O(n\delta + \log(n))$.
4) To achieve an error of $\exp(-n\delta)$ using an embezzling state, we need to take it to have $n\exp(n\delta)$ qubits. This is exorbitant, but in some secenarios, such as our proof of the strong converse in the next section, the size of the entangled ancilla that we use is irrelevant.

To summarize, our error scales as $\exp(-n\delta^2/8\log^2(d) + o(n))$.

E. Converses and strong converses

In information theory, a "converse" is the statement of asymptotic resource optimality of a coding theorem (which is often called "direct part"). A "strong converse" is a statement of the form that with too little resources the error parameter in any protocol approaches 1 asymptotically. In channel simulations, the first and foremost resource is forward communication, but other resources of interest are the amount of entanglement and specifically the amount of entanglement spread.

Here, we first show that as with the classical reverse Shannon theorem, the existence of a coding theorem (this time for entanglement-assisted capacity [14]) means that no better simulation is possible. Indeed, such matching coding theorems generally give us *strong* converses, implying that attempting to simulate a channel at rate $C_E - \delta$ or lower results in an error $\geq 1 - \exp(-n\delta')$ for some $\delta' > 0$. At the same time, they give us strong converses for coding theorems, proving that attempting to code at a rate $C_E + \delta$ results in the probability of successful decoding being $\leq \exp(-n\delta')$, again for some δ' depending on δ. Second, we use arguments from the theory of entanglement spread (cf. Sec. II-C) to show that our simulations for non-IID inputs do require either embezzling states, or – if only maximally entangled states are available – the use of extra communication (which may be forward or backward directed), to create entanglement spread.

1) Strong converse for forward communication: The general principle behind these strong converses is based on the fact that m forward cbits, assisted by arbitrary back communication and entanglement, can transmit $m + k$ bits only with success

probability $\leq 2^{-k}$. The proof is folklore.[12] We call this principle *the guessing bound*; it is also sometimes referred to as "causality." To apply the guessing bound, suppose we have a coding theorem that allows us to use $\mathcal{N}^{\otimes n}$ (perhaps also with auxiliary resources, such as shared entanglement) to send $n(C-\delta)$ bits with success probability $1-\epsilon_{n,\delta}$. [Typically, $\epsilon_{n,\delta}$ will be of the form $\exp(-O(n\delta^2))$.] Now assume that there exists a simulation of $\mathcal{N}^{\otimes n}$ (using any auxiliary resources that are not capable of forward communication) that uses $n(C-\delta')$ bits of communication and achieves error $\epsilon'_{n,\delta'}$. If $\delta' > \delta$, then the guessing bound implies that

$$\epsilon_{n,\delta} + \epsilon'_{n,\delta'} \geq 1 - 2^{-n(\delta'-\delta)}. \tag{90}$$

Thus coding theorems constrain possible simulation theorems. Vice versa, by the same logic, suppose we had a simulation of $\mathcal{N}^{\otimes n}$ using $n(C+\delta)$ cbits of forward communication plus additional resources and error $\epsilon_{n,\delta}$, and consider a coding of $n(C+\delta')$ cbits into n uses of $\mathcal{N}$ and auxiliary resources, achieving error probability $\epsilon'_{n,\delta'}$. Then as before, for $\delta' > \delta$,

$$\epsilon_{n,\delta} + \epsilon'_{n,\delta'} \geq 1 - 2^{-n(\delta'-\delta)}. \tag{91}$$

For the purposes of this argument, any auxiliary resources are permitted as long as they are consistent with the guessing bound. In particular, embezzling states of unlimited size are allowed, and so is backwards quantum communication, and in this way we can also establish whatever type of entangled state we need.

In this case, the arguments of the last section established that an inefficiency of δ in our channel simulation (i.e. spending $n(C_E+\delta)$ bits) allows errors to be bounded by $\leq \exp(-n\delta^2/8\log^2(d) + o(n))$. Similarly, it is known that $n(C_E-\delta)$ bits can be sent through $\mathcal{N}^{\otimes n}$ with error $\leq \exp(-n\delta^2/8\log^2(d))$ if we are allowed a sufficient rate of ebits. This establishes that $C_E(\mathcal{N})$ is the optimal cbit rate for simulation, and of communication, in the strong converse sense, even if arbitrary entangled states and back communication are for free. Previously this was known to hold only when considering product-state inputs [68], [83] or restricted classes of channels [61], [82]. Recently an alternate proof of the entanglement-assisted strong converse has also been given based on a more direct argument involving completely bounded norms [38]. The fact that our strong converse also applies in the setting where free back communication is allowed from Bob to Alice is perhaps surprising given that back communication is known to increase the classical capacity in the unassisted case [72] (although not in the assisted case [18]). One limitation of our strong converses is that they only apply when the entanglement-assisted capacity is exceeded, whereas [68], [83], [61], [82] addressed the Holevo capacity or the ordinary classical capacity.

While the above argument applies to arbitrary use of the channel to communicate (and allows arbitrary input states in the simulation), we can also establish such a strong converse in the case of a known IID input. Here it is not only known [33] that $\langle \mathcal{N}_F : \rho \rangle \geq \frac{1}{2} I(R;B)[q \to q] + \frac{1}{2} I(B;E)[qq]$, but the corresponding protocol can be shown to have error bounded by $2^{-n\delta'}$. Thus, suppose a simulation existed for $\langle \mathcal{N} : \rho \rangle$ that used $\frac{1}{2}(I(R;B)-\delta)[q \to q]$ and an unlimited amount of entanglement and back communication to achieve fidelity f. Then combining this simulation with teleportation and our coding protocol would give a method for using cbits at rate $I(R;B)-\delta$ together with entanglement to simulate cbits at rate $I(R;B)-\delta/2$ with fidelity $\geq f - 2^{-n\delta'}$ for some $\delta' > 0$. By causality, any such simulation must have fidelity $\leq 2^{-n\delta/2}$, and thus we must have $f \leq 2^{-n\delta/2} + 2^{-n\delta'}$.

2) Converses for the use of entanglement and back communication, based on spread: Here we have to distinguish between the channel simulation with and without coherent feedback.

The case *with* coherent feedback is easier to handle as it places more stringent constraints on the protocol, and so the bounds are easier to prove. Thus we begin with this case, which corresponds to part (d) of Theorem 3.

We shall argue that entanglement spread is necessary. In fact, we will show a larger communication cost (forward plus backward) if the only entangled resource consists of ebits (i.e. maximally entangled states). Recall that the simulation theorem for feedback channels uses communication at rate

$$\max_{\rho_1} H(B)_{\rho_1} + \max_{\rho_2} (H(R) - H(E))_{\rho_2} = C_E(\mathcal{N}) + \Delta_{\text{sim}}(\mathcal{N}). \tag{92}$$

In what follows, we will omit $\mathcal{N}$ from our notation. We will show that this rate is optimal by constructing an input on which $U_{\mathcal{N}}^{\otimes n}$ will create $\approx n(C_E + \Delta_{\text{sim}})$ spread.

For $i = 1, 2$, let ρ_i be the states from Eq. (92) and let $|\psi_i\rangle$ be a purification of $\rho_i^{\otimes n}$. Let

$$|\Psi\rangle = \frac{|1\rangle^A |1\rangle^B |\psi_1\rangle^{A'A} |\psi_2\rangle^{AB} + |2\rangle^A |2\rangle^B |\psi_2\rangle^{A'B}}{\sqrt{2}}.$$

Here we use A repeatedly to indicate registers under Alice's control, B to indicate registers owned by Bob, and A' for a register controlled by Alice that will be input to $U_{\mathcal{N}}^{\otimes n}$. We omit describing the $|0\rangle$ registers that should pad Alice and Bob's registers so that each branch of the superposition has the same number of qubits. Let $|\varphi_i\rangle^{RBE} = (U_{\mathcal{N}}^{A' \to BE} \otimes I^R)^{\otimes n} |\psi_i\rangle^{A'R}$. Then,

$$\begin{aligned} |\Theta\rangle &:= U_{\mathcal{N}}^{\otimes n} |\Psi\rangle \\ &= \frac{|1\rangle^A |1\rangle^B |\varphi_1\rangle^{ABE} |\psi_2\rangle^{AB} + |2\rangle^A |2\rangle^B |\varphi_2\rangle^{BBE}}{\sqrt{2}}. \end{aligned} \tag{93}$$

Here, E is again a register controlled by Alice and we observe that $|\varphi_2\rangle$ has two out of its three registers controlled by Bob.

We argue that $\approx n(C_E + \Delta_{\text{sim}})$ spread has been created by applying $U_{\mathcal{N}}^{\otimes n}$. First, observe that $|\Psi\rangle$ is locally equivalent to $\frac{|1,1\rangle + |2,2\rangle}{\sqrt{2}} \otimes |\psi_2\rangle^{AB}$, which has $O(\sqrt{n})$ spread. More precisely,

[12] Here is a sketch of the proof. Suppose a protocol exists that achieves success probability q on a randomly chosen $m+k$-bit input. Modify this protocol so that the bits transmitted from Alice to Bob are replaced by random bits that Bob generates locally. We can think of this as Bob guessing Alice's input. This modified protocol can be simulated locally by Bob and corresponds to him drawing from a fixed distribution independent of Alice's input. On a random $m+k$-bit input, this must have success probability 2^{-m-k}. Our bound on the original protocol also means that this has success probability $\geq q2^{-m}$, since Bob has probability 2^{-m} of correctly guessing Alice's m transmitted bits. Thus we obtain $q \leq 2^{-k}$.

$\Delta_\epsilon(\psi_2^A) \leq O(\sqrt{n\log(1/\epsilon)}\log(d))$, and so $|\Psi\rangle$ can be prepared with error ϵ using this amount of communication [64]. Next, we argue that $|\Theta\rangle$ has a large amount of spread. The part attached to the $|1,1\rangle$ register has entanglement roughly equal to $n(H(B)_{\rho_1} + H(R)_{\rho_2})$ and the part attached to the $|2,2\rangle$ register has entanglement roughly equal to $nH(E)_{\rho_2}$. Since these registers are combinations of i.i.d. states, one can prove (c.f. Theorem 13 of [52], or Proposition 6 of [44]) that for any $\epsilon < 1/2$, $\Delta_\epsilon(\Theta^A) \geq n(H(B)_{\rho_1} + H(R)_{\rho_2} - H(E)_{\rho_2}) - O(\sqrt{n})$. We conclude from Theorem 5 that simulating $U_{\mathcal{N}}^{\otimes n}$ to error lower than a sufficiently small constant (such as 10^{-4}) using unlimited ebits requires communication $\geq n(C_E(\mathcal{N}) + \Delta_{\text{sim}}(\mathcal{N}) - o(1))$. We suspect, but do not prove, that a tighter analysis could prove this lower bound for all $\epsilon < 1/2$. Note that our statement is not a strong converse in the usual sense (which would demand a proof of our bound for all $\epsilon < 1$) but that it still establishes a jump of the error at the optimal rate.

When we consider simulations without feedback, we no longer have additivity, and we are able only to establish regularized coding theorems and weak converses. The zero-entanglement limit is discussed in [46] and the low-entanglement regime (part (b) of Theorem 3) follows similar lines. The main idea is that if only coherent resources (such as qubits and ebits) are used, then the state of the environment is entirely comprised of what Alice and Bob discard. Let E_A (resp. E_B) denote the system that Alice (resp. Bob) discards.

Let $\mathcal{P}$ denote the simulation of $\mathcal{N}^{\otimes n}$ constructed by the protocol. By the above arguments, $\mathcal{P}$ has an isometric extension $U_{\mathcal{P}}^{A^n \to B^n E_A E_B}$, just as $\mathcal{N}^{A\to B}$ has isometric extension $U_{\mathcal{N}}^{A\to BE}$. The fact that the simulation is successful means that $\|\mathcal{P} - \mathcal{N}^{\otimes n}\|_\diamond \leq \epsilon$. We now make use of a generalization of Uhlmann's theorem [62] to show that

$$\|U_P - V^{E^n \to E_A E_B} \circ U_{\mathcal{N}}^{\otimes n}\|_\diamond \leq \sqrt{\epsilon} \tag{94}$$

for some isometry V.

For part (b) of Theorem 3, this allows us to reduce the converse to that for part (a). We obtain Eqs. (20) and (21) from Fannes' inequality. Before discarding E_B, Bob's total state $B^n E_B$ is within ϵ of a state on $q+e$ qubits, and thus has $H(B^n E_B) \leq q + e + O(n\epsilon)$. Similarly, Bob has received only q qubits, so we must have $\frac{1}{2} I(R^n; B^n E_B) \leq q + O(n\epsilon)$.

For part (e), Eq. (94) allows us to reduce the converse to the converse for part (d). Again this is because the ability to simulate $\mathcal{N}^{\otimes n}$ without preserving E is equivalent to the ability to simulate $V \cdot U_{\mathcal{N}}^{\otimes n}$ for *some* choice of $V^{E^n \to E_A E_B}$.

3) The clueless Eve channel: We conclude our discussion of converses with an explicit example of a channel that requires more communication to simulate with ebits than with embezzling states [part (e) of Theorem 3]. This channel is designed so that different inputs create different amounts of entropy for the receiver, but without leaking information about this to the environment. Hence, we call it the "clueless Eve channel."

The channel $\mathcal{N}_d$ maps $d+1 \to d+1$ dimensions. We define it in terms of its isometric extension as follows:

$$U_{\mathcal{N}_d} = |\Phi_d\rangle^{BE}\langle 0|^A + \sum_{i=1}^{d} |0\rangle^B |i\rangle^E \langle i|^A, \tag{95}$$

where $|\Phi_d\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i,i\rangle$. In other words $|0\rangle$ is mapped to the maximally mixed state (over dimensions $1,\ldots,d$) for Bob, while $|i\rangle$ is mapped to the $|0\rangle$ state for $1 \leq i \leq d$. One can show that $C_E(\mathcal{N}_d) = 2Q_E(\mathcal{N}_d) = 1$ independent of d using convexity and symmetry arguments[13] along the lines of [25], [55]. However, the following argument will show that on some (non-tensor-power) inputs, the channel's ebit-assisted simulation cost, even for a non-feedback simulation, strictly exceeds its entanglement-assisted capacity. (This may be contrasted with the case of the amplitude damping channel considered earlier in Fig. 6, where the gap between ebit-assisted simulation cost and C_E is present only for feedback simulation). To see qualitatively why standard ebits are an insufficient entanglement resource to efficiently simulate this channel, consider the purified non-tensor-power input

$$|\Psi\rangle^{RA^n} := \frac{|0^n\rangle^R |0^n\rangle^{A^n} + |\Phi_{d^n}\rangle^{RA^n}}{\sqrt{2}} \tag{96}$$

to n uses of the channel. As usual, R is a reference system and A^n is sent to $B^n E^n$ by $U_{\mathcal{N}}^{\otimes n}$.

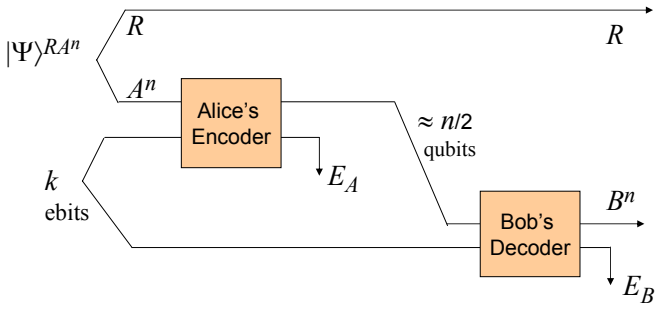

Fig. 14. A would-be simulation of the clueless Eve channel $\mathcal{N}_d$ on the non-tensor-power source $\left|\Psi^{RA^n}\right\rangle$, using around $nQ_E = n/2$ qubits of forward communication and k ordinary ebits as the entanglement resource, deposits different amounts of entropy in Alice's local environment E_A depending on which term in Ψ^{RA^n} is acted upon, thereby decohering the superposition and spoiling the simulation.

In Fig. 14 Alice's encoder, assisted by some number k of ebits, transforms the A^n part of this input into a supposedly small ($\approx n/2$ qubit) quantum message sent to Bob and a residual environment system E_A retained by Alice. By conservation of entropy, if $A^n = 0$, then Alice's environment E_A will have entropy at most $k + n/2$, whereas if $A^n \neq 0$ it will be left in a different state with entropy at least $k + n\log(d) - n/2$. Because (as will be shown in the following

[13] C_E is given by the maximum of $I(A;B)$ over inputs ρ. Due to the structure of the channel, it is invariant under the map $\rho \to U\rho U^\dagger$ for any U satisfying $U|0\rangle = |0\rangle$. Since $I(A:B)$ is concave in the input density matrix ρ, it follows that it can be maximized by ρ that commutes with all such U. The resulting states have the form $p|0\rangle\langle 0| + (1-p)(\sum_{i=1}^{d} |i\rangle\langle i|/d)$ and the resulting one-parameter maximization problem is easily seen to be equivalent to determining the entanglement-assisted capacity of a noiseless classical bit (i.e. totally dephasing) channel.

theorem) these two states become close to orthogonal for large d, Alice's environment will gain information about which term in the superposition Ψ^{RA^n} was present, and consequently will decohere the superposition, which a faithful simulation of the channel would not have done. Carrying through this argument more precisely, we have:

Theorem 15. *Let $\mathcal{P}$ be a protocol using q qubits of communication (total, in either direction) and k ebits. If $\|\mathcal{P}-\mathcal{N}_d^{\otimes n}\|_\diamond \leq \epsilon$ then $q \geq \frac{1}{4}n\log d - 1 - 8\sqrt{\epsilon}n\log d$.*

Proof: We begin by introducing some notation. The proof depends on the assumed closeness between $\mathcal{P}$ and $\mathcal{N}_d^{\otimes n}$ on the non-tensor-power source Ψ^{RA^n}. Applying $U_{\mathcal{N}}^{\otimes n}$ to the A^n part of Ψ^{RA^n} and using the fact that $|\Phi_d\rangle^{\otimes n} = |\Phi_{d^n}\rangle$ we obtain the (ideal) state

$$|\Theta\rangle^{RB^nE^n} := \frac{|0\rangle^R|\Phi_{d^n}\rangle^{B^nE^n} + |0^n\rangle^{B^n}|\Phi_{d^n}\rangle^{RE^n}}{\sqrt{2}} \quad (97)$$

that would result from the operation of the channel. We now purify the simulating protocol $\mathcal{P}$ in a canonical way: all non-unitary operations are replaced by isometries and discarding subsystems, and each subsystem that Alice (resp. Bob) discards is added to a register called E_A (resp. E_B). We can think of E_A, E_B as local environments. Let $U_{\mathcal{P}}$ denote the resulting purification and define the actual protocol output to be

$$\left|\tilde{\Theta}\right\rangle^{RB^nE_AE_B} := (I\otimes U_{\mathcal{P}})|\Psi\rangle. \quad (98)$$

To simulate $\mathcal{N}_d^{\otimes n}$ on input $|\Psi\rangle$ we do not need to approximate $|\Theta\rangle$, but it suffices to approximate the RB^n part of the state. By Uhlmann's theorem, this is equivalent to the claim that there exists an isometry $V : E^n \to E_AE_B$ such that

$$\left|\left\langle\tilde{\Theta}\middle|\Theta_V\right\rangle\right|^2 \geq 1-\epsilon, \quad (99)$$

where $|\Theta_V\rangle^{RB^nE_AE_B} := (I^{RB^n}\otimes V^{E^n\to E_AE_B})|\Theta\rangle$.

To prove the lower bound, we will argue that either $|\Theta_V\rangle$ has high spread for any V, or it has high mutual information between R and B^nE_B. Either way, we obtain a lower bound on the communication required to approximately create it. We first sketch the idea of why this should be true. Let $V|\Phi_{d^n}\rangle^{XE^n} := |\varphi_V\rangle^{XE_AE_B}$, where X can be either R or B^n. Then

$$|\Theta_V\rangle^{RB^nE_AE_B} = \frac{|0^n\rangle^R|\varphi_V\rangle^{B^nE_AE_B} + |0^n\rangle^{B^n}|\varphi_V\rangle^{RE_AE_B}}{\sqrt{2}}$$

$$\Theta_V^{RE_A} = \frac{|0\rangle\langle 0|^R\otimes\varphi_V^{E_A} + \varphi_V^{RE_A}}{2}.$$

To understand the spectrum of $\Theta_V^{RE_A}$, observe that $|0\rangle\langle 0|\otimes\varphi_V^{E_A}$ and $\varphi_V^{RE_A}$ have orthogonal support. Therefore, if $\varphi_V^{E_A}$ and $\varphi_V^{RE_A}$ have spectrum $\alpha = (\alpha_1\ldots,\alpha_a)$ and $\beta = (\beta_1,\ldots,\beta_b)$ respectively, then $\Theta_V^{RE_A}$ has eigenvalues

$$\frac{\alpha_1}{2},\ldots,\frac{\alpha_a}{2},\frac{\beta_1}{2},\ldots,\frac{\beta_b}{2}.$$

Note also that $\varphi_V^{RE_A}$ has the same spectrum as $\varphi_V^{E_B}$. Thus

$$\begin{aligned}1+\max(H_{0,\epsilon}(\varphi_V^{E_A}),H_{0,\epsilon}(\varphi_V^{E_B})) &\geq H_{0,\epsilon}(\Theta_V^{RE_A})\\ &\geq \max(H_{0,2\epsilon}(\varphi_V^{E_A}),H_{0,2\epsilon}(\varphi_V^{E_B})) \quad (100a)\end{aligned}$$

$$\begin{aligned}1+\min(H_{\infty,\epsilon}(\varphi_V^{E_A}),H_{\infty,\epsilon}(\varphi_V^{E_B})) &\leq H_{\infty,\epsilon}(\Theta_V^{RE_A})\\ &\leq 1+\min(H_{\infty,2\epsilon}(\varphi_V^{E_A}),H_{\infty,2\epsilon}(\varphi_V^{E_B})). \quad (100b)\end{aligned}$$

Pretend for a moment that $\epsilon = 0$. In that case $H_0(\Theta_V^{RE_A}) \geq H_0(\varphi_V^{E_A}) \geq H(\varphi_V^{E_A}) = H(\tilde{\Theta}^{E_A})$. Combining this with the fact that $H_\infty \leq S$, we have that $\Delta_0(\Theta_V^{RE_A}) \geq H(\tilde{\Theta}^{E_A}) - H(\tilde{\Theta}^{RE_A}) = -H(R|E_A)_{\tilde{\Theta}} = H(R)_{\tilde{\Theta}} - I(R;B^nE_B)_{\tilde{\Theta}}$. Rearranging we have that the sum of the spread ($\Delta_0(\Theta_V^{RE_A})$) and the mutual information ($I(R;B^nE_B)_{\tilde{\Theta}}$) is at least $H(R)_{\tilde{\Theta}} = 1+\frac{1}{2}n\log d$. Since spread and mutual information are both $\leq 2q$, we obtain the desired result.

The difficulty in extending this argument to the $\epsilon > 0$ case is (a) that spread can vary dramatically under small perturbations in the state (as observed even in situations as simple as entanglement dilution [64], [52], [44]), and (b) that the dimensions of E_A, E_B are unbounded, and so Fannes' inequality is difficult to apply. The second difficulty is easiest to address: we will use a variant of Fannes' inequality known as the Alicki-Fannes inequality, which bounds the variation of $H(R|E_A)$ using only $|R|$ and not $|E_A|$.

Lemma 16 (Alicki-Fannes inequality [6])**.** *If $\epsilon := \frac{1}{2}\|\rho^{XY} - \sigma^{XY}\|_1 < 1/2$ then*

$$|H(X|Y)_\rho - H(X|Y)_\sigma| \leq 8\epsilon\log|X| + 2H_2(2\epsilon) \quad (101)$$

To address the unbounded Lipschitz constant of Δ_0, we will need to look more carefully at how Θ_V and $\tilde{\Theta}$ are related. First, we replace φ_V with a low-spread approximation. From Eq. (35), we obtain nonnegative operators $M_{A,0}, M_{A,\infty}, M_{B_0}, M_{B,\infty}$ whose largest eigenvalues are ≤ 1 and that satisfy

- $1-2\epsilon = \operatorname{Tr} M_{A,0}\varphi^{E_A} = \operatorname{Tr} M_{A,\infty}\varphi^{E_A}$
- $1-2\epsilon = \operatorname{Tr} M_{B,0}\varphi^{E_B} = \operatorname{Tr} M_{B,\infty}\varphi^{E_B}$
- $M_{A,0}, M_{A,\infty}, \varphi^{E_A}$ all commute
- $M_{B,0}, M_{B,\infty}, \varphi^{E_B}$ all commute
- $H_0(M_{A,0}\varphi^{E_A}) = H_{0,2\epsilon}(\varphi^{E_A})$
- $H_\infty(M_{A,\infty}\varphi^{E_A}) = H_{\infty,2\epsilon}(\varphi^{E_A})$
- $H_0(M_{B,0}\varphi^{E_B}) = H_{0,2\epsilon}(\varphi^{E_B})$
- $H_\infty(M_{B,\infty}\varphi^{E_B}) = H_{\infty,2\epsilon}(\varphi^{E_B})$

We can now define

$$|\hat{\varphi}\rangle := \gamma^{-1/2}(I\otimes\sqrt{M_{A,0}M_{A_\infty}}\otimes\sqrt{M_{B,0}M_{B_\infty}})|\varphi_V\rangle, \quad (102)$$

where $\gamma \geq 1-8\epsilon$ is a normalizing constant, chosen so that $\langle\hat{\varphi}|\hat{\varphi}\rangle = 1$. For ease of calculations, we will choose $\epsilon \leq 1/16$, so that $\log(1-8\epsilon) \geq -1$. Then

$$H_0(\hat{\varphi}^{E_A}) \leq H_{0,2\epsilon}(\varphi_V^{E_A}) \quad (103a)$$

$$H_0(\hat{\varphi}^{E_B}) \leq H_{0,2\epsilon}(\varphi_V^{E_B}) \quad (103b)$$

$$H_\infty(\hat{\varphi}^{E_A}) \geq H_{\infty,2\epsilon}(\varphi_V^{E_A}) + 1 \tag{104a}$$
$$H_\infty(\hat{\varphi}^{E_B}) \geq H_{\infty,2\epsilon}(\varphi_V^{E_B}) + 1 \tag{104b}$$

Now we use $|\hat{\varphi}\rangle$ to define

$$\left|\hat{\Theta}\right\rangle^{RB^nE_AE_B} := \frac{|0^n\rangle^R|\hat{\varphi}\rangle^{B^nE_AE_B} + |0^n\rangle^{B^n}|\hat{\varphi}\rangle^{RE_AE_B}}{\sqrt{2}}. \tag{105}$$

Observe that $\left\langle \Theta_V \middle| \hat{\Theta} \right\rangle = \langle \varphi_V | \hat{\varphi} \rangle = \sqrt{\gamma} \geq \sqrt{1-8\epsilon}$, implying $\frac{1}{2}\|\Theta_V - \hat{\Theta}\|_1 \leq \sqrt{8\epsilon}$. Combined with Eq. (99), we obtain

$$\frac{1}{2}\|\tilde{\Theta} - \hat{\Theta}\|_1 \leq 4\sqrt{\epsilon}. \tag{106}$$

The advantage of $\hat{\Theta}$ is that it has exactly the same structure as Θ_V, but with $|\varphi_V\rangle$ replaced with $|\hat{\varphi}\rangle$. Thus it similarly satisfies

$$\hat{\Theta}^{E_A} = \hat{\varphi}^{E_A} \quad \text{and} \quad \hat{\Theta}^{E_B} = \hat{\varphi}^{E_B}, \tag{107}$$

and thus Eq. (100) still holds when Θ_V is replaced with $\hat{\Theta}$ and φ_V is replaced with $\hat{\varphi}$.

We now conclude with a traditional chain of entropic inequalities, with each step labeled by its justification:

$$\begin{aligned}
2q \geq & \Delta_0(\tilde{\Theta}^{RE_A}) \\
\geq & \Delta_\epsilon(\Theta^{RE_A})) && \text{Lemma 6} \\
\geq & H_{0,\epsilon}(\Theta^{RE_A}) - H_{\infty,\epsilon}(\Theta^{RE_A}) && \text{Eq. (34)} \\
\geq & \max(H_{0,2\epsilon}(\varphi_V^{E_A}), H_{0,2\epsilon}(\varphi_V^{E_B})) && \text{Eq. (100a)} \\
& - \min(H_{\infty,2\epsilon}(\varphi_V^{E_A}), H_{\infty,2\epsilon}(\varphi_V^{E_B})) - 1 && \text{Eq. (100b)} \\
\geq & \max(H_0(\hat{\varphi}_V^{E_A}), H_0(\hat{\varphi}_V^{E_B})) && \text{Eq. (103)} \\
& - \min(H_\infty(\hat{\varphi}_V^{E_A}), H_\infty(\hat{\varphi}_V^{E_B})) - 3 && \text{Eq. (104)} \\
\geq & H_0(\hat{\varphi}^{E_A}) - H_\infty(\hat{\Theta}^{RE_A}) - 3 && \text{Eq. (100b)} \\
= & H_0(\hat{\Theta}^{E_A}) - H_\infty(\hat{\Theta}^{RE_A}) - 3 && \text{Eq. (107)} \\
\geq & H(\hat{\Theta}^{E_A}) - H(\hat{\Theta}^{RE_A}) - 3 && H_0 \geq S \geq H_\infty \\
= & -H(R|E_A)_{\hat{\Theta}} - 3 \\
\geq & -H(R|E_A)_{\tilde{\Theta}} - 32\sqrt{\epsilon} n \log d - 5 && \text{Lemma 16} \\
= & -H(R|E_A)_{\tilde{\Theta}} - \delta && \delta := 32\sqrt{\epsilon} n \log d + 5 \\
= & H(R)_{\tilde{\Theta}} - I(R;E_A)_{\tilde{\Theta}} - \delta \\
= & \left(1 + \frac{1}{2} n \log d\right) - I(R;E_A)_{\tilde{\Theta}} - \delta \\
\geq & \left(1 + \frac{1}{2} n \log d\right) - 2q - \delta
\end{aligned}$$

■

V. Conclusion

We conclude by summarizing the operational and technical consequences of our work, as well as some open problems.

Operationally, we establish necessary and sufficient amounts of standard noiseless resources for simulation of discrete memoryless quantum channels, including classical DMCs as a special case. As is usual in Shannon theory, simulations become efficient and faithful only in the limit of large block size, even in cases where the simulation capacity is given by a single-letter formula. We consider both ordinary and feedback simulations, a feedback simulation being one in which the simulating sender coherently retains what the simulated channel would have discarded into its environment. We consider simulations on both tensor power sources (the quantum generalization of classical IID sources) and general sources, which may be correlated or entangled over the multiple inputs, a distinction that becomes important for quantum channels. We also establish conditions for asymptotic equivalence among channels, that is conditions under which channels can simulate one another efficiently and reversibly, so that the capacity for channel $\mathcal{M}$ to simulate $\mathcal{N}$ is the reciprocal of that for performing the simulation in the opposite direction. Such equivalences generally hold only in the presence of some combination of auxiliary resources, which by themselves would have no capacity for channel simulation. In each case, an unlimited supply of the auxiliary resources enables asymptotically reversible cross-simulation. For cross-simulations among classical channels, shared randomness is a necessary and sufficient auxiliary resource. For quantum channels on tensor power sources, ordinary shared entanglement is necessary and sufficient. For quantum channels on general sources, more general entangled states ("entanglement-embezzling states") or combinations of resources, such as entanglement and classical back-communication, are required. Finally, in many cases of interest, we quantify the loss of efficiency and reversibility when an auxiliary resource is insufficient or absent. In this respect, we feel that our Theorem 15 is not giving a tight bound, due to an imperfect proof technique. One problem is that mutual information and spread are not placed on a common footing, as they are in the case when simulating an isometry (feedback case).

On the technical side, we can now understand quantum simulations of quantum channels in terms of three key ingredients:

1) *State splitting* (also known as the reverse of state merging [57], [1]) in which a *known* tripartite state Ψ^{ABC} begins with C held by Alice and ends with C held by Bob. Note that this is a coherent version of measurement compression [85], upon which early QRST proofs were based.
2) *Entanglement spread*, which measures how far a state is from maximally entangled on some subspace [42], and turns out to be necessary when protocols requiring different numbers of ebits need to be executed in superposition.
3) *Dividing the environment between Alice and Bob*, which starts with the "Church of the Larger Hilbert Space" principle that mixed states have purifications, and proceeds to the observation that in a protocol using only noiseless resources any simulated environment must WLOG be divided between the sender and receiver. This form of the idea first appeared in [75] and is necessary to understand the low-entanglement versions of the QRST.

These concepts were known to the quantum information theory community separately in various contexts, but find their common use in the QRST.

A number of interesting open questions remain. On the technical side, we observe that for classical channels, Lemma 7 gives low error in the worst case, but for quantum channels,

Lemma 11 only gives average-case bounds. While this can be addressed by using shared randomness catalytically (and thereby without increasing the overall cost of the protocol), a more direct proof would be preferable.

More ambitiously, we observe that the classical and quantum reverse Shannon theorems are incomparable because the assistance of shared entanglement is stronger than the assistance of shared randomness even for purely classical channels (cf the discussion at the end of Sec. II-B). This is in contrast to the fact that Shannon's noisy coding theorem can be viewed as a special case of the entanglement-assisted capacity theorem. It would be desirable to have a single theorem that stated the cost of simulating a channel given the assistance of an arbitrary rate of randomness and entanglement. Some encouraging progress in this direction is given by [81], [17], which shows that for QC channels (i.e. quantum input, classical output) shared randomness can be used in place of shared entanglement. Another direction for generalization is to consider simulations that use side information, along the lines of [90], [80].

There are also new questions about additivity and regularization that arise when considering low-entanglement simulations of quantum channels, most of which are completely open. For example, the zero-entanglement point on the tradeoff curve corresponds to the entanglement of purification [75] whose additivity properties are still open (but see [20] for recent work suggesting that it is not additive).

VI. Acknowledgments

We wish to acknowledge helpful discussions with Paul Cuff, Patrick Hayden, Jonathan Oppenheim, Graeme Smith, John Smolin and Mark Wilde.

References

[1] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter. The mother of all protocols: Restructuring quantum information's family tree. *Proc. Roc. Soc. A*, 465(2108):2537–2563, 2009, arXiv:quant-ph/0606225.

[2] C. Adami and N. J. Cerf. von Neumann capacity of noisy quantum channels. *Phys. Rev. A*, 56:3470–3483, Nov 1997, arXiv:quant-ph/9609024.

[3] R. Ahlswede and I. Csiszár. Common Randomness in Information Theory and Cryptography Part I: Secret Sharing. *IEEE Trans. Inf. Theory*, 39(4):1121–1132, 1993.

[4] R. Ahlswede and I. Csiszár. Common Randomness in Information Theory and Cryptography Part II: CR capacity. *IEEE Trans. Inf. Theory*, 44(1):225–240, 1998.

[5] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *IEEE Trans. Inf. Theory*, 48(3):569–579, 2002, arXiv:quant-ph/0012127.

[6] R. Alicki and M. Fannes. Continuity of quantum conditional information. *J. Phys. A*, 37:L55–L57, 2004, arXiv:quant-ph/0312081.

[7] K. M. R. Audenaert. A sharp continuity estimate for the von Neumann entropy. *J. Phys. A*, 40(28):8127, 2007. quant-ph/0610146.

[8] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher. Concentrating partial entanglement by local operations. *Phys. Rev. A*, 53:2046–2052, 1996, arXiv:quant-ph/9511030.

[9] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.*, 70:1895–1899, 1993.

[10] C. H. Bennett, I. Devetak, P. W. Shor, and J. A. Smolin. Inequalities and separations among assisted capacities of quantum channels. *Phys. Rev. Lett.*, 96:150502, 2006, arXiv:quant-ph/0406086.

[11] C. H. Bennett, A. W. Harrow, and S. Lloyd. Universal quantum data compression via gentle tomography. *Phys. Rev. A*, 73:032336, 2006, arXiv:quant-ph/0403078.

[12] C. H. Bennett, P. Hayden, D. W. Leung, P. W. Shor, and A. J. Winter. Remote preparation of quantum states. *IEEE Trans. Inf. Theory*, 51(1):56–74, 2005, arXiv:quant-ph/0307100.

[13] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. Thapliyal. Entanglement-assisted classical capacity of noisy quantum channels. *Phys. Rev. Lett.*, 83:3081–3084, 1999, arXiv:quant-ph/9904023.

[14] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. Thapliyal. Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem. *IEEE Trans. Inf. Theory*, 48:2637–2655, 2002, arXiv:quant-ph/0106052.

[15] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Phys. Rev. Lett.*, 69:2881–2884, 1992.

[16] M. Berta, M. Christandl, and R. Renner. A conceptually simple proof of the quantum reverse Shannon theorem. *Comm. Math. Phys.*, 306(3):579–615, 2011, arXiv:0912.3805.

[17] M. Berta, J. M. Renes, and M. M. Wilde. Identifying the information gain of a quantum measurement, 2013, arXiv:1301.1594.

[18] G. Bowen. Quantum feedback channels. *IEEE Trans. Inf. Theory*, 50(10):2429–2434, 2004, arXiv:quant-ph/0209076.

[19] G. Bowen. Feedback in quantum communication. *Int. J. Quant. Info.*, 3(01):123–127, 2005, arXiv:quant-ph/0410191.

[20] J. Chen and A. Winter. Non-additivity of the entanglement of purification (beyond reasonable doubt), 2012, arXiv:1206.1307.

[21] M. Christandl. *The structure of bipartite quantum states: Insights from group theory and cryptography*. PhD thesis, University of Cambridge, 2006, arXiv:quant-ph/0604183.

[22] M. Christandl, R. Koenig, and R. Renner. Post-selection technique for quantum channels with applications to quantum cryptography. *Phys. Rev. Lett.*, 102:020504, 2009, arXiv:0809.3019.

[23] M. Christandl and G. Mitchison. The spectra of density operators and the Kronecker coefficients of the symmetric group. *Commun. Math. Phys.*, 261(3):789–797, 2006, arXiv:quant-ph/0409016.

[24] M. Christandl and A. Winter. Uncertainty, monogamy, and locking of quantum correlations. *Information Theory, IEEE Transactions on*, 51(9):3159 – 3165, sept. 2005, arXiv:quant-ph/0501090.

[25] J. Cortese. Holevo-Schumacher-Westmoreland channel capacity for a class of qudit unital channels. *Phys. Rev. A*, 69:022302, Feb 2004, arXiv:quant-ph/0211093.

[26] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Series in Telecommunication. John Wiley and Sons, New York, 1991.

[27] T. S. Cubitt, D. Leung, W. Matthews, and A. Winter. Zero-error channel capacity and simulation assisted by non-local correlations. *IEEE Trans. Inf. Theory*, 57(8):5509–5523, Aug 2011, arXiv:1003.3195.

[28] P. Cuff. Communication requirements for generating correlated random variables. *Proc. IEEE Symp. on Info. Th.*, 2008, arXiv:0805.0065.

[29] P. W. Cuff, H. H. Permuter, and T. M. Cover. Coordination capacity. *Information Theory, IEEE Transactions on*, 56(9):4181–4206, 2010, arXiv:0909.2408.

[30] S. Daftuar and P. Hayden. Quantum state transformations and the Schubert calculus. *Annals of Physics*, 315(1):80–122, 2005, arXiv:quant-ph/0410052.

[31] N. Datta, M.-H. Hsieh, and M. Wilde. Quantum rate distortion, reverse shannon theorems, and source-channel separation. *IEEE Trans. Inf. Theory*, 59(1):615–630, 2013, arXiv:1108.4940.

[32] I. Devetak. Triangle of dualities between quantum communication protocols. *Phys. Rev. Lett.*, 97(14):140503, Oct 2006, arXiv:quant-ph/0505138.

[33] I. Devetak, A. Harrow, and A. Winter. A resource framework for quantum Shannon theory. *IEEE Trans. Inf. Theory*, 54(10):4587–4618, Oct 2008, arXiv:quant-ph/0512015.

[34] I. Devetak, M. Junge, C. King, and M. B. Ruskai. Multiplicativity of completely bounded p-norms implies a new additivity result. *Commun. Math. Phys*, 266:37–63, 2006, arXiv:quant-ph/0506196.

[35] I. Devetak and J. Yard. Exact cost of redistributing multipartite quantum states. *Phys. Rev. Lett.*, 100:230501, 2008, arXiv:quant-ph/0612050.

[36] M. Fannes. A continuity property of the entropy density for spin lattices. *Commun. Math. Phys.*, 31:291–294, 1973.

[37] R. Goodman and N. Wallach. *Representations and Invariants of the Classical Groups*. Cambridge University Press, 1998.

[38] M. K. Gupta and M. M. Wilde. Multiplicativity of completely bounded p-norms implies a strong converse for entanglement-assisted capacity, 2013, arXiv:1310.7028.

[39] T. S. Han and S. Verdú. Approximation theory of output statistics. *IEEE Trans. Inf. Theory*, 39(3):752–752, 1993.

[40] A. W. Harrow. Coherent communication of classical messages. *Phys. Rev. Lett.*, 92:097902, 2004, arXiv:quant-ph/0307091.

[41] A. W. Harrow. *Applications of coherent classical communication and Schur duality to quantum information theory*. PhD thesis, M.I.T., Cambridge, MA, 2005, arXiv:quant-ph/0512255.

[42] A. W. Harrow. Entanglement spread and clean resource inequalities. In *Proc. 16th Intl. Cong. Math. Phys.*, pages 536–540, 2009, arXiv:0909.1557.

[43] A. W. Harrow, P. Hayden, and D. W. Leung. Superdense coding of quantum states. *Phys. Rev. Lett.*, 92:187901, 2004, arXiv:quant-ph/0307221.

[44] A. W. Harrow and H.-K. Lo. A tight lower bound on the classical communication cost of entanglement dilution. *IEEE Trans. Inf. Theory*, 50(2):319–327, 2004, arXiv:quant-ph/0204096.

[45] M. Hayashi. Exponents of quantum fixed-length pure state source coding. *Phys. Rev. A*, 66:032321, 2002, arXiv:quant-ph/0202002.

[46] M. Hayashi. Optimal visible compression rate for mixed states is determined by entanglement of purificatio. *Phys. Rev. A*, 73:060301(R), 2006, arXiv:quant-ph/0511267.

[47] M. Hayashi. Universal approximation of multi-copy states and universal quantum lossless data compression. *Comm. Math. Phys.*, 293(1):171–183, 2010, arXiv:0806.1091.

[48] M. Hayashi and K. Matsumoto. Variable length universal entanglement concentration by local operations and its application to teleportation and dense coding, 2001, arXiv:quant-ph/0109028.

[49] M. Hayashi and K. Matsumoto. Quantum universal variable-length source coding. *Phys. Rev. A*, 66(2):022311, 2002, arXiv:quant-ph/0202001.

[50] M. Hayashi and K. Matsumoto. Simple construction of quantum universal variable-length source coding. *Quantum Inf. Comput.*, 2:519–529, 2002, arXiv:quant-ph/0209124.

[51] M. Hayashi and K. Matsumoto. Universal distortion-free entanglement concentration. *Phys. Rev. A*, 75:062338, 2007, arXiv:quant-ph/0209030.

[52] P. Hayden and A. Winter. On the communication cost of entanglement transformations. *Phys. Rev. A*, 67:012306, 2003, arXiv:quant-ph/0204092.

[53] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(1):13–30, March 1963.

[54] A. S. Holevo. On entanglement assisted classical capacity. *J. Math. Phys.*, 43(9):4326–4333, 2002, arXiv:quant-ph/0106075.

[55] A. S. Holevo. Remarks on the classical capacity of quantum channel, 2002, arXiv:quant-ph/0212025.

[56] A. S. Holevo. Complementary channels and the additivity problem. *Theory of Probability & Its Applications*, 51(1):92–100, 2007. quant-ph/0509101.

[57] M. Horodecki, J. Oppenheim, and A. Winter. Quantum information can be negative. *Nature*, 436:673—676, 2005, arXiv:quant-ph/0505062.

[58] R. Jozsa and S. Presnell. Universal quantum information compression and degrees of prior knowledge. *Proc. Roy. Soc. London Ser. A*, 459:3061–3077, October 2003, arXiv:quant-ph/0210196.

[59] M. Keyl. Quantum state estimation and large deviations. *Rev. Mod. Phys.*, 18(1):19–60, 2006, arXiv:quant-ph/0412053.

[60] A. Y. Kitaev, A. H. Shen, and M. N. Vyalyi. *Classical and Quantum Computation*, volume 47 of *Graduate Studies in Mathematics*. AMS, 2002.

[61] R. Koenig and S. Wehner. A strong converse for classical channel coding using entangled inputs. *Phys. Rev. Lett.*, 103:070504, 2009, arXiv:0903.2838.

[62] D. Kretschmann, D. Schlingemann, and R. F. Werner. The information-disturbance tradeoff and the continuity of Stinespring's representation. *IEEE Trans. Inf. Theory*, 54(4):1708–1717, April 2006, arXiv:quant-ph/0605009.

[63] D. Kretschmann and R. F. Werner. *Tema Con Variazioni:* quantum channel capacity. *New J. Phys.*, 6:26, 2004, arXiv:quant-ph/0311037.

[64] H.-K. Lo and S. Popescu. The classical communication cost of entanglement manipulation: Is entanglement an inter-convertible resource? *Phys. Rev. Lett.*, 83:1459–1462, 1999, arXiv:quant-ph/9902045.

[65] S. Massar and S. Popescu. Amount of information obtained by a quantum measurement. *Phys. Rev. A*, 61:062303, 2000, arXiv:quant-ph/9907066.

[66] S. Massar and A. Winter. Compression of quantum measurement operations. *Phys. Rev. A*, 64(012311), 2003, arXiv:quant-ph/0012128.

[67] M. A. Nielsen. Conditions for a class of entanglement transformations. *Phys. Rev. Lett.*, 83:436–439, 1999, arXiv:quant-ph/9811053.

[68] T. Ogawa and H. Nagaoka. Strong converse to the quantum channel coding theorem. *IEEE Trans. Inf. Theory*, 45(7):2486–2489, 1999, arXiv:quant-ph/9808063.

[69] V. Paulsen. *Completely Bounded Maps and Operator Algebras*. Cambridge University Press, 2003.

[70] M. Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden-Day, San Francisco, 1964.

[71] C. E. Shannon. A mathematical theory of communication. *Bell System Tech. Jnl.*, 27:379–423, 623–656, 1948.

[72] G. Smith and J. A. Smolin. Extensive nonadditivity of privacy. *Phys. Rev. Lett.*, 103:120503, 2009, arXiv:0904.4050.

[73] Y. Steinberg and S. Verdú. Simulation of random processes and rate-distortion theory. *IEEE Trans. Inf. Theory*, 42(1):63–86, Jan 1996.

[74] M. Takeoka, S. Guha, and M. M. Wilde. The squashed entanglement of a quantum channel, 2013, arXiv:1310.0129.

[75] B. M. Terhal, M. Horodecki, D. W. Leung, and D. P. DiVincenzo. The entanglement of purification. *J. Math. Phys.*, 43(9):4286–4298, 2002, arXiv:quant-ph/0202044.

[76] J. A. Tropp. User-friendly tail bounds for sums of random matrices, 2010, arXiv:1004.4389.

[77] A. Uhlmann. The 'transition probability' in the state space of a $*$-algebra. *Rep. Math. Phys.*, 9:273–279, 1976.

[78] W. van Dam and P. Hayden. Universal entanglement transformations without communication. *Phys. Rev. A*, 67(6):060302(R), 2003, arXiv:quant-ph/0201041.

[79] R. F. Werner and A. S. Holevo. Counterexample to an additivity conjecture for output purity of quantum channels. *Journal of Mathematical Physics*, 43(9):4353–4357, 2002, arXiv:quant-ph/0203003.

[80] M. Wilde, N. Datta, M.-H. Hsieh, and A. Winter. Quantum rate-distortion coding with auxiliary resources. *IEEE Trans. Inf. Theory*, 59(10):6755–6773, Oct 2013, arXiv:1212.5316.

[81] M. M. Wilde, P. Hayden, F. Buscemi, and M.-H. Hsieh. The information-theoretic costs of simulating quantum measurements. *Journal of Physics A: Mathematical and Theoretical*, 45(45):453001, 2012, arXiv:1206.4121.

[82] M. M. Wilde, A. Winter, and D. Yang. Strong converse for the classical capacity of entanglement-breaking and hadamard channels. 2013, arXiv:1306.1586.

[83] A. Winter. Coding theorem and strong converse for quantum channels. *IEEE Trans. Inf. Theory*, 45(7):2481–2485, 1999.

[84] A. Winter. Compression of sources of probability distributions and density operators, 2002, arXiv:quant-ph/0208131.

[85] A. Winter. "Extrinsic" and "intrinsic" data in quantum measurements: asymptotic convex decomposition of positive operator valued measures. *Comm. Math. Phys.*, 244(1):157–185, 2004, arXiv:quant-ph/0109050.

[86] A. Winter. Secret, public and quantum correlation cost of triples of random variables. In *2005 IEEE International Symposium on Information Theory*, pages 2270—2274, 2005.

[87] A. Winter. Identification via quantum channels in the presence of prior correlation and feedback. In *General Theory of Information Transfer and Combinatorics*, volume 4123 of *Lecture Notes in Computer Science*, pages 486–504. Springer Berlin Heidelberg, 2006, arXiv:quant-ph/0403203.

[88] W. Wootters and W. Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982.

[89] A. Wyner. The common information of two dependent random variables. *IEEE Trans. Inf. Theory*, 21(2):163–179, 1975.

[90] J. Yard and I. Devetak. Optimal quantum source coding with quantum information at the encoder and decoder. *IEEE Trans. Inf. Theory*, 55(11):5339–5351, Nov 2009, arXiv:0706.2907.

APPENDIX

In this appendix, we prove Lemma 13 and Lemma 6.

First, we prove Lemma 13, restated below for convenience. We follow the proof of Section 6.4.3 of [41], but simplify and streamline the arguments at the cost of proving a less general claim.

Lemma 13. *Let* $d = \max(d_A, d_B, d_E)$. *For any state* $|\varphi\rangle^{R^n A^n}$

with $|\Psi\rangle = (I \otimes U_{\mathcal{N}})^{\otimes n}|\varphi\rangle$,

$$\left\| |\Psi\rangle - \sum_{(\lambda_A,\lambda_B,\lambda_E)\in T^n_{\mathcal{N},\delta}} I \otimes ((\Pi_{\lambda_B} \otimes \Pi_{\lambda_E}) U_{\mathcal{N}}^{\otimes n} \Pi_{\lambda_A})|\varphi\rangle \right\|_1 \leq n^{O(d^2)} \exp\left(-n \frac{\delta^2}{8\log^2(d)}\right). \quad (109)$$

Proof: By the triangle inequality, the LHS of Eq. (109) is

$$\leq \sum_{(\lambda_A,\lambda_B,\lambda_E)\notin T^n_{\mathcal{N},\delta}} \|(I \otimes (\Pi_{\lambda_B} \otimes \Pi_{\lambda_E}) U_{\mathcal{N}}^{\otimes n} \Pi_{\lambda_A})|\varphi\rangle\|_1.$$

We now consider a particular triple $(\lambda_A, \lambda_B, \lambda_E) \notin T^n_{\mathcal{N},\delta}$. Let

$$\epsilon = \|(I \otimes (\Pi_{\lambda_B} \otimes \Pi_{\lambda_E}) U_{\mathcal{N}}^{\otimes n} \Pi_{\lambda_A})|\varphi\rangle\|_1 \quad (110)$$

$$= \mathrm{Tr}(\Pi_{\lambda_B} \otimes \Pi_{\lambda_E}) U_{\mathcal{N}}^{\otimes n} \Pi_{\lambda_A} \varphi^A \Pi_{\lambda_A} (U_{\mathcal{N}}^\dagger)^{\otimes n} \quad (111)$$

In this last step, we observe that all of the terms commute with collective permutations except for φ^A. Thus, Eq. (111) is unchanged if we replace φ^A with its symmetrized version, $\tilde{\varphi}^A := \frac{1}{n!}\sum_{\pi\in\mathcal{S}_n} \pi\varphi^A\pi^{-1}$. Next, observe that

$$\Pi_{\lambda_A}\tilde{\varphi}^A\Pi_\lambda = |\lambda_A\rangle\langle\lambda_A| \otimes \sigma \otimes \frac{I_{\mathcal{P}_{\lambda_A}}}{\dim \mathcal{P}_{\lambda_A}},$$

where σ is some (subnormalized) density matrix on $\mathcal{Q}^{d_A}_{\lambda_A}$. This implies that

$$\Pi_{\lambda_A}\tilde{\varphi}^A\Pi_\lambda \leq |\lambda_A\rangle\langle\lambda_A| \otimes I_{\mathcal{Q}^{d_A}_{\lambda_A}} \otimes \frac{I_{\mathcal{P}_{\lambda_A}}}{\dim \mathcal{P}_{\lambda_A}} = \frac{\Pi_{\lambda_A}}{\dim \mathcal{P}_{\lambda_A}} \quad (112)$$

Next, define the single-system density matrix $\rho = \sum_{i=1}^{d_A} \bar{\lambda}_{A,i}|i\rangle\langle i|$. By Eq. (76), we have

$$\mathrm{Tr}\,\Pi_{\lambda_A}\rho^{\otimes n} \geq (n+d)^{-d(d+1)/2}.$$

Thus, if we twirl $\rho^{\otimes n}$, we find that

$$\frac{\Pi_{\lambda_A}}{\dim \mathcal{P}_{\lambda_A}} \leq \mathbb{E}_{U\in\mathcal{U}_{d_A}}[(U\rho U^\dagger)^{\otimes n}] \cdot (n+d)^{d(d+1)/2} \dim \mathcal{Q}^{d_A}_{\lambda_A} \leq \mathbb{E}_{U\in\mathcal{U}_{d_A}}[(U\rho U^\dagger)^{\otimes n}](n+d)^{d^2}, \quad (113)$$

where in the second step we have used Eq. (71). Combining this equation with Eq. (112), we obtain the operator inequality

$$\Pi_{\lambda_A}\tilde{\varphi}^A\Pi_\lambda \leq \mathbb{E}_{U\in\mathcal{U}_{d_A}}[(U\rho U^\dagger)^{\otimes n}](n+d)^{d^2}.$$

Let r_B, r_E be the spectra respectively of the B and E parts of $U_{\mathcal{N}}U\rho U^\dagger U_{\mathcal{N}}^\dagger$. Then by the definition of $T^n_{\mathcal{N},\delta}$ we have that $\|r_B - \bar{\lambda}_B\|_1 + \|r_E - \bar{\lambda}_E\|_1 > \delta/\log(d)$. Thus, at least one of these distances must be $> \delta/2\log(d)$. By Pinsker's inequality it follows that either $D(\bar{\lambda}_B\|r_B) \geq \delta^2/8\log^2(d)$ or $D(\bar{\lambda}_E\|r_E) \geq \delta^2/8\log^2(d)$. This in turn means we can bound

$$\begin{aligned}\epsilon &\leq (n+d)^{d^2}\,\mathrm{Tr}(\Pi_{\lambda_B} \otimes \Pi_{\lambda_E})(U_{\mathcal{N}}U\rho U^\dagger U_{\mathcal{N}}^\dagger)^{\otimes n} \\ &\leq (n+d)^{d(3d-1)/2} \exp(-n\max(D(\bar{\lambda}_B\|r_B), D(\bar{\lambda}_E\|r_E))) \\ &\leq (n+d)^{d(3d-1)/2} \exp\left(-n\frac{\delta^2}{8\log^2(d)}\right) \quad (114)\end{aligned}$$

Finally, we sum over all $(\lambda_A, \lambda_B, \lambda_E) \notin T^n_{\mathcal{N},\delta}$ to upper-bound the LHS of Eq. (109) by $|\mathcal{I}_{d,n}|^3(n+d)^{d(3d-1)/2}\exp(-n\delta^2/8\log^2(d)) \leq (n+d)^{\frac{d(3d+5)}{2}}\exp(-n\delta^2/8\log^2(d))$. ∎

Lemma 6.

$$\max(0, \Delta_\epsilon(\rho)) = \min\{\Delta_0(\sigma) : \frac{1}{2}\|\rho-\sigma\|_1 \leq \epsilon, 0 \leq \sigma, \mathrm{Tr}\,\sigma = 1\} \quad (115)$$

Proof: If $\Delta_\epsilon(\rho) = \delta$ then by definition there exists σ satisfying $\Delta_0(\sigma) = \delta$, $0 \leq \sigma \leq \rho$, $\rho\sigma = \sigma\rho$ and $\mathrm{Tr}\,\sigma = 1-\epsilon$, which implies that $\|\rho-\sigma\|_1 = \mathrm{Tr}(\rho-\sigma) = \epsilon$. Let the nonzero eigenvalues of σ be $s_1 \geq \cdots \geq s_d > 0$. Then $ds_1 = 2^\delta$ and $\sum_{i=1}^d s_i = 1-\epsilon$. We can add up to $2^\delta - (1-\epsilon)$ weight to these eigenvalues while keeping them all $\leq s_1$. Thus, if

$$\epsilon \leq 2^\delta - (1-\epsilon), \quad (116)$$

then we can add ϵ weight to σ, thus obtaining a normalized state, without increasing its Δ_0. Call the resulting density matrix ω. Then $\Delta_0(\omega) = \delta$ and $\|\omega-\rho\|_1 \leq 2\epsilon$ by the triangle inequality. This is possible whenever Eq. (116) holds, or equivalently, whenever $\delta \geq 0$.

If $\delta < 0$, then we cannot create a normalized state without increasing the spread, since any normalized state has $\Delta_0 \geq 0$. Instead we will take ω to be the maximally mixed state on $\mathrm{supp}\sigma$. Note that $\sigma \leq \omega$, since $s_1 = 2^\delta/d < 1/d$. Thus $\|\omega-\sigma\|_1 = \mathrm{Tr}(\omega-\sigma) = \epsilon$ and we again have $\|\omega-\rho\|_1 \leq 2\epsilon$.

This establishes that the RHS of Eq. (115) is $\leq$ the LHS. To show the other direction, suppose that there exists a normalized ω satisfying $\frac{1}{2}\|\rho-\omega\|_1 \leq \epsilon$ and $\Delta_0(\omega) = \delta$. Then we can write $\rho-\omega = A-B$ where $A, B \geq 0$ and $\mathrm{Tr}\,A = \mathrm{Tr}\,B = \epsilon$. Define $\sigma = \rho - A = \omega - B$. Then $\mathrm{Tr}\,\sigma = 1-\epsilon$, $0 \leq \sigma \leq \rho$ and $\sigma \leq \omega$, implying $\Delta_0(\sigma) \leq \Delta_0(\omega) = \delta$. ∎