A Quantum Multiparty Packing Lemma and the Relay Channel

Dawei Ding, 1 Hrant Gharibyan, 1 Patrick Hayden, 1 and Michael Walter 2

1 Stanford Institute for Theoretical Physics, Stanford University, Stanford, CA 94305, USA
2 QuSoft, Korteweg-de Vries Institute for Mathematics, Institute for Theoretical Physics, and Institute for Logic, Language and Computation, University of Amsterdam, 1098 XG Amsterdam, The Netherlands

Optimally encoding classical information in a quantum system is one of the oldest and most fundamental challenges of quantum information theory. Holevo’s bound places a hard upper limit on such encodings, while the Holevo-Schumacher-Westmoreland (HSW) theorem addresses the question of how many classical messages can be “packed” into a given quantum system. In this article, we use Sen’s recent quantum joint typicality results to prove a one-shot multiparty quantum packing lemma generalizing the HSW theorem. The lemma is designed to be easily applicable in many network communication scenarios. As an illustration, we use it to straightforwardly obtain quantum generalizations of well-known classical coding schemes for the relay channel: multihop, coherent multihop, decode-forward, and partial decode-forward. We provide both finite blocklength and asymptotic results, the latter matching existing formulas. Given the key role of the classical packing lemma in network information theory, our packing lemma should help open the field to direct quantum generalization.

CONTENTS

I. Introduction  1
II. Preliminaries  4
III. Quantum Multiparty Packing Lemma  5
IV. Application to the Classical-Quantum Relay Channel  10
A. Multihop Scheme  12
B. Coherent Multihop Scheme  16
C. Decode Forward Scheme  17
D. Partial Decode Forward Scheme  19
V. Proof of the Quantum Multiparty Packing Lemma  24
VI. Conclusions  30
Acknowledgments  31
References  31
A. Proof of Cutset Bound  33

I. INTRODUCTION

The packing lemma is one of the central tools used in the construction and analysis of information transmission protocols [1]. It quantifies the asymptotic rate at which messages can be “packed”
reversibly into a medium, in the sense that the probability of a decoding errors vanishes in the limit of large blocklength. For concreteness, consider the following general version of the packing lemma.\footnote{See, e.g., [1]. Our formulation is slightly paraphrased and uses a notation that is more suitable for the following.}

**Lemma 1** (Classical Packing Lemma). Let \((U, X, Y)\) be a triple of random variables with joint distribution \(p_{UXY}\). For each \(n\), let \((\tilde{U}^n, \tilde{X}^n)\) be a pair of arbitrarily distributed random sequences and \(\{X^n(m)\}\) a family of at most \(2^nR\) random sequences such that each \(X^n(m)\) is conditionally independent of \(Y^n\) given \(\tilde{U}^n\) (but arbitrarily dependent on the other \(X^n(m')\) sequences). Further assume that each \(X^n(m)\) is distributed as \(\otimes_{i=1}^{n} p_{X|U=\tilde{U}_i}\) given \(\tilde{U}^n\). Then, there exists \(\delta(\varepsilon)\) that tends to zero as \(\varepsilon \to 0\) such that

\[
\lim_{n \to \infty} \Pr((\tilde{U}^n, \tilde{X}^n(m), \tilde{Y}^n) \in T^{(n)}_{\varepsilon} \text{ for some } m) = 0
\]

if \(R < I(X; Y|U) - \delta(\varepsilon)\), where \(T^{(n)}_{\varepsilon}\) is the set of \(\varepsilon\)-typical strings of length \(n\) with respect to \(p_{UXY}\).

The packing lemma provides a unified approach to many, if not most, of the achievability results in Shannon theory. Despite its broad utility, it is a simple consequence of the union bound and the standard joint typicality lemma with the three variables \(U, X, Y\). The usual channel coding theorem directly follows from taking \(U = \emptyset\) and when \(Y^n \sim p^n_Y\).

For the case when \(U = \emptyset\) and when \(Y^n \sim p^n_Y\), the quantum generalization of the packing lemma is known: the Holevo-Schumacher-Westmoreland (HSW) theorem [2, 3]. This can be proven using a conditional typicality lemma for a classical-quantum state with one classical and one quantum system. However, until recently no such typicality lemma was known for two classical systems and one quantum system, and so a quantum version of Lemma 1 was lacking. Furthermore, while in classical Shannon theory Lemma 1 can be used repeatedly in settings where the message is encoded into multiple random variables, this approach fails in the quantum case due to measurement disturbance, specifically the influence of one decoding on subsequent ones. Hence, while it is sufficient to solve the full multiparty packing problem in the classical case with just two senders and one receiver, a general multiparty packing lemma with \(k \in \mathbb{N}\) senders is required in the quantum case. The bottleneck is again the lack of a general quantum joint typicality lemma with more than two parties. However, we can obtain partial results in the quantum case for some network settings, as we will describe below.

In this paper we use the quantum joint typicality lemma established recently by Sen [4] to prove a quantum one-shot multiparty packing lemma for \(k\) senders. We then demonstrate the wide applicability of the lemma by using it to straightforwardly generalize classical protocols in a specific network communication setting to the quantum case. The lemma allows us to construct and prove the correctness of these simple generalizations and, we believe, should help to open the field of classical network information theory to direct quantum generalization. One feature of the lemma is that it leads naturally to demonstrations of the achievability of rate regions without having to resort to time-sharing, a desirable property known as simultaneous decoding. In network settings, this is often necessary because different receivers could have different effective rate regions and therefore require incompatible time-sharing strategies. Indeed, this is a frequent source of incomplete or incorrect results even in classical information theory [5]. A general construction leading to simultaneous decoding in the quantum setting has therefore been sought for many years [5–11]. Sen’s quantum joint typicality lemma achieves this goal, as does our packing lemma, which can be viewed as a user-friendly presentation of Sen’s lemma.

Recall that network information theory is the study of communication in the setting of multiple parties, a generalization of the conventional single-sender single-receiver two-party scenario, commonly known as point-to-point communication. Common network scenarios include having multiple
senders encoding different messages, as in the case of the multiple access channel [12], multiple receivers decoding the same message, as in the broadcast channel setting [13], or a combination of both, as for the interference channel [14]. However, the above examples are all instances of what is called single hop communication, where the message directly travels from a sender to a receiver. In multihop communication, there are one, or even multiple, intermediate nodes where the message is decoded or partially decoded before being transmitted to the final receiver. Examples of such settings include the relay channel [15], which we will focus on in this paper, and more generally, graphical multi-cast networks [16, 17].

Research in quantum joint typicality has generally been driven by the need to establish quantum generalizations of results in classical network information theory. Examples include the quantum multiple access channel [8, 18], the quantum broadcast channel [19, 20], and the quantum interference channel [9]. Indeed, some partial results on joint typicality had been established or conjectured in order to prove achievability bounds for various network information processing tasks [7, 21]. Subsequent work made some headway on the abstract problem of joint typicality for quantum states, but not enough to affect coding theorems [22, 23] prior to Sen’s breakthrough [4].

The quantum relay channel was studied previously in [24], where the authors constructed a partial decode-forward protocol. Here we develop finite blocklength results for the relay channel in addition to reproducing the earlier conclusions and avoiding a resolvable issue with error accumulation from successive measurements in their partial decode-forward bound. (We construct a joint decoder which obtains all the messages from the multiple rounds of communication at once.) Naturally, our analysis makes extensive use of the quantum multiparty packing lemma. Indeed, once the coding strategy is specified, a direct application of the packing lemma in the asymptotic limit gives a list of inequalities which describe the rate region, which we then simplify using entropy inequalities to the usual rate region of the partial decode-forward lower bound. There has also been related work in [25], which considered concatenated channels, a special case of the more general relay channel model. As noted in [24], work on quantum relay channels may have applications to designing quantum repeaters [26]. Sen has also used his joint typicality lemma to prove achievability results for the quantum multiple access, broadcast, and interference channels [4], but here we give a general packing lemma which can be conveniently used as a black box for quantum network information applications.

Our paper is structured as follows. In Section II, we establish our notation and discuss some preliminaries. In Section III, we describe the setting and state the quantum multiparty packing lemma. The statement will very much resemble a one-shot, multiparty generalization of Lemma 1 but, to reiterate, while the multiparty generalization is trivial in the classical case, it requires the power of a full joint typicality lemma in the quantum case. In Section IV we describe the setting of the classical-quantum (c-q) relay channel and systematically describe the achievability bounds corresponding to known coding schemes in the classical setting: multihop, coherent multihop, decode-forward, and partial decode-forward [27]. It is worthwhile to note that while the first three bounds only require the packing bound with two senders, the last bound is proved by applying multiparty packing for an arbitrary number of senders. In addition to the one-shot bounds, we show that the asymptotic bounds are obtained by taking the limit of large blocklength, thereby obtaining quantum generalizations of known capacity lower bounds for the classical case. In Section V we prove the quantum multiparty packing lemma via Sen’s quantum joint typicality lemma [4]. For convenience, we restate a special case of the Sen’s joint typicality lemma and suppress some of the details. In Section VI we give a conclusion, including an evaluation of the method proposed in this paper as well as possible directions for future work.
II. PRELIMINARIES

We first establish some notation and recall some basic results.

**Classical and quantum systems:** A classical system $X$ is identified with an alphabet $\mathcal{X}$ and a Hilbert space of dimension $|\mathcal{X}|$, while a quantum system $B$ is given by a Hilbert space of dimension $d_B$. Classical states are modeled by diagonal density operators such as $\rho_X = \sum_{x \in \mathcal{X}} p_X(x) |x\rangle \langle x|_X$, where $p_X$ is a probability distribution, quantum states are described by density operator $\rho_A$ etc, and classical-quantum states are described by density operators of the form

$$\rho_{XB} = \sum_{x \in \mathcal{X}} p_X(x) |x\rangle \langle x|_X \otimes \rho_B^{(x)}.$$  \hfill (1)

**Probability bound:** Denote by $E_1$, $E_2$ two events. We will use the following inequality repeatedly in the paper:

$$\Pr(E_1) = \Pr(E_1|E_2) \Pr(E_2) + \Pr(E_1|\overline{E_2}) \Pr(\overline{E_2}) \leq \Pr(E_2) + \Pr(E_1|\overline{E_2}),$$ \hfill (2)

where we use $\overline{E_2}$ to denote the complement of $E_2$ and used the fact that $\Pr(E_2), \Pr(E_1|\overline{E_2}) \leq 1$.

**Hypothesis-testing relative entropy:** The hypothesis-testing relative entropy is defined as

$$D_H^\varepsilon(\rho\parallel \sigma) = \max_{0 \leq \Pi \leq I} \frac{1}{\text{tr}(\Pi \rho)} - \log \text{tr}(\Pi \sigma).$$

For $n$ copies of states $\rho$ and $\sigma$, [28] establishes the following inequalities:

$$D(\rho\parallel \sigma) - \frac{F_1(\varepsilon)}{\sqrt{n}} \leq \frac{1}{n} D_H^\varepsilon(\rho^{\otimes n}\parallel \sigma^{\otimes n}) \leq D(\rho\parallel \sigma) + \frac{F_2(\varepsilon)}{\sqrt{n}},$$ \hfill (3)

where $F_1(\varepsilon), F_2(\varepsilon) \geq 0$ are given by $F_1(\varepsilon) \equiv 4\sqrt{2} \log \frac{1}{\varepsilon} \log \eta$ and $F_2(\varepsilon) \equiv 4\sqrt{2} \log \frac{1}{1-\varepsilon} \log \eta$, with $\eta \equiv 1+\text{tr} \rho^{3/2} \sigma^{-1/2} + \text{tr} \rho^{1/2} \sigma^{1/2}$. In the limit of large $n$, we obtain the quantum Stein’s lemma [29, 30]:

$$\lim_{n \to \infty} \frac{1}{n} D_H^\varepsilon(\rho^{\otimes n}\parallel \sigma^{\otimes n}) = D(\rho\parallel \sigma).$$ \hfill (4)

**Conditional density operators:** Let a classical system $X$ consist of subsystems $X_v$, for $v$ in some index set $V$, with alphabet $\mathcal{X} = \bigwedge_{v \in V} \mathcal{X}_v$. Consider a classical-quantum state $\rho_{XB}$ as in Eq. (1) and a subset $S \subseteq V$. We can write

$$\rho_{XB} = \sum_{x_\mathcal{X}} p_{X_\mathcal{X}}(x_\mathcal{X}) |x_\mathcal{X}\rangle \langle x_\mathcal{X}|_{X_\mathcal{X}} \otimes \rho_{XSB}^{(x_\mathcal{X})},$$ \hfill (5)

where

$$\rho_{XSB}^{(x_\mathcal{X})} \equiv \sum_{x_S} p_{XS|x_\mathcal{X}}(x_S|x_\mathcal{X}) |x_S\rangle \langle x_S|_{X_S} \otimes \rho_{BS}^{(x_S,x_\mathcal{X})}.$$

We can interpret $\rho_{XSB}^{(x_\mathcal{X})}$ as a “conditional” density operator. We further define $\rho_{XB}^{(\{X_S,B\})}$ by replacing the conditional density operator in Eq. (5) by the tensor product of its marginals:

$$\rho_{XB}^{(\{X_S,B\})} = \sum_{x_\mathcal{X}} p_{X_\mathcal{X}}(x_\mathcal{X}) |x_\mathcal{X}\rangle \langle x_\mathcal{X}|_{X_\mathcal{X}} \otimes \rho_{XS}^{(x_\mathcal{X})} \otimes \rho_{BS}^{(x_\mathcal{X})}.$$
A multiparty packing lemma is concerned with packing classical messages via an encoding that those messages, in turn, may be generated in a correlated fashion. Suppose for the purpose of this formulation lets us obtain the conditional mutual information as an asymptotic limit of the hypothesis testing relative entropy; by Eq. (4),

\[
\lim_{n \to \infty} \frac{1}{n} D_H^e \left( \rho_{XB}^{\otimes n} \parallel \left( \nu_{S} \otimes \rho_{B}^{\otimes n} \right) \right) = D(\rho_{XB} \parallel \nu_{XB}) - \sum_{x} p_{X}(x) D_{XB}(\nu_{X} \parallel \rho_{X})
\]

\[
= \sum_{x} p_{X}(x) I(X; B)_{\rho_{X}} = I(X; B|X)_{\rho_{X}}.
\]

This formulation lets us obtain the conditional mutual information as an asymptotic limit of the hypothesis testing relative entropy; by Eq. (4),

\[
\lim_{n \to \infty} \frac{1}{n} D_{H}^{e} \left( \rho_{XB}^{\otimes n} \parallel \left( \nu_{S} \otimes \rho_{B}^{\otimes n} \right) \right) = D(\rho_{XB} \parallel \nu_{XB}) - \sum_{x} p_{X}(x) D_{XB}(\nu_{X} \parallel \rho_{X})
\]

\[
= \sum_{x} p_{X}(x) I(X; B)_{\rho_{X}} = I(X; B|X)_{\rho_{X}}.
\]

III. QUANTUM MULTIPARTY PACKING LEMMA

In this section, we formulate a general multiparty packing lemma for quantum Shannon theory that can be conveniently used as a black box for random coding constructions. The goal is to “pack” as many classical messages as possible into our quantum system while retaining distinguishability. A multiparty packing lemma is concerned with packing classical messages via an encoding that involves multiple classical systems. As mentioned in the introduction, this is necessary in quantum information theory due to measurement disturbance. That is, while in classical information theory one can do consecutive decoding operations with impunity, in quantum information theory a decoding operation can change the system and thereby affect a subsequent operation. For example, while classically it is possible to check whether the output of a channel is typical for a tuple of input random variables simply by verifying typicality pair by pair, quantumly this method can be problematic. Hence, we would like to combine a set of decoding operations into one simultaneous decoding. We obtain a construction of this flavor in Lemma 2. Its asymptotic version, Lemma 3, states that the decoding error vanishes provided that a set of inequalities on the rate of transmission is satisfied, as opposed to a single one as in Lemma 1. This is exactly what we expect from a simultaneous decoding operation.

In order to motivate the formal statements to come, it is helpful to have an example in mind. In network coding scenarios, it is often necessary to have multiple message sets, representing in the simplest cases transmissions to and from different users or in different rounds of communication. Those messages, in turn, may be generated in a correlated fashion. Suppose for the purpose of illustration that we have three message sets $M_1$, $M_2$, and $M_3$ and a family of density operators $\rho^{(x_1,x_2,x_3)}$. To generate a code, we could choose $x_1(m_1)$ for $m_1 \in M_1$ according to $P_{X_1}$, next generate $x_2(m_1,m_2)$ for each $m_1$ according to $p_{X_2|x_1=x_1(m_1)}$, and lastly draw $x_3(m_1,m_2,m_3)$ according to $P_{X_3|x_1=x_1(m_1),x_2=x_2(m_2|m_1)}$ for each pair $m_1$, $m_2$. This arrangement can be represented graphically by a structure that we call a multiplex Bayesian network (Fig. 1, explained below). This structure is key to the technical setup of our multiparty packing lemma.

Let the random variable $X$ be a Bayesian network with respect to a directed acyclic graph (DAG) $G = (V, E)$. The random variable $X$ is composed of random variables $X_v$ with alphabet $X_v$ for each $v \in V$. For $v \in V$, let

\[
\text{pa}(v) = \{ v' \in V \mid (v', v) \in E \}
\]

denote the set of parents of $v$, corresponding to the random variables that $X_v$ is conditioned on. Below, we will use the Bayesian network to generate codewords $x(m)$ with components $x_v(m)$ for
We can visualize a multiplex Bayesian network by adjoining to the DAG \( G \) that is, where \( x \) vertices represent components of the codewords and the graph visualization of the example with three random variables, see Fig. 1. Therefore define the following algorithm: perform the for loops below, but do not impact the joint distribution of the codewords. We can for any two disjoint subsets \( J \), induces a lexicographical ordering on their Cartesian products, which we denote by \( X_{J'} \) for any \( J' \subseteq J \). We define \( M_{\emptyset} = \{\emptyset\} \) as a singleton set so that we can identify \( M_J \) for any two disjoint subsets \( J' \). These total orderings determine the order in which we perform the for loops below, but do not impact the joint distribution of the codewords. We can therefore define the following algorithm:

\[
\text{ind}(v') \subseteq \text{ind}(v) \text{ for every } v' \in \text{pa}(v).
\]

In the example, this captures the fact that the random variable \( x_3(m) \) is defined conditional on the value of \( x_2(m_1, m_2) \) and therefore must necessarily depend on \( m_1 \) and \( m_2 \); similarly for \( x_2 \) and \( m_2, m_1 \).

We will call the tuple \( B = (G, X, M, \text{ind}) \), where \( M \equiv X_{\cup J} \), a multiplex Bayesian network. We can visualize a multiplex Bayesian network by adjoining to the DAG \( G \) additional vertices \( M_j \), one for each \( j \in J \), and edges that connect each \( M_j \) to those \( X_v \) such that \( j \in \text{ind}(v) \). For a visualization of the example with three random variables, see Fig. 1.

![Figure 1](https://example.com/figure1.png)

**Figure 1. An example of a multiplex Bayesian network with vertices \( X_1, X_2 \) and \( X_3 \) and message sets \( M_1 \), \( M_2 \) and \( M_3 \). This network can be used to generate a code, where we choose \( x_1(m_1) \) for \( m_1 \in M_1 \) according to \( p_{X_1} \), \( x_2(m_2|m_1) \) according to \( p_{X_2|X_1=x_1(m_1)} \) and \( x_3(m_3|m_1, m_2) \) according to \( p_{X_3|X_1=x_1(m_1), X_2=x_2(m_2|m_1)}. \)

Fix a multiplex Bayesian network \( B = (G, X, M, \text{ind}) \). We would like to produce a random codebook

\[
\{x_v(m)\}_{v \in V, m \in M},
\]

where \( x_v \) is a random variable with alphabet \( X_v \). We will generate a random codebook via an algorithm implemented with respect to the multiplex Bayesian network being considered. The vertices represent components of the codewords and the graph \( G \) will be the Bayesian network describing the dependencies between the components of the random codewords. Moreover, each component \( x_v(m) \) will only depend on those parts \( m_j \in M_j \) of the message for which \( j \in \text{ind}(v) \). That is, \( x_v(m) \) and \( x_v(m') \) will be equal as random variables provided \( m_j = m'_j \) for every \( j \in \text{ind}(v) \).

We now give the algorithm for generating the random codebook. Since \( G \) is a DAG, it has a topological ordering, that is, a total ordering on \( V \) such that for every \((v', v) \in E\), \( v' \) precedes \( v \) in the ordering. We also pick an arbitrary total ordering on \( J \) and on \( M_j \) for every \( j \in J \). This then induces a lexicographical ordering on their Cartesian products, which we denote by \( M_{J'} := X_{\cup J'} \). We define \( M_{\emptyset} = \{\emptyset\} \) as a singleton set so that we can identify \( M_J \times M_{J''} = M_{J' \cup J''} \) for any two disjoint subsets \( J', J'' \subseteq J \). These total orderings determine the order in which we perform the for loops below, but do not impact the joint distribution of the codewords. We can therefore define the following algorithm:
Algorithm 1: Codebook generation from multiplex Bayesian network

for $v \in V$ do 
    for $m_v \in M_{\text{ind}(v)}$ do 
        generate $x_v(m_v)$ according to $p_{X_v|X_{\text{pa}(v)}}(\cdot|x_{\text{pa}(v)}(m_{\text{pa}(v)}))$ 
    for $\bar{m}_v \in M_{\text{ind}(v)}$ do 
        $x_v(m_v, \bar{m}_v) = x_v(m_v)$ 
    end for 
end for

Here, $\text{ind}(v) \equiv J \setminus \text{ind}(v)$, $m_{\text{pa}(v)}$ is the restriction of $m_v$ to $M_{\text{ind}(\text{pa}(v))}$ (this makes sense by Eq. (7)), $X_{\text{pa}(v)} = (X_{v'})_{v' \in \text{pa}(v)}$ and similarly for $x_{\text{pa}(v)}(m_{\text{pa}(v)})$, and the pair $(m_v, \bar{m}_v)$ is interpreted as an element of $M$ with the appropriate components. The topological ordering on $V$ ensures that $x_{\text{pa}(v)}(m_{\text{pa}(v)})$ is generated before $x_v(m_v)$, so this algorithm can be run. We thus obtain a random codebook as in Eq. (8).

We make a few observations.

1. By construction, for all $m \in M$ and $\xi \in \mathcal{X}$,
   \[
   \Pr(x(m) = \xi) = p_X(\xi) \equiv \prod_{v \in V} p_{X_v|X_{\text{pa}(v)}}(\xi_v|\xi_{\text{pa}(v)}). \]
   That is, $x(m)$ is a Bayesian network with respect to $G$ equal in distribution to $X$.

2. By construction, given $v \in V$ and $m_v \in M_{\text{ind}(v)}$, all $x_v(m_v, \bar{m}_v)$ for $\bar{m}_v \in M_{\text{ind}(v)}$ are equal as random variables.

3. Generalizing observation 1, the joint distribution of all codewords can be split into factors in a simple manner. Specifically, given $\xi(m) \in \mathcal{X}$ for every $m \in M$, we have
   \[
   \Pr(x(m) = \xi(m)) \text{ for all } m \in M = \prod_{v \in V} \prod_{m_v \in M_{\text{ind}(v)}} p_{X_v|X_{\text{pa}(v)}}(\xi_v(m_v)|\xi_{\text{pa}(v)}(m_{\text{pa}(v)}))
   \]
   provided $\xi_v(m_v) = \xi_v(m'_v)$ for all $m, m'$ with $m_v = m'_v$. Otherwise, the joint probability is zero.

We will use Algorithm 1 on $B$ to obtain a codebook for which we would like to construct multiple different quantum decoders. More precisely, let $H$ be the induced subgraph of $G$ for some $V_H \subseteq V$ where for all $v \in V_H$, $\text{pa}(v) \subseteq V_H$. We call $H$ an ancestral subgraph. Then, we can naturally define $X_H$ to be the set of random variables corresponding to $V_H$, $J_H \equiv \bigcup_{v \in V_H} \text{ind}(v) \subseteq J$, $M_H \equiv \bigtimes_{j \in J_H} M_j$, and $C_H \equiv \{x_H(m_H)\}_{m_H \in M_H}$. We will then use a quantum encoding $\{\rho_B^{(x_H)}\}_{x_H \in \mathcal{X}_H}$ where $B$ is some quantum system. Furthermore, the receiver will also only need to decode a subset of the components of the message $D \subseteq J_H$ since they might in general have a guess for the other components $\overline{D} \equiv J_H \setminus D$. This is a very general construction for classical-quantum network communication settings, where $J$ and $X$ will respectively correspond to the messages and classical inputs to the classical-quantum channel on different rounds of communication. $H$ would then be the inputs on a particular round, and $\overline{D}$ would be the decoder’s message estimates from previous rounds.

We can now state our quantum multiparty packing lemma:

---

3 Note that by the definition of $M_H$ we only need $m_H$ to identify $x_H$ up to equality as random variables.
Lemma 2 (One-shot quantum multiparty packing lemma). Let $\mathcal{B} = (G, X, M, \text{ind})$ be a multiplex Bayesian network and run Algorithm 1 to obtain a random codebook $C = \{x(m)\}_{m \in M}$. Let $H \subseteq G$ be an ancestral subgraph, $\{\rho_B^{x_B}\}_{x_B \in X_H}$ a family of quantum states, $D \subseteq J_H$, and $\varepsilon \in (0,1)$. Then there exists a POVM\(^4\) $\{Q_B^{(m_D, m_{\overline{D}})}\}_{m_D \in M_D}$ for each $m_{\overline{D}} \in M_{\overline{D}}$ such that, for all $(m_D, m_{\overline{D}}) \in M_H$, $\mathbb{E}_C \left[ \text{tr} \left[ (I - Q_B^{(m_D, m_{\overline{D}})}) \rho_B^{x_B} \right] \right] \leq f(|V_H|, \varepsilon) + 4 \sum_{\emptyset \neq T \subseteq D} 2^{(\sum_{i \in T} R_i) - D_H(\rho_{X_B}^{(X_{\overline{B}})}, \rho_{X_B}^{(X_{\overline{B}})}))}. \quad (9)$

Here, $\mathbb{E}_C$ denotes the expectation over the random codebook $C_H = \{x_H(m_H)\}_{m_H \in M_H}$, $R_i \equiv \log |M_i|$, and $S_T \equiv \{v \in V_H \mid \text{ind}(v) \cap T \neq \emptyset \}$, and

$$\rho_{X_H B} \equiv \sum_{x_H \in X_H} p_{X_H}(x_H) |x_H\rangle \langle x_H|_{X_H} \otimes \rho_B^{x_H}. \quad$$

Furthermore, $f(k, \varepsilon)$ is a universal function (independent of our setup) that tends to zero as $\varepsilon \to 0$.

Remark. The bound in Eq. (9) can also be written as

$$\mathbb{E}_C \left[ \text{tr} \left[ (I - Q_B^{(m_D, m_{\overline{D}})}) \rho_B^{x_B} \right] \right] \leq f(|V_H|, \varepsilon) + 4 \sum_{m'_D \neq m_D} 2^{-D_H(\rho_{X_B}^{(X_{\overline{B}})}, \rho_{X_B}^{(X_{\overline{B}})}))}, \quad (10)$$

where $S \equiv \{v \in V_H \mid \exists j \in D \cap \text{ind}(v) \text{ such that } (m_D)_j \neq (m'_D)_j \}.$

In words, $S$ is the set of random codewords that depend on a part of the message that differs between $m_D$ and $m'_D$. This is similar to decoding error bounds obtained with conventional methods, such as the Hayashi-Nagaoka lemma [31]. We obtain Eq. (9) from Eq. (10) by parametrizing the different $m'_D$ with respect to the components that differ from $m_D$.

Remark. Note Eq. (9) assumes that the decoder’s guess of $m_{\overline{D}}$ is correct. That is, they choose the POVM $\{Q_B^{(m_D, m_{\overline{D}})}\}_{m_D \in D'}$ where $m_{\overline{D}}$ is exactly the $m_{\overline{D}}$ in the encoded state $\rho_B^{x_B(m_D, m_{\overline{D}})}$. If the decoder’s guess is incorrect, then this bound will not hold in general. In applications, $m_{\overline{D}}$ will typically correspond to message estimates of previous rounds, which we will assume to be correct by invoking a union bound. That is, we bound the total probability of error by summing the probabilities of error of a decoding assuming that all previous decodings were correct.

Using Lemma 2 and Eq. (6), we can naturally obtain the asymptotic version where we simply repeat the encoding-decoding procedure $n \in \mathbb{N}$ times and take the limit of large $n$. By the quantum Stein’s lemma Eq. (4), the error in Eq. (9) will vanish if the rates of encoding are bounded by conditional mutual information quantities. We present this as a self-contained statement.

Lemma 3 (Asymptotic quantum multiparty packing lemma). Let $\mathcal{B} = (G, X, M, \text{ind})$ be a multiplex Bayesian network. Run Algorithm 1 $n$ times to obtain a random codebook $C^n = \{x^n(m) \in X^n\}_{m \in M}$.\(^4\) These POVMs depend on the codebook $C_H$ and are hence involved in the averaging in Eq. (9). This will be important in the analyses below.
Let $H \subseteq G$ be an ancestral subgraph, $\{f^{(x_H)}_B\}_{x_H \in X_H}$ a family of quantum states, and $D \subseteq J_H$. Then there exists a POVM $\{Q^{(m_D,m_{\overline{S}})}_B\}_{m_D \in M_D}$ for each $m_{\overline{S}} \in M_{\overline{S}}$ such that, for all $(m_D, m_{\overline{S}}) \in M_H$,

$$\lim_{n \to \infty} \mathbb{E}_{C^m_H} \left[ \operatorname{tr} \left( (I - Q^{(m_D,m_{\overline{S}})}_B) \bigotimes_{i=1}^n \rho_{B_i}^{(x_i, H(m_D,m_{\overline{S}}))} \right) \right] = 0,$$

provided that

$$\sum_{t \in T} R_t < n I(X_{S_T}; B|X_{S_T})_\rho \quad \text{for all } \emptyset \neq T \subseteq D.$$

Above, $\mathbb{E}_{C^m_H}$ is the expectation over the random codebook $C^m_H = \{x^m_H(m_H)\}_{m_H \in M_H}$, $R_t = \log |M_t|$, $S_T = \{ v \in V_H \mid \text{ind}(v) \cap T \neq \emptyset \}$, and

$$\rho_{X_H B} = \sum_{x_H \in X_H} p_{X_H}(x_H) |x_H\rangle \langle x_H|_{X_H} \otimes \rho_B^{(x_H)}.$$  

**Example.** To clarify the definitions and illustrate the application of Lemma 3 we give a concrete example of a multiparty packing setting. Consider the multiplex Bayesian network given in Fig. 1. Then, choosing $H = G$ and $D = \{2,3\}$, we obtain a POVM $\{Q^{(m_2,m_3)}_B\}_{m_2 \in M_2, m_3 \in M_3}$ for each $m_1 \in M_1$. The mapping from $T \subseteq D$ to $S_T \subseteq V = \{1,2,3\}$ is given in Table I. Hence, we obtain vanishing error in the asymptotic limit if

$$R_2 < n I(X_2 X_3; B|X_1)_\rho,$$

$$R_3 < n I(X_3; B|X_1 X_2)_\rho,$$

$$R_2 + R_3 < n I(X_2 X_3; B|X_1)_\rho,$$

where

$$\rho_{X_1 X_2 X_3 B} = \sum_{x_1, x_2, x_3} p_{X_1 X_2 X_3}(x_1, x_2, x_3) |x_1, x_2, x_3\rangle \langle x_1, x_2, x_3|_{X_1 X_2 X_3} \otimes \rho_B^{(x_1, x_2, x_3)}.$$  

Note that the third inequality subsumes the first.

| $T$ | $S_T$ |
|-----|------|
| $\{2\}$ | $\{2,3\}$ |
| $\{3\}$ | $\{3\}$ |
| $\{2,3\}$ | $\{2,3\}$ |

Table I. $S_T \subseteq V$ for various $\emptyset \neq T \subseteq D$.

We expect that Lemma 3 can be used in a variety of settings to directly generalize results from classical network information theory, which often hinge on Lemma 1, to the quantum case.

In fact, it is not too difficult to see that an i.i.d. variant$^5$ of Lemma 1 can be derived from Lemma 3. More precisely, let $(U, X, Y) \sim p_{U|XY}$ be a triple of random variables as in the former. Consider a

$^5$ This is because we assume i.i.d. codewords in Lemma 3, which is sufficient for, e.g., relay, multiple access [4], and broadcast channels [32].
DAG $G$ consisting of two vertices, corresponding to random variables $U$ and $X$ with joint distribution $p_{UX}$, and an edge going from the former to the latter. We set $J = \{1\}$, ind($X$) = $\{1\}$, and $M_1 = M$ as the message set. A visualization of this simple multiplex Bayesian network $(G,(U,X),M,\text{ind})$ is given in Fig. 2.

By running Algorithm 1 $n$ times, we obtain codewords which we can identify as $\tilde{U}^n$ and $\tilde{X}^n(m)$. Conditioned on $\tilde{U}^n$, it is clear that for each $m \in M$, $\tilde{X}^n(m) \sim \bigotimes_{i=1}^n p_{X|U=\tilde{U}^i}$. Next, choose the subgraph to be all of $G$, set of quantum states the classical states

$$\rho_{\tilde{Y}_i}^{(u,x)} = \sum_{\tilde{y} \in Y} p_{Y|UX}(\tilde{y}|u,x) |\tilde{y}\rangle \langle \tilde{y}|_{\tilde{Y}_i}$$

and decoding subset $D = \{1\}$, corresponding to $M$. We see that if we consider the entire system consisting of $\tilde{U}^n$, $\tilde{X}^n(m)$ and $\bigotimes_{i=1}^n \rho_{\tilde{Y}_i}^{(\tilde{U},\tilde{X}^n(m'))}$ for $m' \neq m$, it is clear that $\tilde{X}^n(m)$ is conditionally independent of $\tilde{Y}^n$ due to the conditional independence of $X^n(m)$ and $X^n(m')$ given $\tilde{U}^n$. By Lemma 3, we obtain a POVM $\{Q_{\tilde{Y}_n}^{(m)}\}_{m \in M}$ such that, for all $m \in M$,

$$\lim_{n \to \infty} E_{C^n} \left[ \text{tr} \left( \left( I - Q_{\tilde{Y}_n}^{(m)} \right) \bigotimes_{i=1}^n \rho_{\tilde{Y}_i}^{(\tilde{u}_i x, m))} \right) \right] = 0$$

provided $R < I(X;Y|U)$, which is analogous to Lemma 1 if we “identify” the POVM measurement with the typicality test.

In Section V we will prove Lemma 2 using Sen’s quantum joint typicality lemma with $|V|$ classical systems and a single quantum system. We will then prove Lemma 3. In the proof of our packing lemma, we will actually prove a more general, albeit more abstract, statement.

### IV. APPLICATION TO THE CLASSICAL-QUANTUM RELAY CHANNEL

To illustrate the wide applicability of Lemma 2 and demonstrate its power, we will use it to prove achievability results for the classical-quantum relay channel. The first three results make use of the packing lemma in situations where the number of random variables involved in the decoding is at most two ($|V_H| \leq 2$). This situation can be dealt with using existing techniques [24]. The final partial decode-forward lower bound, however, applies the packing lemma with $|V_H|$ unbounded with increasing blocklength, thus requiring its full strength. These lower bounds are well-known for classical relay channels [1], and that our packing lemma allows us to straightforwardly generalize them to the quantum and even finite blocklength case.\(^6\)

---

\(^6\) Note that in this case the one-shot capacity reduces to the point-to-point scenario, as the relay lags behind the sender.
After the completion of the sender transmits to their protocol with received systems. On round and number of messages sender's the receiver's. channels because the relay's transmission also affects the system that the relay obtains and the setup is shown in Fig. 3. Note that this is much more general than the setting of two concatenated classical-quantum channel was first established in [24].

First we give some definitions. Recall that a classical-quantum relay channel [24, 25] is a generalization of mutual information. Note that the partial decode-forward asymptotic bound for lower bounds on the capacity, which match exactly those of the classical setting with the quantum quantum relay channel.

Figure 3. The relay channel $\mathcal{N}_{X_1X_2\rightarrow B_2B_3}$ is a family of quantum states which defines the classical-quantum relay channel.

lower bounds on the capacity, which match exactly those of the classical setting with the quantum generalization of mutual information. Note that the partial decode-forward asymptotic bound for the classical-quantum relay channel was first established in [24].

We now define what comprises a general code for the classical-quantum relay channel. Let $n \in \mathbb{N}$, $R \in \mathbb{R}_{\geq 0}$. A $(n, 2^{nR})$ code for classical-quantum relay channel $\mathcal{N}_{X_1X_2\rightarrow B_2B_3}$ for $n$ uses of the channel and number of messages $2^{nR}$ consists of

1. A message set $\mathcal{M}$ with cardinality $2^{nR}$.
2. An encoding $\mathcal{X}_n^m(m) \in \mathcal{X}_n$ for each $m \in \mathcal{M}$.
3. A relay encoding and decoding $\mathcal{R}_{(B_2)j-1(B_2')j-1\rightarrow X_2j}(B_2', B_2)$ for $j \in [n]$. Here, $(B_2)_j$ is isomorphic to $B_2$ and $(X_2)_j$ isomorphic to $X_2$ while $(B_2')_j$ is some arbitrary quantum system. The relay starts with some trivial (dimension 0) quantum system $(B_2)_0(B_2')_0$.
4. A receiver decoding POVM $\{Q^{(m)}_{B_2'}\}_{m \in \mathcal{M}}$.

On round $j$, the sender transmits $(x_1)_j(m)$ while the relay applies the $\mathcal{R}_{(B_2j-1)(B_2')j-1\rightarrow(X_2j)(B_2')_j}$ to their $(B_2j-1)(B_2')_j$ system and transmits the $(X_2j)$ state while keeping the $(B_2')_j$ system. After the completion of $n$ rounds, the receiver applies the decoding POVM $\{Q^{(m)}_{B_2'}\}_{m \in \mathcal{M}}$ on their received systems $\rho_{B_2'}(m)$ to obtain their estimate for the message. See Fig. 4 for a visualization of a protocol with $n = 3$ rounds. The average probability of error of a general protocol is given by

\[
p_e = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \text{tr} \left[ \left( I - Q^{(m)}_{B_2'} \right) \rho_{B_2'}(m) \right].
\]

\footnote{Here $\mathcal{R}_{(B_2j-1)(B_2')j-1\rightarrow(X_2j)(B_2')_j}$ has label $j$ that we will not write explicitly since systems $X_2$, $B_2$ and $B_2'$ are already labeled.}
Figure 4. A 3-round protocol for the classical-quantum relay channel. Here $R_{j-1}$ denotes the relay operation applied on round $j$, and $(B'_j)_j$ denotes the state left behind by the relay operation. The decoding operator $D$ is applied to all systems $(B_3)_j$ simultaneously.

In the protocols we give below, we use random codebooks. We can derandomize in the usual way to conform to the above definition of a code. Furthermore, in our protocols the relay only leaves behind a classical system when decoding. Since our relay channels are classical-quantum, it is not clear that this is suboptimal.

Given $R \in \mathbb{R}_{\geq 0}$, $n \in \mathbb{N}$, $\delta \in [0,1]$, we say that a triple $(R,n,\delta)$ is achievable for a relay channel if there exists a $(n,2^{nR'})$ code such that

$$R' \geq R \quad \text{and} \quad p_e \leq \delta.$$ 

The capacity of the classical-quantum relay channel $N_{X_1X_2\rightarrow B_2B_3}$ is then defined as

$$C(N) \equiv \lim_{\delta \rightarrow 0} \lim_{b \rightarrow \infty} \inf \{R : (R,n,\delta) \text{ is achievable for } N \}.$$ 

Now, before looking at specific coding schemes, we first give a general upper bound, a direct generalization of the cutset bound for the classical relay channel:

**Proposition 4** (Cutset Bound). Given a classical-quantum relay channel $N_{X_1X_2\rightarrow B_2B_3}$, its capacity is bounded from above by

$$C(N_{X_1X_2\rightarrow B_2B_3}) \leq \max_{p_{X_1X_2}} \min \{I(X_1X_2; B_3), I(X_1; B_2B_3|X_2)\}.$$ (11) 

**Proof.** See Appendix A. \hfill \Box

For some special relay channels, this along with some of the lower bounds proven below will be sufficient to determine the capacity.

**A. Multihop Scheme**

The multihop lower bound is obtained by a simple two-step process where the sender transmits the message to the relay and the relay then transmits it to the receiver. That is, the relay simply “relays” the message. The protocol we give below is exactly analogous to the classical case [1], right
We also define \( b \) when \( m \rightarrow M \equiv \{ 0 : 2^b \} \), where we restricted to the message components the codewords are dependent on via \( \rho \). We will show that we can achieve the triple \( \langle \frac{1}{2} R, b, \delta \rangle \) for some \( \delta \) a function of \( R, b, \varepsilon \). Let \( p_{X_1}, p_{X_2} \) be probability distributions over \( X_1, X_2 \), respectively. Throughout, we will use \( \rho_{X_1, X_2} = \sum_{x_1, x_2} p_{X_1}(x_1) p_{X_2}(x_2) |x_1 x_2\rangle \langle x_1 x_2| \otimes \rho_{B_1 B_2}^{(x_1 x_2)} \).

We also define \( \rho_{B_3}^{(x_2)} = \sum_{x_1} p_{X_1}(x_1) \rho_{B_3}^{(x_1 x_2)} \) to be the reduced state on \( B_3 \) induced by tracing out \( X_1 B_2 \) and fixing \( X_2 \). We will use random coding with a one-shot block Markov scheme.

**Code:** Throughout, \( j \in [b] \). Let \( G \) be a graph with \( 2b \) vertices corresponding to independent random variables \( (X_1)_j \sim p_{X_1}, (X_2)_j \sim p_{X_2} \). Since all the random variables are independent, there are no edges. Furthermore, let \( M_0, M_j \) be index sets, where \( |M_0| = 1 \) and \( |M_j| = 2^R \). That is, \( J = \{ 0 : b \} \). The \( M_j \) will be the sets from which the messages for each round will be taken. We use single element message set \( M_0 \) to make the effect of the first and the last blocks more explicit. Finally, the function \( \text{ind} \) maps \( (X_1)_j \) to \( \{ j \} \) and \( (X_2)_j \) to \( \{ j - 1 \} \). Then, letting \( X \equiv X_1^i X_2^j \) and \( M \equiv \{ j \}_{j=0}^b M_j \), \( B \equiv (G, X, M, \text{ind}) \) is a multiplex Bayesian network. See Fig. 5 for a visualization when \( b = 3 \). Now, run Algorithm 1 with \( B \) as the argument. This will return a random codebook

\[
C = \bigcup_{j=1}^b \{(x_1)_j(m_j), (x_2)_j m_{j-1})\}_{m_j \in M_j, m_{j-1} \in M_{j-1}},
\]

where we restricted to the message components the codewords are dependent on via \( \text{ind} \). For decoding we will apply Lemma 2 with this codebook and use the assortment of POVMs that are given for different ancestral subgraphs and other parameters.

**Encoding:** On the \( j \)-th transmission, the sender transmits a message \( m_j \in M_j \) via \( (x_1)_j(m_j) \in C \).

**Relay encoding:** Set \( \tilde{m}_0 \) to be the sole element of \( M_0 \). On the \( j \)-th transmission, the relay sends their estimate \( \tilde{m}_{j-1} \) via \( (x_2)_j(\tilde{m}_{j-1}) \in C \). Note that this is the relay’s estimate of the message \( m_{j-1} \) transmitted by the sender on the \( (j - 1) \)-th transmission.
Relay decoding: Consider the $j$-th transmission. We invoke Lemma 2 with the ancestral subgraph containing the two vertices $(X_1)_j$ and $(X_2)_j$, the set of quantum states $\{\rho^{(x_2)}_{B_2}\}_{x_2 \in X_2}$, decoding subset $\{j\} \subseteq \{j-1, j\}$, and small parameter $\varepsilon \in (0, 1)$. The relay picks the POVM corresponding to the message estimate for the previous round $\tilde{m}_{j-1}$, which is denoted by $\{Q^{(m'_j|\tilde{m}_{j-1})}_{B_2}\}_{m' \in M_j}$. He applies this on their received state to obtain a measurement result $\tilde{m}_j$. Note that this is the relay’s estimate for message $m_j$.

Decoding: On the $j$-th transmission, we again invoke Lemma 2 and the receiver will use the POVM corresponding to the ancestral subgraph containing just the vertex $(X_2)_j$, the set of quantum states $\{\rho^{(x_2)}_{B_2}\}_{x_2 \in X_2}$, decoding subset $\{j-1\} \subseteq \{j-1\}$, and small parameter $\varepsilon$. Note that we don’t have a message guess here since the decoding subset is not a proper subset. In this case where $p \equiv (m_0, \ldots, m_{b-1})$. Note that $m_b$ is never decoded by the receiver since it is the message sent in the last block and thus, we can ignore it without loss of generality. Let $\tilde{m} \equiv (\tilde{m}_0, \ldots, \tilde{m}_{b-1})$, $\tilde{m} \equiv (\tilde{m}_0, \ldots, \tilde{m}_{b-1})$ denote the aggregation of the message estimates of the relay and receiver, respectively. The probability of error averaged over the random codebook $C$ is given by

$$p_e(C) = \mathbb{E}_C[p(\tilde{m} \neq m)],$$

where $p$ here denotes the probability for a fixed codebook. Now, by Eq. (2),

$$p_e(C) \leq \mathbb{E}_C[p(\tilde{m} \neq m)] + \mathbb{E}_C[p(\tilde{m} \neq m|\tilde{m} = m)].$$

We consider the first term corresponding to the relay decoding. By the union bound,

$$\mathbb{E}_C[p(\tilde{m} \neq m)] \leq \mathbb{E}_C[p(\tilde{m}_0 \neq m_0)] + \sum_{j=1}^{b-1} \mathbb{E}_C[p(\tilde{m}_j \neq m_j|\tilde{m}_{j-1} = m_{j-1})].$$

By the definition of $\tilde{m}_0$, the first term is zero. Now, we can apply Eq. (9) to bound each summand in the second term as follows:8

$$\mathbb{E}_C[p(\tilde{m}_j \neq m_j|\tilde{m}_{j-1} = m_{j-1})] = \mathbb{E}_C\left[\text{tr}\left((I - Q^{(m_j|m_{j-1})}_{B_2})\rho^{(x_1)_j,(m_j)(x_2)_j}(m_{j-1})\rho^{(x_1)_j,(m_j)(x_2)_j}(m_{j-1})\right)\right]$$

$$= \mathbb{E}_{C_{(X_1)_j(x_2)_j}}\left[\text{tr}\left((I - Q^{(m_j|m_{j-1})}_{B_2})\rho^{(x_1)_j,(m_j)(x_2)_j}(m_{j-1})\right)\right]$$

$$\leq f(2, \varepsilon) + 4 \sum_{T = \{j\}} 2^{-R_{D_H}(\rho^{(x_1)_j(x_2)_j,b_2}||\rho^{(x_1)_j(x_2)_j,b_2})}$$

$$= f(2, \varepsilon) + 4 \times 2^{-R_{D_H}(\rho^{(x_1)_j(x_2)_j,b_2}||\rho^{(x_1)_j(x_2)_j,b_2})},$$

8 The careful reader would notice that the conditioning on $\tilde{m}_{j-1} = m_{j-1}$ is not necessary here since the probability of decoding $m_j$ correctly at the relay is independent of whether $m_{j-1}$ was decoded successfully. However, this will be necessary for the other schemes we give.
where $C_{(X_1)_j(X_2)_j}$ is the corresponding subset of the codebook $C$, and we used $S_{(j)} = \{(X_1)_j\}$. We dropped the index $j$ in the last equality since $(X_1)_j(X_2)_j \sim p_{X_1} \times p_{X_2}$. Hence, overall,

$$
E_C[p(\tilde{m} \neq m)] \leq b \left[ f_2(\varepsilon) + 4 \times 2^{R - D_H^c(\rho_{X_1 X_2 B_2})} \right].
$$

We now consider the second term in Eq. (12), corresponding to the receiver decoding. By the union bound,

$$
E_C[p(\tilde{m} \neq m|m = m)] \leq \sum_{j=1}^{b-1} E_C[p(\hat{m}_j \neq m_j|m = m)].
$$

Again by definition, the first term vanishes. Now, the receiver on the $(j+1)$-th transmission obtains the state $\rho_{B_3}^{((x_1)_{j+1}(m_{j+1})(x_2)_{j+1}(m_{j+1}))}$. Averaging over $(x_1)_{j+1}(m_{j+1})$, this becomes $\rho_{B_3}^{((x_2)_{j+1}(m_{j}))}$. Hence, the summands in second term are also bounded via Eq. (9):

$$
E_C[p(\hat{m}_j \neq m_j|m = m)] = E_C\left[ \left| \text{tr} \left[ (I - Q_{B_3}^{(m_j)}) \rho_{B_3}^{((x_1)_{j+1}(m_{j+1})(x_2)_{j+1}(m_{j}))} \right] \right| \right]
$$

$$
= E_C (X_1)_{j+1}(X_2)_{j+1} \left[ \left| \text{tr} \left[ (I - Q_{B_3}^{(m_j)}) \rho_{B_3}^{((x_1)(x_2)(m_j))} \right] \right| \right]
$$

$$
= E_C (X_2)_{j+1} \left[ \text{tr} \left[ (I - Q_{B_3}^{(m_j)}) \rho_{B_3}^{((x_2)(m_j))} \right] \right]
$$

$$
\leq f(1, \varepsilon) + 4 \sum_{T = (j)} 2^{R - D_H^c(\rho_{B_3}^{((X_2)(m_j))})},
$$

where we used $S_{(j)} = \{(X_2)_{j+1}\}$ and again dropped indices in the last inequality. Hence, overall

$$
E_C[p(\tilde{m} \neq m|m = m)] \leq b \left[ f(1, \varepsilon) + 4 \times 2^{R - D_H^c(\rho_{X_2 B_3})} \right].
$$

We have therefore established the following:

**Proposition 5 (Multihop).** Given $R \in \mathbb{R}_{\geq 0}$, $\varepsilon \in (0, 1)$, $b \in \mathbb{N}$, the triple $(b^{-1}R, b, \delta)$, is achievable for the classical-quantum relay channel, where

$$
\delta = b \left[ f(1, \varepsilon) + f(2, \varepsilon) + 4 \times 2^{R - D_H^c(\rho_{X_2 B_3})} \right] + 4 \times 2^{R - D_H^c(\rho_{X_1 X_2 B_2 B_3})}.
$$

At this point it would be useful to give the explicit form of $f(k, \varepsilon)$ for $k \in \mathbb{N}$ from [4], and our proof of the packing lemma in Section V:

$$
f(k, \varepsilon) = \varepsilon + 2^{k+1} \varepsilon^4 / 4 \left( 2^{2k+3} - 1 / 2 + 1 \right).
$$

Note that some coarse approximations are made to obtain a simple expression.

In the asymptotic limit we use the channel $n/b$ times in each of the $b$ blocks. The protocol will be analogous to one-shot protocol, except the relay channel will have a tensor product form $\Lambda_{X_1 X_2 \rightarrow B_2 B_3}^{\otimes n/b}$, characterized by a family of quantum states $\rho_{B_2}^{((X_1)^{n/b})(X_2)^{n/b})}$. The codebook will be $C^{n/b}$ and for finite

\footnote{Note that we need $R, \varepsilon$ to be sufficiently small so that $\delta \in [0,1]$. Otherwise we can simply take the minimum between the expression and 1. For large $b$, a more useful bound can be obtained by using the channel for finite $n$ times for each of the $b$ blocks.}
$b$ and large $n$ we will invoke Lemma 3 (instead of Lemma 2) to construct POVM’s for the relay and the receiver such that the decoding error vanishes if $R < \min \{ I(X_1; B_2|X_2)_{\rho}, I(X_2; B_3)_{\rho} \}$, thereby obtaining the quantum equivalent of the classical multihop bound for sufficiently large $b$:

$$C \geq \max_{p_{X_1}p_{X_2}} \min \{ I(X_1; B_2|X_2)_{\rho}, I(X_2; B_3)_{\rho} \}.$$  

(13)

### B. Coherent Multihop Scheme

In the multihop scheme, we obtained a rate optimized over product distributions, specifically Eq. (13). For the coherent multihop scheme we will obtain the same rate except optimized over all possible two-variable distributions $p_{X_1 X_2}$ by conditioning codewords on each other.

Again, let $R \geq 0$ be our rate, $\varepsilon \in (0, 1)$, and total blocklength $b \in \mathbb{N}$. We will show that we can achieve the triple $(\frac{b-1}{b} R, b, \delta)$ for some $\delta$ a function of $R, b, \varepsilon$. Let $p_{X_1 X_2}$ be probability distributions over $X_1 \times X_2$. Throughout, we will use

$$\rho_{X_1 X_2 B_2 B_3} \equiv \sum_{x_1, x_2} p_{X_1 X_2}(x_1, x_2) |x_1 x_2\rangle\langle x_1 x_2|_{X_1 X_2} \otimes \rho_{B_2 B_3}^{(x_1 x_2)}.$$  

We also again define $\rho_{B_3}^{(x_2)} \equiv \sum_{x_1} p_{X_1 X_2}(x_1|x_2) \rho_{B_3}^{(x_1 x_2)}$ to be the reduced state on $B_3$ by tracing out $X_1 B_2$ and fixing $X_2$. Our coding scheme will be very similar to that of the multihop.

**Code:** Let $G$ be a graph with $2b$ vertices corresponding to random variables $(X_1)_j (X_2)_j \sim p_{X_1 X_2}$, independent of other pairs, with edges from $(X_1)_j$ to $(X_2)_j$. Furthermore, let $M_0, M_j$ be index sets, where $|M_0| = 1$ and $|M_j| = 2^R$. Finally, the function ind maps $(X_1)_j$ to $\{j\}$ and $(X_2)_j$ to $\{j-1\}$. Then, letting $X \equiv X_0^b X_0^b$ and $M \equiv \gamma_{j=0}^b M_j$, it is easy to see that $B \equiv (G, X, M, \text{ind})$ is a multiplex Bayesian network. See Fig. 6 for a visualization when $b = 3$. Now, run Algorithm 1 with $B$ as the argument. This will return a random codebook

$$C = \bigcup_{j=1}^b \{(x_1)_j (m_{j-1}, m_j), (x_2)_j (m_{j-1})\}_{m_j \in M_j, m_{j-1} \in M_{j-1}};$$

where we restricted to the message components the codewords are dependent on via ind. For decoding we will apply Lemma 2 with this codebook and use the assortment of POVMs that are given for different ancestral subgraphs and other parameters.

---

10 Note that our rate is $\frac{b-1}{b} R$. To achieve rate $R$ we need $\frac{b-1}{b} \to 1$, and so we take the large $n$ limit followed by the large $b$ limit.
Encoding: Set $m_0$ to be the sole element of $M_0$. On the $j$-th transmission, the sender transmits a message $m_j \in M_j$ via $(x_1)_j(m_{j-1}, m_j) \in C$.

Relay encoding: Same as multihop.

Relay decoding: Same as multihop.

Decoding: Same as multihop.

Error analysis: With an analysis essentially identical to that of the multihop protocol we arrive at the following.

**Proposition 6 (Coherent Multihop).** Given $R \in \mathbb{R}_{\geq 0}$, $\varepsilon \in (0, 1)$, $b \in \mathbb{N}$, the triple $(\frac{b-1}{b} R, b, \delta)$ is achievable for the classical-quantum relay channel, where

$$
\delta = b\left[f_1(\varepsilon) + f_2(\varepsilon) + 4 \times 2^{-\mathcal{D}_H(\rho_{X_2B_3}\|\rho_{X_2B_3})} + 4 \times 2^{-\mathcal{D}_H(\rho_{X_1X_2B}\|\rho_{X_1X_2B})}\right].
$$

Asymptotically, this vanishes if $R < \min\{I(X_1; B_2 X_2)_{\rho}, I(X_2; B_3)_{\rho}\}$, thereby obtaining the quantum equivalent of the coherent multihop bound for sufficiently large $b$:

$$
C \geq \max_{\rho_{X_1X_2}} \min\{I(X_1; B_2 X_2)_{\rho}, I(X_2; B_3)_{\rho}\}.
$$

C. Decode Forward Scheme

In the decode-forward protocol we make an incremental improvement on the coherent multihop protocol by letting the receiver’s decoding also involve $X_1$.

Again, let $R \geq 0$ be our rate, $\varepsilon \in (0, 1)$, and total blocklength $b \in \mathbb{N}$. The classical-quantum state $\rho_{X_1X_2B_2B_3}$ is identical to that of the coherent multihop scenario.

**Code:** The codebook is generated in the same way as in the coherent multihop protocol save with the index set $M_b$ having cardinality 1 to take into account boundary effects for the backward decoding protocol\(^\text{12}\) we will implement.

**Encoding:** Set $m_0$ to be the sole element of $M_0$. On the $j$-th transmission, the sender transmits the message $m_j \in M_j$ via $(x_1)_j(m_{j-1}, m_j) \in C$. Note that there is only one message $m_b \in M_b$ they can choose on the $b$-th round.

**Relay encoding:** Same as that of coherent multihop.

**Relay decoding:** Same as that of coherent multihop. However, note that on $b$-th round, since $|M_b| = 1$, the decoding is trivial and the estimate $\hat{m}_b$ will be the sole element of $M_b$.

**Decoding:** The receiver waits until all $b$ transmissions are finished. Then, they implement a backward decoding protocol, that is, starting with the last system they obtain. Set $\hat{m}_b$ to be the

---

\(^{11}\) Note, however, that the POVM the relay uses from Lemma 2 will not be the same as that of the multihop case since the multiplex Bayesian networks are not the same.

\(^{12}\) In \cite{1} multiple decoding protocols are given. We here give the quantum generalization of the backward decoding protocol.
sole element of $M_0$. On the $j$-th system they use the POVM corresponding to the ancestral subgraph containing vertices $(X_1)_j$ and $(X_2)_j$, the set of quantum states $\{\rho_{B_3}^{(x_1,x_2)}\}_{x_1 \in X_1,x_2 \in X_2}$, decoding subset \{j-1\} $\subseteq$ \{j-1,j\}, and small parameter $\varepsilon$. We denote the POVM by $\{Q_{B_3}^{(m'_{j-1},m_j)}\}_{m'_{j-1} \in M_{j-1}}$, where we use the estimate $\hat{m}_j$, and the obtained measurement result $\hat{m}_{j-1}$. Note that trivially $\hat{m}_0$ is the sole element of $M_0$.

**Error analysis:** Fix some $m = (m_0,\ldots,m_b) \in M$. Let $\hat{m} = (\hat{m}_0,\ldots,\hat{m}_b), \tilde{m} = (\hat{m}_0,\ldots,\hat{m}_b)$ denote the aggregation of the messages estimates of the relay and receiver, respectively. Then, the probability of error averaged over $C$ is given by

\[ p_e(C) = \mathbb{E}_C [p(\hat{m} \neq m)]. \]

Again, by the bound in Eq. (2),

\[ p_e(C) \leq \mathbb{E}_C [p(\hat{m} \neq m)] + \mathbb{E}_C [p(\hat{m} \neq m|\tilde{m} = m)]. \]

The bound on the first term is identical to that of the coherent multihop protocol and is given by

\[ \mathbb{E}_C [p(\hat{m} \neq m)] \leq b \left[f_2(\varepsilon) + 4 \times 2^{R - D_H^r(\rho_{X_1X_2B3})} \right]. \]

For the second term, we first apply the union bound:

\[ \mathbb{E}_C [p(\hat{m} \neq m|\tilde{m} = m)] \leq \mathbb{E}_C \left[ \sum_{j=1}^{b-1} p(\hat{m}_j \neq m_j | \hat{m}_{j+1} = m_{j+1} \land \tilde{m} = m) \right], \]

where we take into account that the terms corresponding to 0 and $b$ vanish by definition. Each of the summands can be bounded via Lemma 2:

\[
\mathbb{E}_C [p(\hat{m}_j \neq m_j | \hat{m}_{j+1} = m_{j+1} \land \tilde{m} = m)] \\
= \mathbb{E}_C \left[ \text{tr} \left[ (I - Q_{B_3}^{(m_j,m_{j+1})}) \rho_{B_3}^{(x_1,x_2)} \right] \right] \\
= \mathbb{E}_C_{(X_1)_{j+1}(X_2)_{j+1}} \left[ \text{tr} \left[ (I - Q_{B_3}^{(m_j,m_{j+1})}) \rho_{B_3}^{(x_1,x_2)} \right] \right] \\
\leq f_2(\varepsilon) + 4 \sum_{T = \{j\}} 2^{R - D_H^r(\rho_{X_1X_2B3})} \| \rho_{X_1X_2B3} \|, \\
\leq f_2(\varepsilon) + 4 \times 2^{R - D_H^r(\rho_{X_1X_2B3})},
\]

where we use that $S_{\{j\}} = \{(X_1)_{j+1}(X_2)_{j+1}\}$. Hence, we conclude that

\[ \mathbb{E}_C [p(\hat{m} \neq m|\tilde{m} = m)] \leq b \left[ f_2(\varepsilon) + 4 \times 2^{R - D_H^r(\rho_{X_1X_2B3})} \right]. \]

We conclude the following.

**Proposition 7** (Decode Forward). Given $R \in \mathbb{R}_{>0}$, $\varepsilon \in (0,1)$, $b \in \mathbb{N}$, the triple $(\frac{1}{b+1}R, b, \delta)$ is achievable for the classical-quantum relay channel where

\[ \delta = b \left[ 2f_2(\varepsilon) + 4 \times \left( 2^{R - D_H^r(\rho_{X_1X_2B3})} + 2^{R - D_H^r(\rho_{X_1X_2B3})} \right) \right]. \]

Asymptotically, this vanishes if $R < \min \{I(X_1;B_2|X_2)_\rho, I(X_1X_2;B_3)_\rho \}$, thereby obtaining the decode-forward lower bound for sufficiently large $b$: 

\[ C \geq \max_{p_{X_1X_2}} \min \{I(X_1;B_2|X_2)_\rho, I(X_1X_2;B_3)_\rho \}. \]
D. Partial Decode Forward Scheme

We now derive the partial decode-forward lower bound. This will actually require the full power of Lemma 2 as the receiver will decode all the messages simultaneously by performing a joint measurement on all b blocks. Intuitively, the partial decode-forward builds on the decode-forward by letting the relay only decode and pass on a part, what we will call P, of the overall message.

We will split the message into two parts P and Q with respective rates $R_P, R_Q \geq 0$. Let $\varepsilon \in (0, 1)$ and $b \in \mathbb{N}$ be the total blocklength. Choose some distribution $p_{X_1X_2}$ but also a random variable $U$ correlated with $X_1X_2$ so that the overall distribution is $p_{UX_1X_2}$. The classical-quantum state of interest will be

$$\rho_{UX_1X_2B_2B_3} \equiv \sum_{u,x_1,x_2} p_{UX_1X_2}(u,x_1,x_2) \langle u x_1 x_2 | U X_1 X_2 \otimes \rho_{B_2B_3}^{(x_1,x_2)} \rangle.$$  \hfill (14)

Note that $\rho_{B_2B_3}^{(x_1,x_2)}$ does not depend on $u$, but sometimes we will $\rho_{B_2B_3}^{(u x_1 x_2)} = \rho_{B_2B_3}^{(x_1 x_2)}$ to keep notation simple. However, if we trace over $X_1$, we will induce a $u$ dependence via the correlation between $U$ and $X_1X_2$:

$$\rho_{UX_2B_2B_3} = \sum_{u,x_2} p_{UX_2}(u,x_2) \langle u x_2 | U X_2 \otimes \rho_{B_2B_3}^{(x_2)} \rangle,$$

where

$$\rho_{B_2B_3}^{(u x_2)} \equiv \sum_{x_1} p_{X_1|UX_2}(x_1|u,x_2) \rho_{B_2B_3}^{(x_1,x_2)}.$$  

This will be important for the relay decoding.

**Code:** Let $G$ be a graph with $3b$ vertices corresponding to random variables $(U)_j (X_1)_j (X_2)_j \sim p_{UX_1X_2}$. The graph has edges going from $(X_2)_j$ to $(U)_j$ and $(U)_j$ to $(X_1)_j$ for all $j$ and no edges going across blocks with different $j$’s. Furthermore, let $P_b, P_j$ and $Q_j$ be index sets, so that $J = [0 : b] \sqcup [b]$, where $|P_b| = |P_0| = |Q_b| = 1$, $|P_j| = 2^{R_P}$ and $|Q_j| = 2^{R_Q}$ otherwise. Finally, the function ind maps $(X_1)_j \text{ to }^{13} \{P_j, Q_j, P_{j-1}\}$, $(U)_j$ to $\{P_j, P_{j-1}\}$, and $(X_2)_j$ to $\{P_{j-1}\}$. Then, letting $X = U^b X^b_1 X^b_2$, $M_p = \bigcup_{j=0}^b P_j$, $M_q = \bigcup_{j=1}^b Q_j$ and $M = M_p \times M_q$, it is easy to see that $\mathcal{B} \equiv (G, X, M, \text{ind})$ is a multiplex Bayesian network. See Fig. 7 for a visualization when $b = 3$. Now, run Algorithm 1 with $\mathcal{B}$ as the argument. This will return a random codebook

$$C = \bigcup_{j=1}^b \{(x_1)_j (p_{j-1}, p_j, q_j), (u)_j (p_{j-1}, p_j), (x_2)_j (p_{j-1})\}_{p_j \in P_j, p_{j-1} \in P_{j-1}, q_j \in Q_j},$$

where we restricted to the message components the codewords are dependent on via ind. For decoding we will apply Lemma 2 with this codebook and use the assortment of POVMs that are given for different ancestoral subgraphs and other parameters.

**Encoding:** Set $p_0$ to be the sole element of $P_0$. On the $j$-th transmission, the sender transmits the two-part message $(p_j, q_j) \in P_j \times Q_j$ via $(x_1)_j (p_{j-1}, p_j, q_j) \in C$. Note that on the $b$-th transmission the sender has to send a fixed message $(p_b, q_b)$ being the sole element of $P_b \times Q_b$.

\hfill \footnote{For convenience we will denote the elements of $J$ by the index sets they correspond to.}
Since the only index sets which are not included in the part to be decoded are all of cardinality 1, we omit the index sets.

We can bound the first term just as we did for the other protocols. First, use the union bound.

Relay encoding: Let \( \tilde{p}_0 \) be the sole element of \( P_0 \). On the \( j \)-th transmission, the relay sends \( \tilde{p}_{j-1} \) via \( (x_2)_j(\tilde{p}_{j-1}) \) from codebook \( C \). Note that this is the relay’s estimate of the message sent by the sender on the \( j - 1 \)-th transmission.

Relay decoding: The relay will try to recover the \( p \)-part of the sender’s message using the same technique as in the previous protocols. On the \( j \)-th transmission the relay will use the POVM corresponding to the ancestral subgraph containing the two vertices \( (U)_j \) and \( (X_2)_j \), the set of quantum states \( \{ \rho_{B_2}^{(u,x_2)} \} \), decoding subset \( \{ P_j \} \subseteq \{ P_{j-1}, P_j \} \), and small parameter \( \varepsilon \). The POVM is denoted by \( \{ Q_{B_2}^{(p_j|p_{j-1})} \} \), where we use the estimate \( \tilde{p}_{j-1} \), and the relay applies this on their received state to obtain a measurement result \( \tilde{p}_j \). Note that \( \tilde{p}_b \) is trivially the sole element of \( P_b \).

Decoding: The decoder waits until all \( b \) transmissions are completed. The receiver will use the POVM corresponding to the ancestral subgraph the entire graph \( G \), the set of quantum states \( \{ \times_{j=1}^b \rho_{B_3}^{(u)} \} \), where the \( (u) \) dependence here is trivial, decoding set \( \times_{j=1}^{b-1} P_j \times \times_{j=1}^{b-1} Q_j \), and small parameter \( \varepsilon \). We denote the POVM by \( \{ Q_{B_3}^{p_i|p_{i-1}} \} \), to their received state on \( B_3 \) to obtain their estimate of the entire string of messages, which we call \( \tilde{m}_p \equiv (\tilde{p}_0, \ldots, \tilde{p}_b) \), \( \tilde{m}_q \equiv (q_1, \ldots, q_b) \), where \( \tilde{p}_0, \tilde{p}_b, q_b \) are set to be the sole elements of the respective index sets.

Error analysis: We fix the strings of messages \( m_p = (p_0, \ldots, p_b) \) and \( m_q = (q_1, \ldots, q_b) \). By the bound in Eq. (2),

\[
p_e(C) \equiv E_C[p(\tilde{m}_p \neq m_p m_q)] \leq E_C[p(\tilde{m}_p \neq m_p)] + E_C[p(\tilde{m}_p \neq m_p m_q | \tilde{m}_p = m_p)].
\]

We can bound the first term just as we did for the other protocols. First, use the union bound.

\[
E_C[p(\tilde{m}_p \neq m_p)] \leq \sum_{j=1}^{b-1} E_C[p(\tilde{p}_j \neq p_j | \tilde{p}_{j-1} = p_{j-1})].
\]

Since the only index sets which are not included in the part to be decoded are all of cardinality 1, we omit the conditioning for conciseness.
By Lemma 2 we can bound each summand as follows:

\[
\mathbb{E}_C[p(\tilde{p}_j\neq p_j)j_{\tilde{p}_j-1} = p_{j-1})] = \mathbb{E}_C\left[ \text{tr}\left( (I - Q_B)_{p(B_2)}(1)_{(p_j(p_j-1)j_{(x_j)}(x_j(p_j-1)j_1)}) \right) \right]
\]

\[
= \mathbb{E}_C[(U_j(x_j)_{(x_j)_{j-1}}(p_j(p_j-1)j_{(x_j)}(x_j(p_j-1)j_1))]
\]

\[
\leq f_2(\varepsilon) + 4 \sum_{T=\{p\}} 2^{R_p-\delta_H^2(U_{p_x}(x_j)_{(x_j)_{j-1}}(p_j(p_j-1)j_{(x_j)}(x_j(p_j-1)j_1))]
\]

\[
= f_2(\varepsilon) + 4 \times 2^{R_p-\delta_H^2(U_{p_x}(x_j)_{(x_j)_{j-1}}(p_j(p_j-1)j_{(x_j)}(x_j(p_j-1)j_1))]
\]

where we used \(S_{\{p\}} = \{(U_j)_{j}\}\). We dropped the index \(j\) in the last equality since \(U_j(x_j)_{(x_j)_{j-1}}(p_j(p_j-1)j_{(x_j)}(x_j(p_j-1)j_1))\). Hence, overall,

\[
\mathbb{E}_C[p(\tilde{m}_p \neq m_p)] \leq b \left[ f_2(\varepsilon) + 4 \times 2^{R_p-\delta_H^2(U_{p_x}(x_j)_{(x_j)_{j-1}}(p_j(p_j-1)j_{(x_j)}(x_j(p_j-1)j_1))]
\]

For the second term, we again invoke Lemma 2:

\[
\mathbb{E}_C[p(\tilde{m}_p \neq m_p)\tilde{m}_q = m_q = m_p)] =
\]

\[
= \mathbb{E}_C\left[ \text{tr}\left( (I - Q_B)_{p(B_2)}(1)_{(p_j(p_j-1)j_{(x_j)}(x_j(p_j-1)j_1))} \right) \right]
\]

\[
= \mathbb{E}_C\left[ \text{tr}\left( (I - Q_B)_{p(B_2)}(1)_{(p_j(p_j-1)j_{(x_j)}(x_j(p_j-1)j_1))} \right) \right]
\]

\[
\leq f_3(\varepsilon) + 4 \sum_{J_p, J_q \subseteq [b-1]:j_p + j_q > 0} 2^{j_p R_p + j_q R_q - \delta_H^2(U_{p_x}(x_j)_{(x_j)_{j-1}}(p_j(p_j-1)j_{(x_j)}(x_j(p_j-1)j_1))]
\]

where \(S_{\{j_p, j_q\}} = \{X_1^f X_2^f U_{(j_p)(j_q)}\}\). Here we used the following definitions \(J \equiv J_p \cup J_q, J_p \equiv J_p \cup J_p', J_q \equiv (j \in [b] j - 1 \in J_p)\), and \(J_p \equiv J_p, J_q \equiv J_q\). Also, note that \(\rho_{U_{p_x}X_1^{p_x}X_2^{p_x}} = \rho_{U_{p_x}X_1X_2B_3}^b\). Thus, overall, we have proved

**Proposition 8.** Given \(R_p, R_q \in \mathbb{R}_{\geq 0}, \varepsilon \in (0, 1), b \in \mathbb{N}\), the triple \((\frac{b-1}{b}(R_p + R_q), b, \delta)\) is achievable for the classical-quantum relay channel, where

\[
\delta = b \left[ f_2(\varepsilon) + 4 \times 2^{R_p-\delta_H^2(U_{p_x}X_1^{p_x}X_2^{p_x})} \right]
\]

\[
+ f_3(\varepsilon) + 4 \sum_{J_p, J_q \subseteq [b-1]:j_p + j_q > 0} 2^{j_p R_p + j_q R_q - \delta_H^2(U_{p_x}X_1^{p_x}X_2^{p_x})}.
\]

In the asymptotic limit, the error vanishes provided

\[
R_p < I(U; B_2|X_2)
\]

and, for all \(J_p, J_q \subseteq [b-1],

\[
j_p R_p + j_q R_q < I(X_1^f X_2^f U_{(j_p)(j_q)}; B_3^b|X_1^f X_2^f U_{(j_p)(j_q)})_{\rho_{U_{p_x}X_1^{p_x}X_2^{p_x}}^b}.
\]

(15)
Note $J_p, J_q \subseteq [b - 1]$ and $J'_p \subseteq [2 : b]$. However, we will use the convention that all complementary sets are with respect to largest containing set [15] [b]. We can simplify Eq. (16) via a general lemma:

**Lemma 9.** Let $\rho_{B_1 \ldots B_m}$ be $m$-partite quantum state. We consider the state $\rho_{B_1 \ldots B_m}^{\otimes n}$ for some $n \in \mathbb{N}$. Now, let $B, B', C$ be disjoint subsystems of $(B_1 \ldots B_m)^{\otimes n}$ and such that $B, B'$ are supported on disjoint tensor factors. Then,

$$I(B; B'|C) = 0.$$ 

**Proof.** We prove this by the definition of the conditional mutual information and the fact that $\rho_{B_1 \ldots B_m}^{\otimes n}$ is a tensor product state:

$$I(B; B'|C) = S(BC) + S(B'C) - S(BB'C) - S(C)$$

$$= S(BC_B) + S(C_{\overline{B}}) + S(B'C_{B'}) + S(C_{\overline{B'B'}}) - S(BB'C) - S(B'C_{B'}) - S(C_{\overline{B'B'}}) - S(C)$$

$$= 0.$$ 

where $C_B$ is the subsystem of $C$ supported on the tensor factors that support $B$ and $C_{\overline{B}}$ is the rest of $C$. \hfill \square

Thus, using this and the chain rule, for any conditional mutual information quantity we can remove conditioning systems which are supported on tensor factors disjoint from those that support the non-conditioning systems. This will be key in the following analyses. For instance, in Eq. (16), $\mathcal{J}$ and $\mathcal{J} \cup J'_p \cup J_p = \mathcal{J}$ are supported on disjoint tensor factors, and so we can remove the conditioning on the $X_1^{\mathcal{J}}$ system:

$$I(X_1^{\mathcal{J}} X_2^{J_p} U_{J_p}; B_3|X_2^{\mathcal{J}} X_2^{T_p} U_{J_p'}) = I(X_1^{\mathcal{J}} X_2^{J_p} U_{J_p}; B_3|X_2^{\mathcal{J}} X_2^{T_p} U_{J_p'}) + I(X_1^{\mathcal{J}} X_2^{J_p} U_{J_p}; X_2^{\mathcal{J}} B_3 X_2^{\mathcal{T_p} U_{J_p'}}) - I(X_1^{\mathcal{J}} X_2^{J_p} U_{J_p}; X_2^{\mathcal{J}} X_2^{T_p} U_{J_p'})$$

Thus, Eq. (16) reduces to

$$j_p R_p + j_q R_q < I(X_1^{\mathcal{J}} X_2^{J_p} U_{J_p}; B_3|X_2^{T_p} U_{J_p'}).$$

We claim that the set of pairs $(R_p, R_q)$ that satisfy these bounds gives the classical partial decode-forward lower bound with quantum mutual information quantities in the limit of large $b$.\footnote{This will also cause $\frac{b - 1}{b} \rightarrow 1$ so that the rate we achieve really is $R_p + R_q$.}

In particular, we show:

**Proposition 10.** Let

$$S(b) \equiv \left\{ (R_p, R_q) \in \mathbb{R}^2_{\geq 0} \mid \forall J_p, J_q \subseteq [b - 1] \text{ such that } j_p + j_q > 0, j_p R_p + j_q R_q < I(X_1^{\mathcal{J}} X_2^{J_p} U_{J_p}; B_3|X_2^{T_p} U_{J_p'})_{\rho_{UX_1 X_2 B_3}} \right\}$$

and

$$S \equiv \left\{ (R_p, R_q) \in \mathbb{R}^2_{\geq 0} \mid R_q < I(X_1; B_3|X_2)_{\rho_{UX_1 X_2 B_3}}, R_p + R_q < I(X_1 X_2; B_3)_{\rho_{UX_1 X_2 B_3}} \right\},$$

where $\rho_{UX_1 X_2 B_2 B_3}$ is given by Eq. (14). Then, $\lim_{b \rightarrow \infty} S(b)$ exists and is equal to $S$.\footnote{The $b$-th messages and estimates will match, but in general the $b$-th $x_1, x_2, u$ depend also on the $b - 1$-th messages and estimates.}
Note that the bounds that define $S$ do not match the bounds given for instance in [1] since we do not first decode $P$ and thereby $Q$, but instead jointly decode to obtain all of $P, Q$. However, in the end we will still obtain the same lower bound on the capacity.

**Proof.** For reference, we list the bounds:

$$j_p R_p + j_q R_q < I(X_1 X_2; B_3 | J_p U; X_{J_p} J_q)$$

and

$$R_q < I(X_1 | X_2; B_3)$$

$$R_p + R_q < I(X_1 X_2; B_3)$$

We first claim $\limsup_{b \to \infty} S(b) \subseteq S$. Consider $J_p, J_q = [b-1]$, in which case Eq. (17) becomes

$$(b-1)(R_p + R_q) < I(X_1(X_2)_b^{b+1} | B_3(X_2)_{b+1})$$

which, using Lemma 9, can be manipulated into

$$R_p + R_q < \frac{b}{b-1} I(X_1 X_2; B_3) - \frac{1}{b-1} I(X_2; B_3)$$

$$= I(X_1 X_2; B_3) + \frac{1}{b-1} I(X_1; B_3 X_2).$$

In the limit of large $b$, this becomes Eq. (19). To obtain Eq. (18), take $j_p = 0$. Then, Eq. (17) becomes by Lemma 9

$$j_q R_q < I(X_1^{J_q} | B_3 X_2^{b+1} U; B_3) = j_q I(X_1; B_3 X_2 U).$$

Now, since $j_p = 0$, $j_q$ cannot be zero, so this is equivalent to

$$R_q < I(X_1; B_3 X_2 U).$$
The claim thus follows. We next claim $S(b) \supseteq S$ for all $b$ and so $\liminf_{b \to \infty} S(b) \supseteq S$. We only need to consider when $j_p > 0$ since otherwise we obtain $\text{Eq. (18)}$ as shown above, which holds for all $b$. Now, interpret each of the inequalities above as a linear bound on an $R_p-R_q$ diagram. We will show that none of the lines corresponding to $\text{Eq. (17)}$ cuts into $S$. First, fixing $j_p, j_q \leq [b-1]$, we find the $R_p$ intercept of said line

$$
\frac{1}{j_p} I(X_1^J X_2^J U^J; B_3^J | X_2^J U^J) = \frac{1}{j_p} \left( I(X_1^J X_2^J \overline{U}^J; B_3^J | X_2^J U^J) + \cdots \right)
$$

$$
\geq \frac{1}{j_p} I(X_1^J X_2^J \overline{U}^J; B_3^J)
$$

$$
= I(X_1 X_2 U; B_3) = I(X_1 X_2; B_3),
$$

where $\cdots$ stands for some conditional mutual information quantity and therefore is non-negative. Thus, the $R_p$ intercept is at least as large as that of $\text{Eq. (19)}$, as shown in Fig. 8. This determines one of the points of the line.

We now find another point. We observe that $I(X_1; B_3 | X_2 U) \leq I(X_1 X_2 U; B_3)$ so the line associated with $\text{Eq. (18)}$ intersects that of $\text{Eq. (19)}$ in $\mathbb{R}_{\geq 0}^2$. Hence, it is sufficient to show the bound on $R_p$ when $R_q = I(X_1; B_3 | X_2 U)$ in $\text{Eq. (17)}$ is weaker than $I(X_1 X_2 U; B_3) - I(X_1; B_3 | X_2 U) = I(X_2 U; B_3)$. To see this, we plug in $R_q = I(X_1; B_3 | X_2 U)$ into $\text{Eq. (17)}$:

$$
\begin{align*}
&j_p R_p + j_q I(X_1; B_3 | X_2 U) \\ &\quad \leq I(X_1^J X_2^J U^J; B_3^J | X_2^J U^J) \\ &\quad = I(X_1^J; B_3^J | X_1^J U^J X_2^J U^J) + I(X_1^J \setminus J_0) X_2^J U^J; B_3^J | X_2^J U^J) \\ &\quad = I(X_1^J; B_3^J | X_2^J U^J) + I(X_2^J U^J; B_3^J | X_2^J U^J) + \cdots \\ &\quad = j_q I(X_1; B_3 | X_2 U) + j_p I(X_2 U; B_3) + \cdots .
\end{align*}
$$

This establishes our claim and completes the proof.

Therefore, combining the bounds $\text{Eqs. (15), (18) and (19)}$, the overall rate $R_p + R_q$ of the entire protocol is achievable if

$$
R_p + R_q < \min \{ I(X_1; B_3 | U X_2)_\rho + I(U; B_2 | X_2)_\rho, I(X_1 X_2; B_3)_\rho \}. 
$$

This is sufficient since if it holds we can choose $R_p, R_q$ to satisfy the bounds. It is also necessary since if it is violated, then one of the bounds has to be violated. We can optimize over $p_{UX_1X_2}$ so we obtain the partial decode-forward lower bound:

$$
C \geq \max_{p_{UX_1X_2}} \min \{ I(X_1; B_3 | U X_2)_\rho + I(U; B_2 | X_2)_\rho, I(X_1 X_2; B_3)_\rho \}. 
$$

(20)

**Remark.** This coding scheme is optimal in the case when $\mathcal{N}_{X_1 X_2} \rightarrow B_2 B_3$ is semideterministic, namely $B_2$ is classical and $p_{B_2}^{(x_1 x_2)}$ is pure for all $x_1, x_2$. This is because in this case the partial decode-forward lower bound $\text{Eq. (20)}$ with $U = B_2$ as random variables matches the cutset upper bound $\text{Eq. (11)}$. This is possible because of the purity condition, which essentially means $B_2$ is a deterministic function of $X_1, X_2$. The semideterministic classical relay channel was defined and analyzed in [33].

V. PROOF OF THE QUANTUM MULTIPARTY PACKING LEMMA

In this section we prove Lemma 2 via Sen’s joint typicality lemma [4]. We then use Lemma 2 to prove the asymptotic version, Lemma 3. We shall state a special case of the joint typicality lemma,
the $t = 1$ intersection case in the notation of [4], as a theorem. For the sake of conciseness, we suppress some of the detailed expressions.

We first give some definitions. A subpartition $L$ of some set $S$ is a collection of nonempty, pairwise disjoint subsets of $S$. We define $\bigcup(L)$ to be their union, that is, $\bigcup(L) = \bigcup_{L \in L} L$. Note that $\bigcup(L) \subseteq S$. We say a subpartition $L$ of $S$ covers $T \subseteq S$ if $T \subseteq \bigcup(L)$.

**Theorem 11** (One-shot Quantum Joint Typicality Lemma [4]). Let

$$\rho_{XA} = \sum_x p_X(x) |x\rangle_x \otimes \rho_A^{(x)}$$

be a classical-quantum state where $A \equiv A_1 \ldots A_N$ and $X \equiv X_1 \ldots X_M$. Let $\varepsilon \in (0, 1)$ and let $Y = Y_1 \ldots Y_{N+M}$ consist of $N + M$ identical copies of some classical system, with total dimension $d_Y$. Then there exist quantum systems $\hat{A}_k$ and isometries $\hat{J}_k : \hat{A}_k \to \hat{A}_k$ for $k \in [N]$, as well as a cqc-state of the form

$$\hat{\rho}_{XAY} = \frac{1}{d_Y} \sum_{x,y} p_X(x) |x\rangle_x \otimes \hat{\rho}_A^{(x,y)} \otimes |y\rangle_y \langle y|_Y,$$

and a cqc-POVM $\hat{\Pi}_{XAY}$, such that, with $\hat{J} \equiv \bigotimes_{k \in [N]} \hat{J}_k$,

1. $\|\hat{\rho}_{XAY} - (1_X \otimes \hat{J})\rho_{XA}(1_X \otimes \hat{J})^\dagger \otimes \tau_Y\|_1 \leq f(N, M, \varepsilon)$, where $\tau_Y = \frac{1}{d_Y} \sum_y |y\rangle \langle y|_Y$ denotes the maximally mixed state on $Y$,

2. $\tr[\hat{\Pi}_{XAY} \hat{\rho}_{XAY}] \geq 1 - g(N, M, \varepsilon)$.

3. Let $\mathcal{L}$ be a subpartition of $[M] \cup [N]$ that covers $[N]$. Define $Y_L := Y_{\bigcup(L)}$, $S \equiv [M] \cap \bigcup(L)$, $\mathcal{S} \equiv [M] \setminus S$ and the “conditional” quantum states

$$\hat{\rho}_{XSA}^{(x,y)} = \frac{1}{d_Y} \sum_{s,y \in \mathcal{L}} p_X(s|\mathcal{L}) \rho_{X_S}^{(x,y)} |s\rangle_{X_S} \otimes \hat{\rho}_A^{(x,y)} \otimes |y\rangle_{\mathcal{L}} \langle y|_{\mathcal{L}}$$

$$\rho_{XSA}^{(x)} = \sum_x p_X(s|\mathcal{L}) \rho_{X_S}^{(x)} |s\rangle_{X_S} \otimes \rho_A^{(x)}.$$

We can now define

$$\hat{\rho}_{XAY}^{(L)} = \frac{1}{d_Y} \sum_{x,y \in \mathcal{L}} p_{X_S}^{(x,y)} |x\rangle_{X_S} \otimes \bigotimes_{L \in L} p_{X_L|\mathcal{L}}^{(x)} p_{A_L|\mathcal{L}}^{(y)} |y\rangle_{Y_L} \langle y|_{Y_L}$$

$$\rho_{XAY}^{(L)} = \sum_x p_{X_S}^{(x)} |x\rangle_{X_S} \otimes \bigotimes_{L \in L} p_{X_L|\mathcal{L}}^{(x)} p_{A_L|\mathcal{L}}^{(y)} |y\rangle_{Y_L} \langle y|_{Y_L}$$

in terms of the reduced density matrices of the states $\hat{\rho}_{XSA}^{(x,y)}$ and $\rho_{XSA}^{(x)}$ defined above. Then,

$$\tr[\hat{\Pi}_{XAY} (\hat{\rho}_{XAY}^{(L)})] \leq 2^{-D_{\mathcal{H}}(\rho_{XAY}^{(L)} \| \hat{\rho}_{XAY})} + h(N, M, d_A, d_Y).$$

Here, $f(N, M, \varepsilon)$, $g(N, M, \varepsilon)$, $h(N, M, d_A, d_Y)$ are universal functions (independent of the setup) such that

$$\lim_{\varepsilon \to 0} f(N, M, \varepsilon) = \lim_{\varepsilon \to 0} g(N, M, \varepsilon) = \lim_{d_Y \to \infty} h(N, M, d_A, d_Y) = 0.$$
Proof. This follows readily from Sen’s Lemma 1 in [4] with an appropriate change of notation and suitable simplifications. We will use Sen’s terminology and notation. We choose \(k_{Sen} \equiv N, \epsilon_{Sen} \equiv M, L_{Sen}\) a system isomorphic to our \(Y_k, \delta_{Sen} = \epsilon_{Sen}^{1/4N}\), and the same error \(\epsilon\) for each pseudosubpartition of \([M] \sqcup [N]\). We denote \(A_k \equiv (A_k^n)_{Sen}\), so that \((A_k^n)_{Sen} = A_kY_k\) and \((X_k^n)_{Sen} = X_kY_k\); that is, we explicitly include the augmenting systems in our notation. We also write \(J_k\) for the natural embedding \(A_k \to A'_k\). Then Sen’s lemma yields a state \(\hat{\rho}_{X_AY} = \rho_{Sen}'\) and a POVM \(\hat{\Pi}_{X_AY} = \Pi_{Sen}'\) that satisfies all desired properties. First, statement 1 in Sen’s lemma asserts that \(\hat{\rho}_{X_AY}\) and \(\hat{\Pi}_{X_AY}\) are cqc. Next, our properties 1 and 2 are direct restatements of his statements 2 and 3, with \(f(N, M, \epsilon) = 2^{(N+M)/2 + 1}\epsilon^{1/4N}\) and \(g(N, M, \epsilon) = 2^{2MN + 4(N+1)^N} 2^{(N+M)^2} \epsilon^{1/2} + 2(N+M)^2 \epsilon^{1/4N}\). Finally, we apply statement 4 in Sen’s lemma to a subpartition \(\mathcal{L}\) covering \([N]\) and the probability distribution \(q_{Sen}(x) = p_{X_A}(x)\prod_{L \in \mathcal{L}} p_{X_L \mid [M]}(x_L \mid [x_M])\). Then our \(\rho_{X_A}^{(C)}\) is Sen’s \(\rho(s_1, ..., s_l)\) and our \(\rho_{X_A}^{(C)}\) is Sen’s \(\rho'_{(s_1, ..., s_l)}\), so we obtain property 3 with \(h(N, M, d_A, d_Y) = 32^N d_A d_Y^{-1/2(N+M)}\). \(\square\)

Now, we will prove a generalization of Lemma 2 which takes greater advantage of the power of Theorem 11 by abstracting the properties that the random codebook \(C\) needs to satisfy for the multiparty packing lemma to hold. We will use the notation \(X = X_1 \ldots X_k\) to denote set of \(k \in \mathbb{N}\) systems.

**Lemma 12.** Let \(\{p_X, \rho_B^x\}\) be an ensemble of quantum states, where \(X \equiv X_1 \ldots X_k\) with \(k \in \mathbb{N}\), \(\mathcal{I} = \mathcal{I}_1 \times \mathcal{I}_2\) an index set, and \(\epsilon \in (0, 1)\) a small parameter. Now, let \(\mathcal{C} = \{x(i)\}_{i \in \mathcal{I}}\) be a family of random variables such that for every \(i \in \mathcal{I}, x(i) \sim p_X, j = 1 \ldots k\), and there exists a map\(^\dagger\) \(\Psi : \mathcal{I} \times \mathcal{I} \to \mathcal{P}([k])\) such that for every \(i, i' \in \mathcal{I}\), letting \(T \equiv \Psi(i, i')\),

1. \(x_T(i) = x_T(i')\) as random variables
2. \(x_T(i), x_T(i')\) are independent conditioned on \(x_T(i) = x_T(i')\),

where \(T = [k] \setminus T\). Then, for each \(i_1 \in \mathcal{I}_1\) there exists a POVM \(\{Q_{B}^{(i_2 i_1)}\}_{i_2 \in \mathcal{I}_2}\) dependent on the random variables \(C\) such that for all \(i = (i_1, i_2) \in \mathcal{I}\),

\[
\mathbb{E}_C \left[ \text{tr}\left( (I - Q_B^{(i_2 i_1)}) \rho_B^{(x(i_1, i_2))} \right) \right] \leq f(k, \epsilon) + 4 \sum_{i_2' \neq i_2} 2^{-D_H(\rho_B^{(x(i_1, i_2) \otimes x(i_1, i_2'))})},
\]

where \(\mathbb{E}_C\) is the expectation over the random variables in \(\mathcal{C}\), \(S \equiv \Psi((i_1, i_2), (i_1, i_2'))\), and

\[\rho_{XB} \equiv \sum_x p_X(x) \left| x \right> \left< x \otimes \rho_B^x\right].\]

Furthermore, \(f(k, \epsilon)\) is a universal function such that \(\lim_{\epsilon \to 0} f(k, \epsilon) = 0\).

Before we prove Lemma 12, we first show that Lemma 2 follows from Lemma 12 by establishing that the random codebook generated by Algorithm 1 satisfies the required properties.

**Proof of Lemma 2.** Fix subgraph \(H, \{\rho_B^{x_H}\}_{x_H \in x_H}, D \subseteq J_H, \epsilon \in (0, 1)\). We invoke Lemma 12 with the ensemble \(\{p_{X_H}, \rho_B^{x_H}\}\) with \(k = [V_H], I_1 = M_H, I_2 = M_D, \text{ the same } \epsilon, \text{ and the family of random variables } \mathcal{C} = C_H\). We thus identify \(\mathcal{I} = M_H = M_D \times M_H\). We also define an arbitrary ordering on \(V_H\) such that we can identify it with \([k]\).

We check that \(C_H\) satisfies the required properties using the observations we made regarding Algorithm 1. First, for every \(m_H \in M_H, x_H(m_H) \sim p_{X_H}\) by observation 1 on p. 7.

\(^\dagger\) Note that the bound does not depend on the specific choice of the map.
Next, we claim the map
\[ \Psi(m_H, m_H') = \{ v \in V_H \mid \exists j \in \text{ind}(v) \text{ such that } (m_D)_j \neq (m_D')_j \} \]
satisfies the required conditions. Let \( m_H, m_H' \in M_H \) and \( T = \Psi(m_H, m_H') \). By definition, given \( v \in T \), for all \( j \in \text{ind}(v) \), \( (m_H)_j = (m_H')_j \). Hence, \( m_H|_{\text{ind}(v)} = m_H'|_{\text{ind}(v)} \), so by observation 2 on p. 7, \( x_v(m_H) = x_v(m_H') \) as random variables. Thus, \( x_T(m_H) = x_T(m_H') \) as random variables, so we have established condition 1.

We now prove the conditional independence statement in condition 2 is satisfied. For \( \xi_T \in X_T \), observation 1 shows that
\[ \Pr( x_T(m_H) = \xi_T ) = \prod_{v \in T} p_{X_v|X_{\text{pa}(v)}}(\xi_v|\text{pa}(v)) , \]
where we used that \( \text{pa}(T) \subseteq \overline{T} \) as a consequence of Eq. (7). Next, observation 3 implies that the joint distribution of \( x_T(m_H), x_T(m_H') \), and \( x_T(m_H) \) is given as follows. For \( \xi, \xi' \in X \) such that \( \xi_T = \xi_T' \),
\[ \Pr( x_T(m_H) = \xi_T, x_T(m_H') = \xi_T', x_T(m_H) = \xi_T ) = \Pr( x(m_H) = \xi, x(m_H') = \xi' ) \]
\[ = \left( \prod_{v \in T} p_{X_v|X_{\text{pa}(v)}}(\xi|\text{pa}(v)) \right) \left( \prod_{v \in T} p_{X_v|X_{\text{pa}(v)}}(\xi'|\text{pa}(v)) \right) \] .
Hence, \( x_T(m_H) \) and \( x_T(m_H') \) are independent conditional on \( x_T(m_H) \). Lemma 2 in the form given in Eq. (10) then directly follows from applying Lemma 12.

Next, we prove that Lemma 3 follows from Lemma 2.

Proof of Lemma 3. This follows from Lemma 2 by replacing \( X \) with \( n \in \mathbb{N} \) i.i.d. copies of itself, \( X^n \). Then, associating each \( v \in V \) with \( X^n_v \), \( (G, X^n, M, \text{ind}) \) is a multiplex Bayesian network.

We now apply Algorithm 1 with \( (G, X^n, M, \text{ind}) \) as input. This is equivalent to applying it with \( (G, X, M, \text{ind}) \) \( n \) times. Then, applying Lemma 2 with inputs \( H, \{ \bigotimes_{i=1}^n \rho^{(x_i,H)}_{B_i} \}_{x^n \in X^n_H} \), \( D, \varepsilon(n) \in (0, 1) \), we obtain a POVM \( \{ Q^{(m_D,m_{\overline{D}})}_{B^n} \}_{m_D \in M_D} \) for each \( m_T \in M_T \) such that, for \( (m_D, m_{\overline{D}}) \in M_H \),
\[ \mathbb{E}_G^n \left[ \tr \left( (I - Q^{(m_D,m_{\overline{D}})}_{B^n}) \bigotimes_{i=1}^n \rho^{(x_i,H)}_{B_i} (m_D,m_{\overline{D}}) \right) \right] \]
\[ \leq f(|V_H|, \varepsilon(n)) + 4 \sum_{\emptyset \neq T \subseteq D} 2^{(\sum_{i \in T} R_i) - D_H^{\varepsilon(n)}(\rho_{X_H^n B^n} \| \rho_{X_{\overline{T}}^n B^n})} . \]

Consider now
\[ \rho_{X_H^n B^n} = \sum_{x^n_H} \rho^{(x^n_H)}_{X_H^n | x^n_H} \langle x^n_H | X^n_H \rangle \bigotimes_{i=1}^n \rho^{(x_i,H)}_{B_i} \]
and
\[ \rho^{(x^n_{\overline{T}})}_{X_{\overline{T}}^n B^n} = \sum_{x^n_H} \rho^{(x^n_H)}_{X_H^n | x^n_H} \langle x^n_H | X^n_H \rangle \bigotimes_{i=1}^n \rho^{(x^n_{\overline{T}})}_{B_i} . \]
We will use a conventional derandomization argument to eliminate it from the statement. The extra which conveniently justifies this slight abuse of notation. Furthermore, considering

This concludes the proof.

The conclusion therefore follows by Eq. (6) where we choose \( \varepsilon(n) \) such that \( \varepsilon(n) \to 0 \) so that \( f(\varepsilon(n)) \to 0 \) and \( \frac{1}{n}D_H^c(n) (\rho^\otimes n || \sigma^\otimes n) \to D(\rho || \sigma) \). Given Eq. (3), one possibility is \( \varepsilon(n) = 1/n \). This concludes the proof. \( \square \)

Finally, we prove Lemma 12. Note that Theorem 11 gives a pair \( \hat{\rho}, \hat{\Pi} \) that satisfy joint typicality properties but live in a larger Hilbert space. In order to prove Lemma 2, which claims the existence of a POVM on the original Hilbert space, we will need to construct the corresponding POVM in the larger Hilbert space and then appropriately invert the isometry. There is also an extra classical system \( Y \) associated with the \( X \) systems, which we can interpret as an additional random codebook. We will use a conventional derandomization argument to eliminate it from the statement. The extra \( Y \)'s associated with the \( B \) systems we will simply trace over.

Proof of Lemma 12. We invoke Theorem 11 with inputs the \( \rho_{XB}, \varepsilon \), and a classical system \( YZ \). Here \( X = X_1 \ldots X_k, Y = Y_1 \ldots Y_k \) and \( Z \) is a classical system associated with \( B \), to obtain a quantum state \( \hat{\rho}_{XYBZ} \) and POVM \( \hat{\Pi}_{XYBZ} \) which we can expand as follows:

\[
\hat{\rho}_{XYBZ} = \bigoplus_{x,y} p_X(x) |x \rangle_X \otimes \frac{1}{d_Y} |y \rangle_Y \otimes \hat{\rho}_{BZ}^{(x,y)} \]

\[
\hat{\Pi}_{XYBZ} = \bigoplus_{x,y} |x \rangle_X \otimes |y \rangle_Y \langle y | \otimes \hat{\Pi}_{BZ}^{(x,y)} .
\]

Now, for every \( x_j \in X_j \), draw \( y_j(x_j) \) uniformly at random from \( Y_j \), and consider the random vectors \( y(x) := (y_1(x_1), \ldots, y_k(x_k)) \). We use these random vectors and the codebook \( \mathcal{C} = \{x(i)\}_{i \in \mathcal{I}} \) to define a codebook \( \mathcal{C}' = \{y(i)\}_{i \in \mathcal{I}} \), where we set \( y(i) = y(x(i)) \). We also define the joint codebook \( \mathcal{C}'' = \{x(i)y(i)\}_{i \in \mathcal{I}} \). Then, for every \( i, i' \in \mathcal{I} \), letting \( T \equiv \Psi(i, i') \), the following holds:

1. \( x_T(i)y_T(i) = x_T(i')y_T(i') \) as random variables,

2. \( x_T(i)y_T(i) \) and \( x_T(i')y_T(i') \) are independent conditioned on \( x_T(i)y_T(i) = x_T(i')y_T(i') \),

with probabilities

\[
p_{X_TY_T|X_TY_T}(x_T, y_T | x_T, y_T) = \frac{1}{d_{Y_T}} p_{X_T}(x_T)
\]

\[
p_{X_TY_T|X_T}(x_T, y_T | x_T) = \frac{1}{d_{Y_T}} p_{X_T}(x_T),
\]

It is not difficult to see that

\[
\rho_{X_B}^{\otimes n} = \left( \sum_{x_B} p_{X_B}(x_B) |x_B \rangle_X \langle x_B | \otimes \hat{\rho}_B(x_B) \right)^{\otimes n} = \rho_{X_B}^{\otimes n},
\]

which conveniently justifies this slight abuse of notation. Furthermore, considering

\[
\rho_{B}^{\otimes n} = \sum_{x_B} p_{X_B}(x_B) |x_B \rangle_{X_B} \langle x_B | \otimes \hat{\rho}_B(x_B) = \bigotimes_{i=1}^n \rho_{B_i},
\]

we likewise conclude

\[
\rho_{X_B}^{\otimes n} = \left( \rho_{X_B}^{\otimes n} \right)^{\otimes n}.
\]

The conclusion therefore follows by Eq. (6) where we choose \( \varepsilon(n) \) such that \( \varepsilon(n) \to 0 \) so that \( f(\varepsilon(n)) \to 0 \) and \( \frac{1}{n}D_H^c(n) (\rho^\otimes n || \sigma^\otimes n) \to D(\rho || \sigma) \). Given Eq. (3), one possibility is \( \varepsilon(n) = 1/n \). This concludes the proof. \( \square \)
Define the indexed objects:
\[ \tilde{r}^{(i)}_{BZ} \equiv \tilde{r}^{(x(i), y(i))}_{BZ} \text{ and } \tilde{\Pi}^{(i)}_{BZ} \equiv \tilde{\Pi}^{(x(i), y(i))}_{BZ}. \]

We then define the square-root measurement
\[ \tilde{Q}^{(i)}_{BZ} = \left( \sum_{i_2' \in J_2} \tilde{\Pi}^{(i_1, i_2')}_{BZ} \right)^{-1/2} \tilde{\Pi}^{(i_1, i_2)}_{BZ} \left( \sum_{i_2' \in J_2} \tilde{\Pi}^{(i_1, i_2')}_{BZ} \right)^{-1/2} \]
and “invert” the isometry \( \hat{J} \) to obtain the following family of POVM’s on the original Hilbert space:
\[ Q^{(i)}_B = Q^{(i_2 | i_1)}_B = \frac{1}{d_Z} (\hat{J}_{B \to B})^\dagger \text{tr} \left[ \tilde{Q}^{(i)}_{BZ} \right] \hat{J}_{B \to B}. \]

Note that we have a POVM for each value of \( i_1 \) and these POVM’s are dependent on our random encoding \( x(i) \) and random choice of \( y(i) \).

Now, fixing \( i = (i_1, i_2) \in J \), we compute the probability of error averaged over the random choice of \( x(i) \) and \( y(i) \), denoting this by \( \mathbb{E} \equiv \mathbb{E}_{X'} \):
\[
\mathbb{E} \text{tr} \left[ (I - Q^{(i)}_B) \rho^{(i)}_B \right]
\leq 1 - \mathbb{E} \text{tr} \left[ Q^{(i)}_B \rho^{(i)}_B \right]
\leq 1 - \mathbb{E} \text{tr} \left[ \hat{Q}^{(i)}_{BZ} \left( \hat{J}_{B \to B} \langle B' \rangle \hat{J}^\dagger_{B \to B} \otimes \tau_Z \right) \right]
\leq 1 - \mathbb{E} \text{tr} \left[ \hat{Q}^{(i)}_{BZ} \rho^{(i)}_{BZ} \right] + \mathbb{E} \left\| \hat{J}_{B \to B} \langle B' \rangle \hat{J}^\dagger_{B \to B} \otimes \tau_Z - \hat{\rho}^{(i)}_{BZ} \right\|_1
\leq 1 - \mathbb{E} \text{tr} \left[ \hat{Q}^{(i)}_{BZ} \rho^{(i)}_{BZ} \right] + f(1, k, \varepsilon)
\leq 2 \left( 1 - \mathbb{E} \text{tr} \left[ \tilde{\Pi}^{(i)}_{BZ} \hat{\rho}^{(i)}_{BZ} \right] \right) + 4 \sum_{i_2' \neq i_2} \mathbb{E} \text{tr} \left[ \tilde{\Pi}^{(i_1, i_2')}_{BZ} \hat{\rho}^{(i_1, i_2)}_{BZ} \right] + f(1, k, \varepsilon)
\leq 4 \sum_{i_2' \neq i_2} \mathbb{E} \text{tr} \left[ \tilde{\Pi}^{(i_1, i_2')}_{BZ} \hat{\rho}^{(i_1, i_2)}_{BZ} \right] + f(1, k, \varepsilon) + 2g(1, k, \varepsilon),
\]
where in the last three inequalities we used Theorem 11 and the Hayashi-Nagaoka lemma [31, 34].

We consider the first term. Let \( S = \Psi((i_1, i_2), (i_1, i_2')) \). Note that by our conditions on the random codebook, the codewords are equal as random variables on \( S \) and hence,
\[
4 \sum_{i_2' \neq i_2} \mathbb{E} \text{tr} \left[ \tilde{\Pi}^{(i_1, i_2')}_{BZ} \hat{\rho}^{(i_1, i_2)}_{BZ} \right]
= 4 \sum_{i_2' \neq i_2} \mathbb{E}_{X'Y'Y'} \text{tr} \left[ \tilde{\Pi}^{(i_1, i_2')}_{BZ} \hat{\rho}^{(i_1, i_2)}_{BZ} \right]
= 4 \sum_{i_2' \neq i_2} \text{tr} \left[ \mathbb{E}_{X'Y'} \left[ \mathbb{E}_{X''Y''} \left[ \tilde{\Pi}^{(i_1, i_2')}_{BZ} \mathbb{E}_{X''Y''} \left( \tilde{\Pi}^{(i_1, i_2)}_{BZ} \right) \right] \right] \right]
= 4 \sum_{i_2' \neq i_2} \text{tr} \left[ \sum_{x \in X} p(x) \frac{1}{d_{Y''}} \sum_{x' : s | x} p(x' | s) \sum_{y : y' | y} p(y' | y) \tilde{\Pi}^{(x', x', y' | y)}_{BZ} \sum_{x : s | x} p(x) \frac{1}{d_{Y''}} \hat{\rho}^{(x, s, y, y') \dagger}_{BZ} \right].
\]
In the first two equalities we use the notation $X \equiv x(i_1, i_2)$, $X' \equiv x(i_1, i'_2)$ and similarly for $Y, Y'$. In the fourth equality $\hat{\rho}_{XBYZ}^{(x,y)}$ is the marginal of the conditional density operator $\hat{\rho}_{XBYSZ}^{(x,y)}$. In the last inequality we use Theorem 11 and choose the dimensions of $Y, Z$ to be sufficiently large so that $h(1, k, d_B, d_Y d_Z) \leq \varepsilon$.

Finally, we can invoke the usual derandomization argument to remove the dependency of our POVM on the choice of $y(i)$. That is, we know that

$$\mathbb{E} \text{tr} \left[ (I - Q_B^{(i)}) p_B^{(i)} \right] = \mathbb{E}_C \mathbb{E}_C \text{tr} \left[ (I - Q_B^{(i)}) p_B^{(i)} \right] \leq \varepsilon + f(1, k, \varepsilon) + 2g(1, k, \varepsilon) + 4 \sum_{i'_2 \neq i_2} 2^{-D_H(\rho_{XB}^{(i)} \| \rho_{XB}^{(1)})}.$$

Hence, there is a particular choice of $y(i)$ such that the corresponding POVM $Q_B^{(i_2|i_1)}$ satisfies the bound in Lemma 12, with

$$f(k, \varepsilon) = \varepsilon + f(1, k, \varepsilon) + 2g(1, k, \varepsilon).$$

\[ \square \]

**VI. CONCLUSIONS**

The packing lemma is a cornerstone of classical network information theory, used as a black box in the analyses of all kinds of network communication protocols. At its core, the packing lemma follows from properties of the set of jointly typical sequences for multiple random variables. In this letter, we provide an analogous statement in the quantum setting that we believe can serve a similar purpose for quantum network information theory. We illustrate this by using it as a black box to prove achievability results for the classical-quantum relay channel. Our result is based on a joint typicality lemma recently proved by Sen [4]. This result, at a high level, provides a single POVM which achieves the hypothesis testing bound for all possible divisions of a multiparty state into a tensor product of its marginals. This result allows for the construction of finite blocklength protocols for quantum multiple access, relay, broadcast, and interference channels [32].

Two alternative formulations of joint typicality were proposed in [7] and [22]. In the first work, the author conjectured the existence of the jointly typical state that is close to an i.i.d. multiparty state but with marginals whose purities satisfy certain bounds. This notion of typicality was then used in the analysis of multiparty state merging and assisted entanglement distillation protocols. In the second work, the authors provided a similar statement for the one-shot case. Specifically, for a given multiparty state, they conjectured the existence of a state that is close to the initial state
but has a min-entropy bounded by the smoothed min-entropy of the initial state for all marginals. In a follow up paper we will try to understand the relationship between these various notions of quantum joint typicality and whether Sen’s results can be extended to prove the other notions or to realize the applications they are designed for.

Also, as noted in the corresponding section, our protocol for the partial decode-forward bound is not a straightforward generalization of the classical protocol in [1]. Our algorithm involves a joint measurement of all the transmitted blocks instead of performing a backward decoding followed by a forward decoding as in the classical case. The problem arises from the fact that the classical protocol makes multiple measurements on a single system but also intermediate measurements on other systems. Hence, a direct application of our packing lemma has to combine these different measurements and the intermediate ones into one joint measurement. This results in a set of inequalities for the rate region that has to be simplified to obtain the desired bound. This is a step that might be necessary in other applications of our packing lemma.

There are still several interesting questions that remain open regarding quantum relay channels. The most obvious one is proving converses for the given achievability lower bounds. There are known converses for special classical relay channels, and it would be interesting to extend them to the quantum case as we did for semideterministic relay channels. Another, albeit less trivial, direction is to prove a quantum equivalent of the compress-forward lower bound [1]. We might need to analyze this in the entanglement assisted case since it is only then that a single-letter quantum rate-distortion theorem is known [35]. Another idea is to study networks of relay channels, where the relays are operating in series or in parallel. Some preliminary work was done in [25], and the most general notion of this in the classical literature is a multicast network [1]. Lastly, relay channels with feedback would also be interesting to investigate.

ACKNOWLEDGMENTS

We thank Pranab Sen for interesting discussions and for sharing his draft [4] with us. We would also like to thank Mario Berta, Philippe Faist, and Mark Wilde for inspiring discussions. PH was supported by AFOSR (FA9550-16-1-0082), CIFAR and the Simons Foundation. HG was supported in part by NSF grant PHY-1720397. MW acknowledges financial support by the NWO through Veni grant no. 680-47-459. DD is supported by the Stanford Graduate Fellowship and the National Defense Science and Engineering Graduate Fellowship. DD would like to thank God for all of His provisions.

[1] Abbas El Gamal and Young-Han Kim. Network Information Theory. Cambridge University Press, 2011.
[2] Alexander S Holevo. The capacity of the quantum channel with general signal states. IEEE Transactions on Information Theory, 44(1):269–273, 1998. doi:10.1109/18.651037.
[3] Benjamin Schumacher and Michael D Westmoreland. Sending classical information via noisy quantum channels. Physical Review A, 56(1):131, 1997. doi:10.1103/PhysRevA.56.131.
[4] Pranab Sen. A one-shot quantum joint typicality lemma. 2018. arXiv:1806.07278.
[5] Omar Fawzi and Ivan Savov. Rate-splitting in the presence of multiple receivers. CoRR, abs/1207.0543, 2012. arXiv:1207.0543.
[6] Omar Fawzi, Patrick Hayden, Ivan Savov, Pranab Sen, and Mark M Wilde. Classical communication over a quantum interference channel. IEEE Transactions on Information Theory, 58(6):3670–3691, 2012. doi:10.1109/TIT.2012.2188620.
[7] Nicolas Dutil. Multiparty quantum protocols for assisted entanglement distillation. arXiv preprint arXiv:1105.4657, 2011. arXiv:1105.4657.
Appendix A: Proof of Cutset Bound

We give a proof of Proposition 4, essentially identical to that of [1].

Proof. Consider an $(n, 2^{nR})$ code for $\mathcal{N}_{X_1X_2\rightarrow B_2B_3}$. Suppose we have a uniform distribution over the message set $M$, and denote the final classical system obtained by Bob from the POVM measurement by $\hat{M}$. By the classical Fano’s inequality,

$$nR = H(M) = I(M; \hat{M}) + H(M|\hat{M}) \leq I(M; \hat{M}) + n\delta(n),$$

where $\delta(n)$ satisfies $\lim_{n \to \infty} \delta(n) = 0$ if the decoding error is to vanish in asymptotic limit.

We denote by $(X_1)_j, (X_2)_j, (B_2)_j, (B_3)_j$ the respective classical and quantum systems induced by our protocol. We argue

$$I(M; \hat{M}) \leq I(M; B_3^n) = \sum_{j=1}^{n} I(M; (B_3)_j|B_3^{j-1})$$

$$\leq \sum_{j=1}^{n} I(MB_3^{j-1}; (B_3)_j)$$

$$\leq \sum_{j=1}^{n} I((X_1)_j(X_2)_jMB_3^{j-1}; (B_3)_j)$$

$$= \sum_{j=1}^{n} I((X_1)_j(X_2)_j; (B_3)_j).$$

The last step follows from the i.i.d. nature of the $n$ channel uses and the channel is classical-quantum. More explicitly, we can write out the overall state as the protocol progresses, and since the input to the channel on each round is classical, it is not difficult to see that given $(X_1)_j(X_2)_j, (B_2)_j(B_3)_j$ is in tensor product with the other systems. This would not hold if the channel takes quantum inputs, for which we would expect an upper bound that involves regularization. Now, similarly,

$$I(M; \hat{M}) \leq I(M; B_3^n) \leq I(M; B_2^nB_3^n)$$

$$= \sum_{j=1}^{n} I(M; (B_2)_j(B_3)_j|B_2^{j-1}B_3^{j-1})$$

$$= \sum_{j=1}^{n} I(M; (B_2)_j(B_3)_j|B_2^{j-1}B_3^{j-1}(X_2)_j).$$
\[
\leq \sum_{j=1}^{n} I(MB_j^{j-1}B_3^{j-1}; (B_2)_j(B_3)_j|X_2)_j) \\
\leq \sum_{j=1}^{n} I((X_1)_j MB_j^{j-1}B_3^{j-1}; (B_2)_j(B_3)_j|X_2)_j) \\
= \sum_{j=1}^{n} I((X_1)_j; (B_2)_j(B_3)_j|X_2)_j),
\]
where the second equality follows since given \(B_2^{j-1}\), one can obtain \((X_2)_j\) by a series of \(R\) operations (Note that \((B_2)_0(B_3)_0\) is a trivial system and thus independent of the code.).

Define the state
\[
\sigma_{X_1X_2B_2B_3} \equiv \frac{1}{n} \sum_{q=1}^{n} |q\rangle \langle q| \otimes \sigma^{(q)}_{X_1X_2B_2B_3},
\]
where \(\sigma^{(q)}\) is the classical-quantum state on the \(q\)-th round of the protocol, that is, the state on the system \((X_1)_q(X_2)_q(B_2)_q(B_3)_q\). Now, \(I(B_2B_3; Q|X_1X_2)_\sigma = 0\), so
\[
\sum_{j=1}^{n} I((X_1)_j(X_2)_j; (B_3)_j) = bI(X_1X_2; B_3|Q)_\sigma \\
\leq nI(X_1X_2Q; B_3)_\sigma \\
= nI(X_1X_2; B_3)_\sigma
\]
and similarly
\[
\sum_{j=1}^{n} I((X_1)_j; (B_2)_j(B_3)_j|X_2)_j) = nI(X_1; B_2B_3|X_2Q)_\sigma \\
\leq nI(X_1; B_2B_3|X_2)_\sigma \\
= nI(X_1; B_2B_3|X_2)_\sigma.
\]
Hence,
\[
R \leq \min\{I(X_1X_2; B_3)_\sigma, I(X_1; B_2B_3|X_2)_\sigma\} + \delta(n).
\]
Now, \(\sigma_{X_1X_2B_2B_3}\) is simply a uniform average of all the classical-quantum states from each round of the protocol, it is also a possible classical-quantum state induced by \(\mathcal{N}_{X_1X_2B_2B_3}\) acting on some classical input distribution \(p_{X_1X_2}\). In particular, \(R\) is therefore upper bounded by the input distribution which maximizes the quantity on the right-hand side:
\[
R \leq \max_{p_{X_1X_2}} \min\{I(X_1X_2; B_3), I(X_1; B_2B_3|X_2)\} + \delta(n).
\]
Taking the \(n \to \infty\) limit completes the proof. \(\square\)