UC Irvine
Recent Work

Title
Non-signaling Assistance in Lossy Source Coding with Side Information: Virtual Signaling of Decoder Side Information

Permalink
https://escholarship.org/uc/item/4kj3500n

Authors
Yao, Yuhang
Jafar, Syed A

Publication Date
2026-11-11

Copyright Information
This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at https://creativecommons.org/licenses/by-nc-nd/4.0/

eScholarship.org Powered by the California Digital Library
University of California

Non-signaling Assistance in Lossy Source Coding with Side Information: Virtual Signaling of Decoder Side Information

Yuhang Yao, Syed A. Jafar
Nhu Department of Electrical Engineering and Computer Science
University of California Irvine, Irvine, CA 92697
Email: {yuhangy5, syed}@uci.edu

Abstract

We explore how non-signaling correlations, including quantum correlations, improve the fundamental limit of classical lossy source coding. For the problem of lossy source coding with decoder side information, also known as the Wyner-Ziv problem, we show that non-signaling assistance between the encoder and the decoder yields the same rate–distortion trade-off as the classical setting when the side information is also given to the encoder. Thus, non-signaling assistance virtually signals the decoder side information to the encoder. This result can be viewed as a source-coding counterpart of the 'virtual signaling of channel state information at the transmitter (CSIT)' previously identified in channel coding with transmitter side information. We also construct source and side information distributions for which entanglement assistance asymptotically recovers the entire rate advantage brought by non-signaling assistance at the low-rate regime.

1 Introduction

Entangled quantum systems enable two spatially separate terminals to generate correlations that cannot be reproduced by classical shared randomness alone without additional communication. Remarkably, these correlations cannot by themselves be used to transmit information. More generally, the class of *non-signaling* correlations encompasses all correlations that do not by themselves permit communication between the terminals. Despite their inability to transmit information on their own, such correlations have been shown to enhance communication efficiency when used to assist classical communication resources. Examples include increasing the number of messages that can be transmitted with zero error over a classical channel, simulating a prescribed channel using fewer communication bits [1], and even increasing the capacity of some of the most elementary multi-user communication channels [2].

These findings raise a fundamental question in information theory: What are the ultimate limits of communication when non-signaling correlations are available as an assisting resource? At the heart of information theory lies the characterization of optimal coding rates. Two of its most fundamental classes of problems are channel coding and source coding, which share an intriguing dual relationship [3, Sec. 11.3.3]. Both involve a transmitter (encoder) and a receiver (decoder), but the roles of the communication resource and the communication objective are essentially reversed.

In channel coding with transmitter side information, the transmitter seeks to communicate a message reliably over a noisy channel, aided by knowledge of the state governing the channel

statistics. The coding rate is measured in message bits per channel use, with higher rates corresponding to greater communication efficiency. The fundamental objective is to characterize the channel capacity, defined as the supremum of achievable rates. In contrast, in lossy source coding with decoder side information, the encoder uses a noiseless communication link to enable the decoder to reconstruct a source within a prescribed distortion level, aided by correlated side information available only at the decoder. Here, the coding rate is measured in communication bits per source symbol, and lower rates indicate greater efficiency. The objective is to characterize the rate–distortion trade-off, namely, the minimum achievable rate for each prescribed distortion level. Thus, while channel coding converts a noisy communication resource into reliably transmitted bits, lossy source coding uses reliable communication bits to convey an approximate reconstruction of a source. Since neither formulation imposes computational complexity constraints on the encoder or decoder, their optimal coding rates are traditionally regarded as fundamental limits of communication.

The classical solutions to both problems have been known for decades [4,5]. However, the possibility of non-signaling assistance prompts us to revisit a fundamental question: Do these classical characterizations remain the ultimate limits when arbitrary non-communicating resources are available for free? Although the classical results impose no computational constraints on the encoder and decoder and remain valid in the presence of shared randomness, they do not account for more general non-signaling resources shared between the communicating terminals. Quantum entanglement is one such resource, and more general non-signaling correlations offer an even broader framework for investigating the limits of assisted communication.

For channel coding with non-causal channel state information at the transmitter (CSIT), [6] establishes the following result: the non-signaling-assisted capacity equals the classical capacity when the receiver also knows the state. This phenomenon is termed "virtual signaling of CSIT." Although non-signaling assistance cannot physically convey the transmitter's state information to the receiver, it can nevertheless eliminate the entire capacity penalty associated with the receiver's lack of state information. It naturally raises a corresponding question on the source coding side of the duality: What analogous capabilities can non-signaling assistance provide in lossy source coding with decoder side information?

1.1 Contribution

We therefore turn to lossy source coding with decoder side information, known as the Wyner–Ziv problem [5]. More specifically, an encoder observes a source X, a decoder observes side information Y, and the encoder sends a message that enables reconstruction of X within a prescribed average distortion. A code may operate on a sequence of n i.i.d. copies of (X, Y), denoted as (X^n, Y^n). A classical encoder maps X^n to a message M, and a decoder maps (M, Y^n) to a reconstruction Z^n, so that the expected value of the amortized distortion between X^n and Z^n approaches certain prescribed threshold. One may also consider general distortion measures that also depend on the side information Y^n. In the classical setting, it is known that giving Y to the encoder can strictly reduce the required communication rate [5, 7]. In this paper, we show that non-signaling assistance serves exactly the same purpose: the non-signaling-assisted rate–distortion region equals the classical region with side information Y at both terminals. Thus, non-signaling assistance enables *virtual signaling of decoder side information*. This provides a source-coding counterpart to the "*virtual signaling of CSIT*" identified in [6], and another operational manifestation of the source–channel duality with side information.

This equivalence also raises a question about quantum entanglement assistance: how much of this effect can quantum entanglement recover? We provide a simple class of sources defined by graphs, and we show that in the low-rate limit, entanglement assistance asymptotically recovers the entire advantage brought by non-signaling assistance, while the unassisted optimum is strictly separated away. Remarkably, the class is also simple to describe. The decoder side information Y corresponds to a uniformly chosen edge of the graph, and conditioned on each edge Y, the source X observed by the encoder is uniformly distributed over the two endpoints of Y. Under Hamming distortion, the decoder must determine which endpoint was selected. Thus, a random guess of the two endpoints by the decoder would yield a distortion $1/2$. The definition of this class is conceptually simple, without a reduction from a nonlocal game [8–11]. The class is motivated by the examples constructed in our prior work on channel coding with causal CSIT [12], although here the decoder side information is non-causal. Notably, for the analysis of the entanglement-assisted rate–distortion trade-off, we need Tsirelson's representation of quantum correlations [13].

For every graph in this class, as the distortion $D \uparrow 1/2$, all rates approach 0, so we look at the ratio between the rates. The entanglement-assisted and non-signaling-assisted rates have ratio tending to one. For complete graphs $\mathcal{K}_m$, $m \geq 3$, their ratios to the unassisted classical rate remain bounded away from one: the limiting upper bound on the assisted-to-unassisted rate ratio is $2(m+1)/(3m)$, which approaches $2/3$ as m grows. Thus, the guaranteed multiplicative rate advantage approaches a factor of $3/2$.

1.2 Related works

Beyond the dual relationship with non-signaling-assisted channel coding with transmitter side information, the following problems share common features with our setting.

Non-signaling assistance in channel coding. Non-signaling assistance has been studied in various channel coding problems. The results can be roughly divided into two classes: a negative class and a positive class. In the negative class, it is known that for some channels under certain assumptions on non-signaling assistance, the channel capacity cannot be improved at all by non-signaling assistance. For example, Bennett et al. [14] showed that entanglement assistance cannot increase the capacity of a point-to-point channel, and Matthews [15] extends the result to non-signaling assistance. Pereg et al. [16] showed that entanglement assistance between the receivers cannot increase the broadcast channel capacity region, and Fawzi et al. [17] extended the result to non-signaling assistance. Even full-partite non-signaling assistance cannot improve the capacity region of a deterministic broadcast channel [17].

In the positive class, it is known that entanglement assistance between the transmitters can improve the sum-rate capacity of some interference channels [8, 18] as well as multiple-access channels [9, 19, 20]. Further, general all-partite non-signaling assistance can improve the sum-rate capacity of certain K-receiver broadcast channels [21], and certain K-transmitter multiple-access channels [22], with extremal value of the multiplicative advantage shown to be equal to K in both settings. Entanglement assistance can also provide unbounded multiplicative sum-rate capacity gain in certain multiple-access channels with causal state information at the transmitters [10].

Remote state preparation with side information. Beigi and Gohari [23] show that a transmitter who knows the classical description X of a quantum state ρ_Q^x can remotely prepare it using shared entanglement and classical communication at rate $I(X;Q \mid Y)$, when the decoder has

side information Y. This is a generalization of an earlier result [24] without side information. The connection to our present work is noteworthy. Indeed, we obtain a general entanglement-assisted rate–distortion characterization[1] (Proposition 2) by applying the remote state preparation result [23] and a standard information-theoretic converse.

Quantum rate–distortion theories. There are theories generalizing the classical rate–distortion theory to a quantum one. A comprehensive formulation is [25], incorporating quantum source, quantum side information, and distortion measures defined on quantum states. These codes use quantum messages and entanglement assistance. An alternative characterization of the entanglement-assisted rate–distortion region can then be obtained as a special case by restricting the quantum sources and side information to classical registers and choosing a distortion operator that corresponds to the classical distortion function. Although the alternative characterization measures the communication cost in qubits, the conversion to a cost of classical bits can be simply done by doubling the quantum communication cost. This is because with superdense coding and quantum teleportation, one qubit equals two classical bits in communication (and vise versa).

Communication complexity. Entanglement and non-signaling assistance have also been extensively studied in communication complexity [26]. In its conventional formulation, communication complexity concerns the minimum communication required to compute a function of distributed inputs, typically for a single instance under a zero-error or bounded-error criterion. Although repeated instances are also studied through direct-sum and direct-product theorems, these generally focus on the communication required to solve multiple instances with a prescribed *overall* success probability. Related problems, such as exact channel simulation, instead require reproducing the input–output statistics of a prescribed channel. These operational criteria differ from the asymptotic rate–distortion framework considered here, which allows joint coding over i.i.d. source blocks subject to an *average* per-symbol distortion constraint.

Information causality. Information causality [27] considers a setting where an encoder observes K independent sources $(X_1, X_2, \ldots, X_K)$, communicates a message M to a decoder, allowing the decoder to reconstruct X_b, for every $b \in [K]$, where b is known only to the decoder. A connection to the framework of source coding with side information can be made by modeling the decoder index b as a random side information symbol B, assigning a joint distribution for $(X_1, \ldots, X_K)$ and B, and studying the rate–distortion region.

Prepare-and-measure. Another closely related setting is prepare-and-measure, which considers classical sources and side information with an unassisted quantum message. In prepare-and-measure scenarios, the transmitter's input selects a state to transmit, and the receiver's input selects a measurement. In the one-shot setting, there is no general ordering between an entanglement-assisted classical dit (a d-ary message) and an unassisted qudit (a d-dimensional quantum message). Both directions of advantage already occur in point-to-point scenarios. Tavakoli, Pauwels, Woodhead, and Pironio show that an entanglement-assisted bit can generate correlations unattainable with an unassisted qubit [28, Sec. VI.A], whereas an unassisted qutrit can outperform a classical trit assisted by unrestricted entanglement [28, Sec. VI.B]. The binary case admits a stronger statement: Pauwels et al. [29] show that, with adaptive receiver measurements,

[1] The dimension of the auxiliary quantum system involved in the general characterization is, however, unbounded.

an entanglement-assisted bit can simulate all unassisted qubit prepare-and-measure correlations and is strictly more powerful. This qubit-specific ordering therefore does not extend to arbitrary message dimensions. More recently, Nogueira et al. [30] show that non-signaling-assisted classical dit communication can simulate unassisted qudit communication of the same dimension. These comparisons concern one-shot correlations with a fixed message alphabet or Hilbert-space dimension.

In our asymptotic rate–distortion setting, however, a general ordering does hold: unassisted quantum communication is no stronger than entanglement-assisted classical communication at the same numerical rate, measured in qubits and bits per source symbol, respectively. To state this precisely, let $\mathcal{R}$ denote an achievable rate–distortion region. The superscripts $\mathrm{c\text{-}m}$ and $\mathrm{q\text{-}m}$ indicate classical and quantum messages, respectively. Superscripts $\mathrm{un\text{-}a}$, $\mathrm{ent\text{-}a}$, and $\mathrm{ns\text{-}a}$ indicate no assistance (unassisted), entanglement assistance, and non-signaling assistance, respectively. Proposition 3 gives

$$\mathcal{R}^{\mathrm{c\text{-}m,un\text{-}a}} \subseteq \mathcal{R}^{\mathrm{q\text{-}m,un\text{-}a}} \subseteq \mathcal{R}^{\mathrm{c\text{-}m,ent\text{-}a}} \subseteq \mathcal{R}^{\mathrm{c\text{-}m,ns\text{-}a}}. \tag{1}$$

The inclusion $\mathcal{R}^{\mathrm{q\text{-}m,un\text{-}a}} \subseteq \mathcal{R}^{\mathrm{c\text{-}m,ent\text{-}a}}$ follows by applying remote state preparation to repeated uses of a quantum-message code, with vanishing additional distortion. The change from one-shot simulation to asymptotic rate–distortion is important: coding over many i.i.d. source symbols allows entanglement-assisted classical communication to attain every rate–distortion pair achievable by unassisted quantum communication, even though same-dimension simulation can fail for a single use.

2 Preliminaries

2.1 Notation

Let $\mathbb{N}$ denote the set of positive integers and $\mathbb{R}_+$ the set of non-negative real numbers. For $n \in \mathbb{N}$, write $[n] = \{1, \ldots, n\}$. For a sequence $x_1, \ldots, x_n$, we write $x^n = (x_1, \ldots, x_n)$. The Cartesian product of two sets $\mathcal{A}$ and $\mathcal{B}$ is denoted by $\mathcal{A} \times \mathcal{B}$, and $|\mathcal{A}|$ denotes the cardinality of a finite set $\mathcal{A}$.

For finite sets $\mathcal{A}$ and $\mathcal{B}$, let $\mathcal{P}(\mathcal{A})$ denote the set of probability distributions on $\mathcal{A}$ and $\mathcal{P}(\mathcal{B} \mid \mathcal{A})$ the corresponding set of conditional probability distributions. For a distribution $\mathsf{P}_X \in \mathcal{P}(\mathcal{X})$, we write $X \sim \mathsf{P}_X$ to indicate that the random variable X has distribution P_X. We write $\mathsf{P}_X^{\otimes n}(x^n) = \prod_{i=1}^n \mathsf{P}_X(x_i)$ for the n-fold product distribution. For random variables $(X, Y) \sim \mathsf{P}_{XY}$, we write P_X and P_Y to denote the marginal distributions of X and Y, respectively, and $\mathsf{P}_{Y|X}$ denotes the conditional distribution of Y given X. We use $\Pr(E)$ for the probability of an event E and $\mathbb{E}[X]$ for the expectation of a random variable X. The indicator of a predicate E is denoted by $\mathbb{I}[E]$.

For a finite-dimensional Hilbert space $\mathcal{H}$, let $\mathcal{L}(\mathcal{H})$ denote the space of linear operators on $\mathcal{H}$, and $\mathcal{D}(\mathcal{H})$ the set of density operators on $\mathcal{H}$. For quantum systems A and B with associated Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, the composite system AB is associated with $\mathcal{H}_A \otimes \mathcal{H}_B$. We use I_A for the identity operator on $\mathcal{H}_A$, $\mathrm{Tr}[\cdot]$ for the trace, and $\mathrm{Tr}_A[\cdot]$ for the partial trace over system A. The trace norm of an operator A is denoted by $\|A\|_1$.

A quantum channel $\mathcal{N}\colon \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ is a completely positive and trace-preserving (CPTP) map. The identity channel on a quantum system A is denoted by id_A. A classical-quantum channel from a finite alphabet $\mathcal{X}$ to a quantum system Q associates with every $x \in \mathcal{X}$ a density operator $\rho_Q^x \in \mathcal{D}(\mathcal{H}_Q)$. A positive operator-valued measure (POVM) with outcome alphabet $\mathcal{Z}$ on a quantum system A is a collection $\{\Lambda_z\colon z \in \mathcal{Z}\}$ of positive semidefinite operators on $\mathcal{H}_A$ satisfying

$\sum_{z\in\mathcal{Z}} \Lambda_z = I_A$. When a POVM $\{\Lambda_z\}$ is applied to a state $\rho_A \in \mathcal{D}(\mathcal{H}_A)$, the probability of observing outcome z is $\mathrm{Tr}[\Lambda_z \rho_A]$.

2.2 Bipartite non-signaling correlations

We will use bipartite non-signaling correlations (also referred to as non-signaling boxes) to provide a unified description of codes with different kinds of assistance. In the following definitions, $\mathcal{A}_1, \mathcal{A}_2, \mathcal{B}_1, \mathcal{B}_2$ are finite sets.

Definition 1 (Non-signaling box)**.** *A non-signaling box is specified by a conditional distribution $\mathsf{P}_{B_1B_2|A_1A_2} \in \mathcal{P}(\mathcal{B}_1 \times \mathcal{B}_2 \mid \mathcal{A}_1 \times \mathcal{A}_2)$, such that the marginal distribution $\mathsf{P}_{B_1|A_1A_2}(b_1 \mid a_1, a_2)$ is independent of a_2 for every (a_1, b_1), and the marginal distribution $\mathsf{P}_{B_2|A_1A_2}(b_2 \mid a_1, a_2)$ is independent of a_1 for every (a_2, b_2). The set of such correlations is denoted by $\mathcal{P}^{\mathrm{NS}}(\mathcal{B}_1; \mathcal{B}_2 \mid \mathcal{A}_1; \mathcal{A}_2)$.*

Definition 2 (Entanglement-assisted correlation)**.** *A correlation $\mathsf{P}_{B_1B_2|A_1A_2} \in \mathcal{P}(\mathcal{B}_1 \times \mathcal{B}_2 \mid \mathcal{A}_1 \times \mathcal{A}_2)$ is called an entanglement-assisted correlation if there exist finite-dimensional quantum systems Q_1 and Q_2, a state (density operator) $\rho \in \mathcal{D}(\mathcal{H}_{Q_1} \otimes \mathcal{H}_{Q_2})$, and POVMs $\{E_{b_1|a_1}\}$ and $\{D_{b_2|a_2}\}$ such that $\mathsf{P}_{B_1B_2|A_1A_2}(b_1, b_2 \mid a_1, a_2) = \mathrm{Tr}[(E_{b_1|a_1} \otimes D_{b_2|a_2})\rho]$. The set of such correlations is denoted by $\mathcal{P}^{\mathrm{Ent}}(\mathcal{B}_1; \mathcal{B}_2 \mid \mathcal{A}_1; \mathcal{A}_2)$.*

Quantum entanglement alone cannot be used for signaling. This is known as the no-communication theorem (e.g., [31]). Thus, every entanglement-assisted correlation is non-signaling, i.e., $\mathcal{P}^{\mathrm{Ent}}(\mathcal{B}_1; \mathcal{B}_2 \mid \mathcal{A}_1; \mathcal{A}_2) \subseteq \mathcal{P}^{\mathrm{NS}}(\mathcal{B}_1; \mathcal{B}_2 \mid \mathcal{A}_1; \mathcal{A}_2)$.

3 Problem formulation

3.1 Source coding with side-information-dependent distortion measures

Our main focus is the point-to-point lossy source coding problem with side information available at the decoder, first introduced by Wyner and Ziv in [5]. We consider a more general model with side-information-dependent distortion measures, as in [32].

The lossy source coding model with side-information-dependent distortion is specified by finite alphabets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$, a joint distribution $\mathsf{P}_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, and a distortion function

$$d\colon \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \to \mathbb{R}_+. \tag{2}$$

Specifically, $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ specify the alphabets of the source to be compressed, the decoder side information, and the reconstruction, respectively. The distortion function d maps any triple $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ to a non-negative real number.

A blocklength-n code operates on n i.i.d. source pairs $(X^n, Y^n) \sim \mathsf{P}_{XY}^{\otimes n}$. The encoder observes X^n but does not observe Y^n. Based on X^n, the encoder generates a message M that is transmitted noiselessly to the decoder. The message M may be either classical or quantum, depending on the class of codes allowed. The decoder observes Y^n together with M and produces a reconstruction $Z^n \in \mathcal{Z}^n$. For sequences (X^n, Y^n, Z^n) induced by a code, the average distortion (per source symbol) is defined as $\frac{1}{n}\sum_{i=1}^n \mathbb{E}[d(X_i, Y_i, Z_i)]$.

Our primary focus is the setting where the message transmitted from the transmitter to the receiver is classical, henceforth labeled as $\mathrm{c\text{-}m}$ (classical message). Besides the classical message,

we consider three assistance models: 1) the unassisted case (labeled as $\mathrm{un\text{-}a}$); 2) the entanglement-assisted case (labeled as $\mathrm{ent\text{-}a}$); and 3) the non-signaling-assisted case (labeled as $\mathrm{ns\text{-}a}$). For comparison, we also consider an unassisted setting in which the transmitted message itself is quantum, labeled as $\mathrm{q\text{-}m}$. The formal definitions of these codes are given next.

3.2 Classical-message codes with various forms of non-signaling assistance

A blocklength-n classical-message code assisted by general non-signaling assistance consists of a finite set $\mathcal{M}$ and a bipartite non-signaling box (Definition 1)

$$\mathsf{B}_{MZ^n|X^n\widetilde{M}Y^n} \in \mathcal{P}^{\mathrm{NS}}(\mathcal{M}; \mathcal{Z}^n \mid \mathcal{X}^n; \mathcal{M} \times \mathcal{Y}^n). \tag{3}$$

The two parties of the box are referred to as the encoder and the decoder. At the encoder, X^n denotes the input (source observations), and M denotes the output (a classical message). At the decoder, $\widetilde{M}$ and Y^n denote the inputs (the received classical message and side-information observations, respectively), while Z^n denotes the output (source reconstructions).

The code works as follows. Conditioned on $X^n = x^n$, the encoder generates the classical message M according to the conditional distribution $\mathsf{B}_{M|X^n\widetilde{M}Y^n}(m \mid x^n, (\widetilde{m}, y^n)) \triangleq \mathsf{B}_{M|X^n}(m \mid x^n)$. This is well defined because the non-signaling property ensures that the conditional distribution of M does not depend on $(\widetilde{m}, y^n)$. The message M is then communicated to the decoder. Conditioned on $(X^n = x^n, Y^n = y^n, M = m)$, the decoder produces Z^n according to the conditional distribution $\mathsf{B}_{Z^n|MX^n\widetilde{M}Y^n}(z^n \mid m, x^n, m, y^n)$. Consequently, the joint distribution of (X^n, Y^n, M, Z^n) is given by

$$\begin{aligned}
&\mathsf{P}_{X^nY^nMZ^n}(x^n, y^n, m, z^n) \\
&= \Big(\prod_{i=1}^n \mathsf{P}_{XY}(x_i, y_i)\Big)\mathsf{B}_{M|X^n}(m \mid x^n)\mathsf{B}_{Z^n|MX^n\widetilde{M}Y^n}(z^n \mid m, x^n, m, y^n) &(4)\\
&= \Big(\prod_{i=1}^n \mathsf{P}_{XY}(x_i, y_i)\Big)\mathsf{B}_{MZ^n|X^n\widetilde{M}Y^n}(m, z^n \mid x^n, m, y^n) &(5)
\end{aligned}$$

Moreover, if $\mathsf{B}_{MZ^n|X^n\widetilde{M}Y^n} \in \mathcal{P}^{\mathrm{Ent}}(\mathcal{M}; \mathcal{Z}^n \mid \mathcal{X}^n; \mathcal{M} \times \mathcal{Y}^n)$, the code is also called an entanglement-assisted code. If $\mathsf{B}_{MZ^n|X^n\widetilde{M}Y^n}(m, z^n \mid x^n, \widetilde{m}, y^n)$ can be factorized as $\mathsf{E}(m \mid x^n)\mathsf{D}(z^n \mid \widetilde{m}, y^n)$ for some $\mathsf{E} \in \mathcal{P}(\mathcal{M} \mid \mathcal{X}^n)$ and $\mathsf{D} \in \mathcal{P}(\mathcal{Z}^n \mid \mathcal{M} \times \mathcal{Y}^n)$, the code is also called an unassisted code.

3.3 Quantum-message codes

We also consider codes that transmit a quantum message. Let Q be a finite-dimensional quantum system. A blocklength-n (unassisted) quantum-message code consists of a classical-quantum channel $\{\rho_Q^{x^n} : x^n \in \mathcal{X}^n\}$ and a family of POVMs $\{\{D_{z^n|y^n} : z^n \in \mathcal{Z}^n\} : y^n \in \mathcal{Y}^n\}$. Conditioned on $X^n = x^n$, the encoder prepares Q in the state $\rho_Q^{x^n}$. The quantum system Q is sent to the decoder. Conditioned on $Y^n = y^n$, the decoder applies the POVM $\{D_{z^n|y^n} : z^n \in \mathcal{Z}^n\}$ to Q, and the classical outcome is denoted by Z^n. Consequently, the joint distribution of (X^n, Y^n, Z^n) is given by

$$\mathsf{P}_{X^nY^nZ^n}(x^n, y^n, z^n)$$

$$= \Big(\prod_{i=1}^{n} \mathsf{P}_{XY}(x_i, y_i) \Big) \operatorname{Tr} \left[D_{z^n|y^n} \rho_Q^{x^n} \right] \tag{6}$$

The quantum-message model is not intended to replace the classical-message formulations above. Rather, it serves as a natural benchmark for comparing the communication capabilities of unassisted quantum messages with those of classical messages assisted by entanglement in the lossy source coding setting considered here.

3.4 Rate–distortion regions

We now define the asymptotic rate–distortion regions associated with the classes of codes introduced above. For a blocklength-n code, let $\mathcal{M}^{(n)}$ denote its classical message alphabet or $Q^{(n)}$ its quantum message system, as appropriate. Define its communication rate as

$$R_n \triangleq \begin{cases} \dfrac{\log_2 |\mathcal{M}^{(n)}|}{n}, & \text{for classical-message codes,} \\ \dfrac{\log_2 \dim(Q^{(n)})}{n}, & \text{for quantum-message codes,} \end{cases} \tag{7}$$

and its average distortion as

$$D_n \triangleq \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[d(X_i, Y_i, Z_i)]. \tag{8}$$

For any of the above classes of codes, a rate–distortion pair $(R, D) \in \mathbb{R}_+^2$ is said to be (asymptotically) achievable if there exists a sequence of blocklength-n codes of that class such that

$$\limsup_{n\to\infty} R_n \leq R, \qquad \limsup_{n\to\infty} D_n \leq D. \tag{9}$$

The rate–distortion region of a given class of codes is defined as the set of all achievable rate–distortion pairs. For classical-message codes, let $\mathrm{a} \in \{\text{un-a}, \text{ent-a}, \text{ns-a}\}$ specify the type of assistance. We denote the corresponding rate–distortion region by $\mathcal{R}^{\text{c-m,a}}$, and that of unassisted quantum-message codes by $\mathcal{R}^{\text{q-m,un-a}}$.

For any of the rate–distortion regions $\mathcal{R}$ defined above, the corresponding rate–distortion function is defined as

$$R(D) \triangleq \inf\{R \colon (R, D) \in \mathcal{R}\}. \tag{10}$$

Remark 1 (Derandomization)**.** *We note that shared randomness between the encoder and the decoder cannot enlarge the classical-message unassisted rate–distortion region. In particular, a classical-message code assisted by shared randomness corresponds to the non-signaling box having the form*

$$\sum_{\lambda} p_\lambda \mathsf{E}_\lambda(m \mid x^n) \mathsf{D}_\lambda(z^n \mid \widetilde{m}, y^n), \tag{11}$$

where $(p_\lambda)_\lambda$ forms a probability distribution, and $(\mathsf{E}_\lambda, \mathsf{D}_\lambda)$ correspond to an unassisted code for each λ. The average distortion of the code assisted by shared randomness is thus the expected value of the distortion values of the unassisted codes. Thus, there exists an unassisted code (with the same rate) that achieves a distortion no greater than the distortion of the assisted code.

4 Results

4.1 General setting

The results in this subsection apply to the general source coding model defined above. Recall that a problem is specified by a tuple of finite alphabets $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$, a distribution P_{XY}, and a distortion function $d\colon \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \to \mathbb{R}_+$.

Proposition 1 ([3, Problem 11.7], [32])**.** *Without assistance, the classical-message rate–distortion region is*

$$\mathcal{R}^{\text{c-m,un-a}} = \bigcup_{\mathsf{P}_{U|X}, z\colon \mathcal{U}\times\mathcal{Y}\to\mathcal{Z}} \left\{ \begin{array}{l} (R, D) \in \mathbb{R}_+^2 : \\ R \geq I(X; U \mid Y) \\ D \geq \mathbb{E}[d(X, Y, z(U, Y))] \end{array} \right\} \tag{12}$$

where the union is over all conditional distributions $\mathsf{P}_{U|X}$ and deterministic mappings $z\colon \mathcal{U} \times \mathcal{Y} \to \mathcal{Z}$, with the joint distribution given by $\mathsf{P}_{UXY} = \mathsf{P}_{XY}\mathsf{P}_{U|X}$. Moreover, the cardinality of the auxiliary alphabet may be restricted to $|\mathcal{U}| \leq |\mathcal{X}| + 1$.

Proof. The proposition follows directly from the Wyner–Ziv theorem [3, Thm. 11.3], extended to side-information-dependent distortion measures as described in [3, Problem 11.7]. □

Proposition 2. *With entanglement assistance, the classical-message rate–distortion region is*

$$\mathcal{R}^{\text{c-m,ent-a}} = \operatorname{cl}\left(\bigcup_{\{\rho_Q^x\}, \{\{\Lambda_{z|y}\}_z\}_y} \left\{ \begin{array}{l} (R, D) \in \mathbb{R}_+^2 : \\ R \geq I(X; Q \mid Y) \\ D \geq \mathbb{E}[d(X, Y, Z)] \end{array} \right\} \right) \tag{13}$$

where $\operatorname{cl}(\cdot)$ denotes closure, and the union is over all finite-dimensional quantum systems Q, classical-quantum channels $\{\rho_Q^x\colon x \in \mathcal{X}\}$, and families of POVMs $\{\{\Lambda_{z|y}\colon z \in \mathcal{Z}\}\colon y \in \mathcal{Y}\}$ on Q. The conditional mutual information $I(X; Q \mid Y)$ is evaluated with respect to the classical-quantum state

$$\rho_{XYQ} = \sum_{x,y} \mathsf{P}_{XY}(x, y)\, |x, y\rangle\langle x, y| \otimes \rho_Q^x, \tag{14}$$

which satisfies the Markov condition $Q - X - Y$. The classical random variable Z is obtained by measuring Q using the POVM indexed by Y, yielding the joint distribution

$$\mathsf{P}_{XYZ}(x, y, z) = \mathsf{P}_{XY}(x, y) \operatorname{Tr}[\Lambda_{z|y}\rho_Q^x]. \tag{15}$$

Proof sketch. Achievability follows from the remote state preparation theorem of Beigi and Gohari [23, Thm. 5]. The closure is needed as we do not know whether the union is closed due to a lack of a dimension upper bound on the auxiliary quantum system. Alternatively, specializing [33, Thm. 13] or [25, Sec. II-A] to classical source and side-information registers gives an entanglement-assisted quantum communication rate of $\frac{1}{2} I(X; Q \mid Y)$. When combined with teleportation we obtain the stated classical communication rate. The converse follows from standard information-theoretic arguments using the no-communication theorem followed by single-letterization. For completeness, we provide a detailed proof in Appendix A. □

Theorem 1. *With general non-signaling assistance, the classical-message rate–distortion region is*

$$\mathcal{R}^{\text{c-m,ns-a}} = \bigcup_{\mathsf{P}_{Z|XY}} \left\{ \begin{array}{l} (R, D) \in \mathbb{R}_+^2 : \\ R \geq I(X; Z \mid Y) \\ D \geq \mathbb{E}[d(X, Y, Z)] \end{array} \right\} \tag{16}$$

where the union is over all conditional distributions $\mathsf{P}_{Z|XY}$, with the joint distribution of (X, Y, Z) given by $\mathsf{P}_{XY}\mathsf{P}_{Z|XY}$.

Proof sketch. The full proof is given in Appendix B. The achievability argument parallels the twirling argument in [6, Thm. 2] (also see [1]), with the roles of the encoder and decoder reversed. The key idea is that, given any non-signaling-assisted code with encoder side information, one can construct another non-signaling-assisted code without encoder side information that induces the same conditional reconstruction distribution. Consequently, under non-signaling assistance, the rate–distortion region is the same with or without encoder side information and therefore contains the classical unassisted region with encoder side information. See Remark 2 for further discussion. □

Remark 2 (Virtual signaling of decoder side information)**.** *The non-signaling-assisted classical-message rate–distortion region coincides with the classical region $\mathcal{R}_{X|Y}$ when the side information is available at both the encoder and the decoder. Its boundary is described by the conditional rate–distortion function*

$$R_{X|Y}(D) = \inf_{\substack{\mathsf{P}_{Z|XY}: \\ \mathbb{E}[d(X,Y,Z)] \leq D}} I(X; Z \mid Y). \tag{17}$$

It was used as a classical benchmark in the Wyner–Ziv problem [5]. Zamir [7] established universal bounds on the rate reduction $R^{\text{c-m,un-a}}(D) - R_{X|Y}(D)$ obtainable by additionally providing side information to the encoder for difference or balanced distortion measures, including squared-error and binary Hamming distortion. Linder, Zamir, and Zeger [32] studied conditional and Wyner–Ziv rate–distortion functions for side-information-dependent distortion measures, including their high-resolution behavior and the gap between them. Theorem 1 thus provides an alternative operational interpretation of the conditional rate–distortion function through non-signaling assistance between the encoder and decoder.

The following proposition establishes a general relationship between the classical-message rate–distortion regions and the unassisted quantum-message rate–distortion region.

Proposition 3. $\mathcal{R}^{\text{c-m,un-a}} \subseteq \mathcal{R}^{\text{q-m,un-a}} \subseteq \mathcal{R}^{\text{c-m,ent-a}} \subseteq \mathcal{R}^{\text{c-m,ns-a}}$.

Proof. First, every entanglement-assisted code is also a non-signaling-assisted code, so $\mathcal{R}^{\text{c-m,ent-a}} \subseteq \mathcal{R}^{\text{c-m,ns-a}}$. Next, the inclusion $\mathcal{R}^{\text{c-m,un-a}} \subseteq \mathcal{R}^{\text{q-m,un-a}}$ follows because a d-dimensional quantum system can be used to encode a classical d-ary message. Therefore, any rate–distortion pair achievable by unassisted classical-message codes is achievable by unassisted quantum-message codes. Finally, the remaining inclusion $\mathcal{R}^{\text{q-m,un-a}} \subseteq \mathcal{R}^{\text{c-m,ent-a}}$ follows from the remote state preparation theorem [23, Thm. 5]. The key idea is that an unassisted quantum-message code can be simulated by entanglement-assisted classical communication with asymptotically vanishing overhead in rate and distortion. A detailed proof is given in Appendix C. □

4.2 Graph-based side information and Hamming distortion

We next specialize to a class of sources for which the decoder side information identifies two equally likely source symbols. Each problem in this class is specified by a finite, simple, undirected graph $\mathcal{G} = (\mathcal{X}, \mathcal{Y})$ with at least one edge, where $\mathcal{X}$ and $\mathcal{Y}$ are the vertex and edge sets, respectively. We identify each edge $y \in \mathcal{Y}$ with the unordered pair of its endpoints.

The side information Y is uniformly distributed over $\mathcal{Y}$. Conditioned on $Y = y$, the source X is uniformly distributed over the two endpoints of y. Equivalently, the joint distribution is

$$\mathsf{P}_{XY}(x, y) = \frac{\mathbb{I}[x \in y]}{2|\mathcal{Y}|}, \qquad (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{18}$$

In particular,

$$\mathsf{P}_X(x) = \frac{\deg_{\mathcal{G}}(x)}{2|\mathcal{Y}|}, \tag{19}$$

where $\deg_{\mathcal{G}}(x)$ is the degree of vertex x. Thus, X need not be uniformly distributed over the vertices. Isolated vertices have zero probability and may be omitted without affecting the problem.

The encoder observes the source vertex X but not the edge Y, while the decoder observes Y and must determine which endpoint was selected. At blocklength n, the source consists of n independent copies of this experiment, and the encoder and decoder operate on X^n and Y^n, respectively, as in the general formulation. The reconstruction alphabet is $\mathcal{Z} = \mathcal{X}$, and the distortion measure is the side-information-independent Hamming distortion

$$d(x, y, z) = \mathbb{I}[x \neq z], \qquad (x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}. \tag{20}$$

The average distortion is therefore the average symbol error probability. Without loss of generality, the decoder may restrict each reconstruction Z_i to the endpoints of Y_i: an output outside Y_i is always incorrect and can be replaced by either endpoint without increasing distortion.

Note that X has two equally likely possibilities given each $Y = y$, so $H(X \mid Y) = 1$ bit. Without communication, choosing either endpoint of the observed edge gives distortion $1/2$, which is optimal because the two endpoints are equally likely. Hence all the rate–distortion functions considered here are zero for $D \geq 1/2$, and we focus on $0 \leq D \leq 1/2$. When needed, we indicate the underlying graph by a subscript, as in $R_{\mathcal{G}}^{\text{c-m,un-a}}(D)$, $R_{\mathcal{G}}^{\text{c-m,ent-a}}(D)$, $R_{\mathcal{G}}^{\text{c-m,ns-a}}(D)$, and $R_{\mathcal{G}}^{\text{q-m,un-a}}(D)$.

Example 1 (The triangle graph $\mathcal{K}_3$)**.** *Let $\mathcal{G} = \mathcal{K}_3$, the triangle graph, with*

$$\mathcal{X} = \{1, 2, 3\}, \qquad \mathcal{Y} = \big\{\{1, 2\}, \{2, 3\}, \{3, 1\}\big\}. \tag{21}$$

The joint distribution is

$\mathsf{P}_{XY}(x, y)$	$y = \{1, 2\}$	$y = \{2, 3\}$	$y = \{3, 1\}$
$x = 1$	$1/6$	0	$1/6$
$x = 2$	$1/6$	$1/6$	0
$x = 3$	0	$1/6$	$1/6$

. (22)

Both X and Y are uniform on their respective three-element alphabets. For example, if the decoder observes $Y = \{1, 2\}$, it knows that X is either 1 or 2, each with probability $1/2$, and must distinguish them using the encoder's message and any allowed assistance. If the encoder observes $X = 1$, it does not know whether the decoder sees $\{1, 2\}$ or $\{3, 1\}$.

At $D = 0$, Proposition 1, Theorem 1, and Proposition 3 imply that

$$R_{\mathcal{G}}^{\text{c-m,un-a}}(0) = R_{\mathcal{G}}^{\text{c-m,ent-a}}(0) = R_{\mathcal{G}}^{\text{q-m,un-a}}(0) = R_{\mathcal{G}}^{\text{c-m,ns-a}}(0) = H(X \mid Y) = 1. \tag{23}$$

This equals the Slepian–Wolf rate for asymptotically lossless reconstruction [34]. Thus, any separation of the rates requires $0 < D < 1/2$. The following results characterize the assisted rates near $D = 1/2$ and quantify their advantage over unassisted classical communication for complete graphs. Write $h_2(p) = -p \log_2 p - (1-p)\log_2(1-p)$ for the binary entropy function, with $0 \log_2 0 = 0$. All asymptotic expansions below are taken as $\epsilon \downarrow 0$ with the graph fixed; the remainder terms may depend on the graph.

Theorem 2 (NS-a, graph)**.** *For every finite, simple, undirected graph $\mathcal{G}$ with at least one edge and every $0 \le D \le 1/2$,*

$$R_{\mathcal{G}}^{\text{c-m,ns-a}}(D) = 1 - h_2(D). \tag{24}$$

In particular,

$$R_{\mathcal{G}}^{\text{c-m,ns-a}}\left(\frac{1}{2} - \epsilon\right) = \frac{2}{\ln 2}\epsilon^2 + o(\epsilon^2). \tag{25}$$

Proof. See Appendix D. □

Theorem 3 (Ent-a, graph)**.** *For every finite, simple, undirected graph $\mathcal{G}$ with at least one edge,*

$$R_{\mathcal{G}}^{\text{c-m,ent-a}}\left(\frac{1}{2} - \epsilon\right) = \frac{2}{\ln 2}\epsilon^2 + o(\epsilon^2). \tag{26}$$

Proof. See Appendix E. □

Corollary 1. *For every graph $\mathcal{G}$ in Theorem 3,*

$$\lim_{D \uparrow 1/2} \frac{R_{\mathcal{G}}^{\text{c-m,ns-a}}(D)}{R_{\mathcal{G}}^{\text{c-m,ent-a}}(D)} = 1. \tag{27}$$

Theorem 4 (Un-a, complete graph rate lower bound)**.** *Let $\mathcal{K}_m$ denote the complete graph on $m \ge 3$ vertices. For each fixed m,*

$$R_{\mathcal{K}_m}^{\text{c-m,un-a}}\left(\frac{1}{2} - \epsilon\right) \ge \frac{3m}{(m+1)\ln 2}\epsilon^2 + o(\epsilon^2). \tag{28}$$

Proof. See Appendix F. □

Corollary 2 (Entanglement-assisted rate advantage for complete graphs)**.** *For every $m \ge 3$ and every $\eta > 0$, for all $D < 1/2$ sufficiently close to $1/2$,*

$$R_{\mathcal{K}_m}^{\text{c-m,ent-a}}(D) \le \left(\frac{2(m+1)}{3m} + \eta\right) R_{\mathcal{K}_m}^{\text{c-m,un-a}}(D). \tag{29}$$

Consequently, entanglement assistance strictly reduces the classical-message rate for all $D < 1/2$ sufficiently close to $1/2$. The entanglement-assisted rate is at most a fraction arbitrarily close to $2(m+1)/(3m)$ of the unassisted rate. This factor is $8/9$ for the triangle and tends to $2/3$ as $m \to \infty$. By Corollary 1, non-signaling assistance offers no further multiplicative rate reduction in this limit.

5 Conclusion

For lossy source coding with side information at the decoder, the non-signaling-assisted classical-message rate–distortion region coincides with the classical region when the side information is available at both terminals. Thus, non-signaling assistance virtually signals the decoder side information to the encoder for this task. This can be viewed as the source-coding dual of the previously established virtual signaling of CSIT in channel coding with transmitter side information. Turning to quantum entanglement assistance, we have identified a simple class of sources associated with graphs for which entanglement assistance asymptotically recovers the full advantage of non-signaling assistance in the low-rate limit. Future directions include generalizing graph-based sources to hypergraph-based sources or other forms of dependence between the source and side information, including function computation problems.

A Proof of Proposition 2

We first establish the converse from an arbitrary entanglement-assisted classical-message code, formulated as an explicit process of encoder state preparation followed by a decoder POVM. Achievability follows from the remote state preparation theorem of Beigi and Gohari [23, Thm. 5].

Converse

Step 1: Bounding the rate. Fix a blocklength-n entanglement-assisted code with message size $|\mathcal{M}|$. Let B denote the decoder's share of the entangled quantum system immediately after the encoder's measurement and before the decoder performs any processing based on (M, Y^n). Conditioned on $X^n = x^n$, the encoder's measurement results in a classical-quantum state $\rho_{MB}^{x^n}$, with M a classical register. By the no-communication theorem, the marginal state $\rho_B^{x^n}$ does not depend on x^n. This independence also holds conditioned on any $Y^n = y^n$, since the state is defined before the decoder performs any processing and the initial shared entangled state is independent of (X^n, Y^n). It follows that

$$I(X^n; B \mid Y^n) = 0, \tag{30}$$

and, since M is classical,

$$\begin{aligned}
\log_2 |\mathcal{M}| &\geq H(M) \geq I(X^n; M \mid BY^n) && (31)\\
&\overset{(30)}{=} I(X^n; MB \mid Y^n) && (32)\\
&= \sum_{i=1}^{n} I(X_i; MB \mid X^{i-1}Y^n) && (33)\\
&= \sum_{i=1}^{n} I(X_i; M, B, X^{i-1}, Y^{\setminus i} \mid Y_i) && (34)\\
&\geq \sum_{i=1}^{n} I(X_i; \underbrace{(M, B, Y^{\setminus i})}_{\triangleq Q_i} \mid Y_i) && (35)\\
&= nI(X_T; Q_T \mid Y_T, T) && (36)
\end{aligned}$$

$$= nI(X_T; Q_T, T \mid Y_T) \tag{37}$$

Equality (34) follows from the i.i.d. source distribution, which implies that (X_i, Y_i) is independent of $(X^{i-1}, Y^{\setminus i})$. In (35), for each $i \in [n]$, we introduce the quantum auxiliary system $Q_i \triangleq (M, B, Y^{\setminus i})$, where $Y^{\setminus i} \triangleq (Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n)$. In (36), we introduce a time-sharing random variable T, uniformly distributed over $[n]$ and independent of (X^n, Y^n, M, B). Step (37) holds because T is independent of (X_T, Y_T) as the source is i.i.d.

Step 2: Preserving the distortion. Now consider the decoder's POVM. Conditioned on $(Y^n = y^n, M = m)$, the decoder applies the POVM $\{D_{z^n|y^n,m}\}_{z^n \in \mathcal{Z}^n}$ to the quantum system B, producing the classical reconstruction Z^n. Let $\mathsf{P}_{X^n Y^n T M Z^n}$ denote the joint distribution of the classical random variables induced by the original code.

For each $i \in [n]$, consider the marginal POVM that produces Z_i while ignoring $Z^{\setminus i}$. Since $Q_i = (M, B, Y^{\setminus i})$, this marginal POVM can equivalently be represented as a POVM $\{\Lambda^{(i)}_{z|y}\}_{z \in \mathcal{Z}}$ on Q_i, indexed by $Y_i = y$.

Now consider a hypothetical experiment in which, conditioned on $T = i$ and $Y_T = y$, the POVM $\{\Lambda^{(i)}_{z|y}\}_{z \in \mathcal{Z}}$ is applied to Q_i. Equivalently, this can be viewed as a POVM on the joint system (Q_T, T), conditioned on $Y_T = y$. Denote its classical outcome by Z, and let $\mathsf{P}'_{X^n Y^n T M Z}$ denote the distribution induced by this hypothetical experiment. We emphasize that the systems Q_i for different i do not include separate copies of B. In particular, the hypothetical experiment requires only one measurement corresponding to a selected index i.

By construction, for every $i \in [n]$, $\mathsf{P}'_{X^n Y^n M Z|T=i} = \mathsf{P}_{X^n Y^n M Z_i}$, since conditioned on $T = i$, the hypothetical measurement reproduces the original decoder's marginal POVM for Z_i. In particular, this implies that (X_T, Y_T, Z_T) under P and (X_T, Y_T, Z) under P' have the same joint distribution. Therefore,

$$\begin{aligned}
\mathbb{E}_{\mathsf{P}'}[d(X_T, Y_T, Z)] &= \mathbb{E}_{\mathsf{P}}[d(X_T, Y_T, Z_T)] \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{\mathsf{P}}[d(X_i, Y_i, Z_i)] \qquad (38) \\
&= D_n \qquad (39)
\end{aligned}$$

Define $X \triangleq X_T$, $Y \triangleq Y_T$, and $Q \triangleq (Q_T, T)$. From the i.i.d. source law and the Markov chains $Q_i - X_i - Y_i$, we have $Q - X - Y$. Moreover, the hypothetical measurement defines a family of POVMs on Q, indexed by Y, whose classical outcome is Z.

Together with the previously established mutual information bound (37), we conclude that every blocklength-n code induces a single-letter quantum system Q such that $Q - X - Y$, and a family of POVMs $\{\{\Lambda_{z|y}\}_z\}_y$ satisfying

$$R_n \geq I(X; Q \mid Y), \quad D_n = \mathbb{E}_{\mathsf{P}'}[d(X, Y, Z)], \tag{40}$$

where $(X, Y) \sim \mathsf{P}_{XY}$, and Z is the classical outcome of the POVM $\{\Lambda_{z|y}\}_z$ on Q, indexed by $Y = y$.

Step 3: Completing the converse. Now consider an arbitrary achievable rate–distortion pair $(R, D) \in \mathcal{R}^{\text{c-m,ent-a}}$. By the definition of achievability, there exists a sequence of blocklength-n codes such that

$$\limsup_{n \to \infty} R_n \leq R, \qquad \limsup_{n \to \infty} D_n \leq D. \tag{41}$$

Consequently, for any $\epsilon > 0$ and sufficiently large n, we have

$$R + \epsilon \geq R_n \geq I(X; Q \mid Y), \tag{42}$$

$$D + \epsilon \geq D_n = \mathbb{E}_{\mathsf{P}'}[d(X, Y, Z)]. \tag{43}$$

Thus, $(R + \epsilon, D + \epsilon)$ belongs to the union on the right-hand side of Proposition 2, before taking its closure. Since $\epsilon > 0$ is arbitrary, (R, D) belongs to the closure of that union. This establishes the converse. □

Achievability

Fix a classical-quantum channel $x \mapsto \rho_Q^x$ and POVMs $\{\Lambda_{z|y}\}_{z \in \mathcal{Z}}$ satisfying the distortion constraint. Consider the state

$$\rho_{XYQ} = \sum_{x,y} \mathsf{P}_{XY}(x, y) |x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y \otimes \rho_Q^x. \tag{44}$$

In particular, the Markov condition $Q - X - Y$ holds. By the remote state preparation theorem with classical decoder side information [23, Thm. 5], for any $\epsilon > 0$ and sufficiently large n, Alice can use shared entanglement and a one-way classical message of size at most $2^{n(I(X;Q|Y)+\epsilon)}$ to remotely prepare a quantum system $\widetilde{Q}^n$ at Bob's terminal such that the resulting joint state $\sigma_{X^n Y^n \widetilde{Q}^n}$ satisfies

$$\left\| \sigma_{X^n Y^n \widetilde{Q}^n} - \rho_{XYQ}^{\otimes n} \right\|_1 \to 0 \quad \text{as } n \to \infty. \tag{45}$$

Bob then measures each subsystem $\widetilde{Q}_i$ using the POVM $\{\Lambda_{z|Y_i}\}_{z \in \mathcal{Z}}$ indexed by his side information Y_i, and outputs the measurement outcome Z_i, for each $i \in [n]$.

By the monotonicity of trace distance under completely positive trace-preserving (CPTP) maps [35, Exercise 9.1.9], including measurements, the induced distribution of (X^n, Y^n, Z^n) converges in total variation to the desired i.i.d. distribution $\mathsf{P}_{XYZ}^{\otimes n}$, where

$$\mathsf{P}_{XYZ}(x, y, z) = \mathsf{P}_{XY}(x, y) \operatorname{Tr}[\Lambda_{z|y} \rho_Q^x]. \tag{46}$$

Since d is bounded, the average distortion converges to $\mathbb{E}[d(X, Y, Z)]$. Consequently, every rate–distortion pair (R, D) satisfying

$$R > I(X; Q \mid Y), \qquad D \geq \mathbb{E}[d(X, Y, Z)] \tag{47}$$

is achievable. Taking the closure over all admissible quantum systems and POVMs completes the achievability proof. □

B Proof of Theorem 1

The converse compares the conditional distribution induced by a non-signaling-assisted classical-message code with a reference conditional distribution induced without communication. Achievability follows the idea of "virtual signaling of CSIT via non-signaling assistance" in [6]: at every blocklength, encoder side information can be removed without changing the conditional reconstruction distribution.

Converse

Step 1: Comparing with a distribution without communication. Consider a blocklength-n non-signaling-assisted classical-message code $\mathsf{B}_{MZ^n|X^n\widetilde{M}Y^n}$, or simply B. By non-signaling, the decoder marginal

$$\mathsf{B}_{Z^n|\widetilde{M}Y^n}(z^n \mid \widetilde{m}, y^n) \triangleq \sum_{m\in\mathcal{M}} \mathsf{B}(m, z^n \mid x^n, \widetilde{m}, y^n) \tag{48}$$

is independent of x^n. Define a reference conditional distribution

$$\mathsf{Q}_{Z^n|Y^n}(z^n \mid y^n) = \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \mathsf{B}_{Z^n|\widetilde{M}Y^n}(z^n \mid m, y^n), \tag{49}$$

which corresponds to the distribution of Z^n given Y^n when the decoder locally generates a uniformly distributed message $\widetilde{M}$, independently of the source (X^n, Y^n), and uses it in place of the communicated message.

On the other hand, let $\mathsf{P}_{X^nY^nZ^nM}$ be the joint distribution induced by the code. For every x^n, y^n, z^n, we have

$$\begin{aligned}
\mathsf{P}_{Z^n|X^nY^n}(z^n \mid x^n, y^n) &= \sum_{m\in\mathcal{M}} \mathsf{B}(m, z^n \mid x^n, m, y^n) && (50)\\
&\le \sum_{m\in\mathcal{M}} \sum_{m'\in\mathcal{M}} \mathsf{B}(m', z^n \mid x^n, m, y^n) && (51)\\
&= \sum_{m\in\mathcal{M}} \mathsf{B}_{Z^n|\widetilde{M}Y^n}(z^n \mid m, y^n) && (52)\\
&= |\mathcal{M}| \times \mathsf{Q}_{Z^n|Y^n}(z^n \mid y^n). && (53)
\end{aligned}$$

Consequently, for all x^n, y^n,

$$D_{\mathrm{KL}}\big(\mathsf{P}_{Z^n|x^ny^n}\big\|\mathsf{Q}_{Z^n|y^n}\big) \le \log_2|\mathcal{M}|. \tag{54}$$

Taking the average on both sides,

$$\begin{aligned}
\log_2|\mathcal{M}| &\ge \sum_{x^n,y^n} \mathsf{P}_{X^nY^n}(x^n, y^n) D_{\mathrm{KL}}\big(\mathsf{P}_{Z^n|x^ny^n}\big\|\mathsf{Q}_{Z^n|y^n}\big)\\
&= \sum_{y^n} \mathsf{P}_{Y^n}(y^n) \sum_{x^n} \mathsf{P}_{X^n|Y^n}(x^n \mid y^n) D_{\mathrm{KL}}\big(\mathsf{P}_{Z^n|x^ny^n}\big\|\mathsf{Q}_{Z^n|y^n}\big) && (55)\\
&= I(X^n; Z^n \mid Y^n) + \sum_{y^n} \mathsf{P}_{Y^n}(y^n) D_{\mathrm{KL}}\big(\mathsf{P}_{Z^n|y^n}\big\|\mathsf{Q}_{Z^n|y^n}\big) && (56)\\
&\ge I(X^n; Z^n \mid Y^n). && (57)
\end{aligned}$$

Step 2: Bounding the rate. Combining these inequalities yields

$$\begin{aligned}
&\log_2|\mathcal{M}| \ge I(X^n; Z^n \mid Y^n) && (58)\\
&= H(X^n \mid Y^n) - H(X^n \mid Y^nZ^n) && (59)
\end{aligned}$$

$$= \sum_{i=1}^{n} H(X_i \mid Y_i) - H(X^n \mid Y^n Z^n) \tag{60}$$

$$\geq \sum_{i=1}^{n} H(X_i \mid Y_i) - \sum_{i=1}^{n} H(X_i \mid Y^n Z^n) \tag{61}$$

$$\geq \sum_{i=1}^{n} H(X_i \mid Y_i) - \sum_{i=1}^{n} H(X_i \mid Y_i Z_i) \tag{62}$$

$$= \sum_{i=1}^{n} I(X_i; Z_i \mid Y_i) \tag{63}$$

where equality (60) follows from the i.i.d. source assumption.

Step 3: Obtaining a single-letter distribution. Let $\overline{\mathsf{P}}_{Z|XY} \triangleq \frac{1}{n}\sum_{i=1}^{n} \mathsf{P}_{Z_i|X_iY_i}$, and define the joint distribution $\overline{\mathsf{P}}_{XYZ} \triangleq \mathsf{P}_{XY}\overline{\mathsf{P}}_{Z|XY}$. By the convexity of mutual information in the conditional distribution [36, Thm. 2.7.4], we obtain

$$I_{\overline{\mathsf{P}}}(X; Z \mid Y = y) \leq \frac{1}{n}\sum_{i=1}^{n} I_{\mathsf{P}}(X_i; Z_i \mid Y_i = y). \tag{64}$$

for each $y \in \mathcal{Y}$. Since $\overline{\mathsf{P}}_Y = \mathsf{P}_{Y_i} = \mathsf{P}_Y$,

$$I_{\overline{\mathsf{P}}}(X; Z \mid Y) \leq \frac{1}{n}\sum_{i=1}^{n} I_{\mathsf{P}}(X_i; Z_i \mid Y_i). \tag{65}$$

Therefore,

$$R_n = \frac{\log_2 |\mathcal{M}|}{n} \geq I_{\overline{\mathsf{P}}}(X; Z \mid Y). \tag{66}$$

Moreover, by linearity of expectation, the distortion induced by $\mathsf{P}^{(n)}$ equals the average distortion of the blocklength-n code, i.e.,

$$D_n = \mathbb{E}_{\overline{\mathsf{P}}}[d(X, Y, Z)]. \tag{67}$$

Consequently, every achievable rate–distortion pair (R, D) belongs to the closure of the union in Theorem 1. Since the set of conditional distributions $\mathsf{P}_{Z|XY}$ is compact, and both conditional mutual information and expected distortion are continuous in $\mathsf{P}_{Z|XY}$, this union is already closed. This completes the converse proof. □

Achievability

Step 1: Removing encoder side information. We establish a slightly stronger, finite-blocklength statement. In an augmented model, suppose that the encoder is also provided with the side information Y^n. We refer to this as the encoder-side-information (ESI) model. Consider an arbitrary non-signaling-assisted code in this augmented model, specified by a non-signaling box

$$\mathsf{B}^{\mathrm{ESI}}_{MZ^n|X^nY^n_{\mathsf{E}}\widetilde{M}Y^n_{\mathsf{D}}}(m, z^n \mid x^n, y^n_{\mathsf{E}}, \widetilde{m}, y^n_{\mathsf{D}}), \tag{68}$$

where the subscripts E and D distinguish the copies of the side information supplied to the encoder and decoder. We show next that the encoder's side information can be removed while preserving the conditional distribution of Z^n given (X^n, Y^n).

Without loss of generality, let $\mathcal{M} = \{0, 1, \ldots, |\mathcal{M}|-1\}$, and write $m + t \pmod{|\mathcal{M}|}$ as $m \oplus t$ for $m, t \in \mathcal{M}$. Define a new bipartite box, which uses decoder side information only, by

$$\begin{aligned}&\widetilde{\mathsf{B}}_{MZ^n|X^n\widetilde{M}Y^n}(m, z^n \mid x^n, \widetilde{m}, y^n)\\ &\quad\triangleq \frac{1}{|\mathcal{M}|}\sum_{t\in\mathcal{M}} \mathsf{B}^{\mathrm{ESI}}(m\oplus t, z^n \mid x^n, y^n, \widetilde{m}\oplus t, y^n).\end{aligned} \tag{69}$$

Step 2: Verifying the non-signaling constraints. We next show that $\widetilde{\mathsf{B}} \in \mathcal{P}^{\mathrm{NS}}(\mathcal{M}; \mathcal{Z}^n \mid \mathcal{X}^n; \mathcal{M}\times\mathcal{Y}^n)$. The normalization and nonnegativity of $\widetilde{\mathsf{B}}$ are immediate. For the encoder marginal, non-signaling of $\mathsf{B}^{\mathrm{ESI}}$ gives

$$\sum_{z^n}\widetilde{\mathsf{B}}(m, z^n \mid x^n, \widetilde{m}, y^n) = \frac{1}{|\mathcal{M}|}\sum_{t\in\mathcal{M}} \mathsf{B}^{\mathrm{ESI}}_{M|X^nY^n_{\mathsf{E}}}(m\oplus t \mid x^n, y^n) \tag{70}$$

$$= \frac{1}{|\mathcal{M}|}\sum_{a\in\mathcal{M}} \mathsf{B}^{\mathrm{ESI}}_{M|X^nY^n_{\mathsf{E}}}(a \mid x^n, y^n) \tag{71}$$

$$= \frac{1}{|\mathcal{M}|}, \tag{72}$$

which is independent of $(\widetilde{m}, y^n)$. Similarly, for the decoder marginal,

$$\sum_{m}\widetilde{\mathsf{B}}(m, z^n \mid x^n, \widetilde{m}, y^n) = \frac{1}{|\mathcal{M}|}\sum_{t\in\mathcal{M}} \mathsf{B}^{\mathrm{ESI}}_{Z^n|\widetilde{M}Y^n_{\mathsf{D}}}(z^n \mid \widetilde{m}\oplus t, y^n) \tag{73}$$

which is independent of x^n. Thus $\widetilde{\mathsf{B}}$ is a valid non-signaling box in the original model, without side information at the encoder.

Step 3: Preserving the reconstruction distribution. Furthermore, for every (x^n, y^n, z^n), the conditional reconstruction distribution induced by the new box is

$$\begin{aligned}&\widetilde{\mathsf{P}}_{Z^n|X^nY^n}(z^n \mid x^n, y^n)\\ &= \sum_{m\in\mathcal{M}}\widetilde{\mathsf{B}}(m, z^n \mid x^n, m, y^n)\\ &= \frac{1}{|\mathcal{M}|}\sum_{t\in\mathcal{M}}\sum_{m\in\mathcal{M}} \mathsf{B}^{\mathrm{ESI}}(m\oplus t, z^n \mid x^n, y^n, m\oplus t, y^n) &(74)\\ &= \sum_{a\in\mathcal{M}} \mathsf{B}^{\mathrm{ESI}}(a, z^n \mid x^n, y^n, a, y^n) &(75)\\ &= \mathsf{P}^{\mathrm{ESI}}_{Z^n|X^nY^n}(z^n \mid x^n, y^n). &(76)\end{aligned}$$

The equality holds pointwise for all source and side-information sequences, rather than only in expected distortion. Hence every rate–distortion pair achievable with non-signaling assistance and encoder side information is achievable without encoder side information. The construction parallels the twirling argument of [6].

The achievability proof of Theorem 1 concludes by noting that the region in Theorem 1 is the classical rate–distortion region with Y^n available at both terminals. Since non-signaling assistance allows the receiver-side information to be virtually signaled to the encoder, it also achieves the region in Theorem 1. □

C Proof of Proposition 3

It remains to prove $\mathcal{R}^{\text{q-m,un-a}} \subseteq \mathcal{R}^{\text{c-m,ent-a}}$. Consider any blocklength-n quantum-message code with quantum message Q, encoder states $\{\rho_Q^{x^n}\}$, and decoder POVMs $\{D_{z^n|y^n}\}$. Treat each block of n source symbols as a supersymbol $\mathbf{X} = X^n$, with decoder side information $\mathbf{Y} = Y^n$ and target quantum system Q. The associated classical-quantum state is

$$\rho_{\mathbf{XY}Q} = \sum_{x^n,y^n} \mathsf{P}_{XY}^{\otimes n}(x^n, y^n)|x^n\rangle\langle x^n| \otimes |y^n\rangle\langle y^n| \otimes \rho_Q^{x^n}. \tag{77}$$

Since the Markov condition $Q - \mathbf{X} - \mathbf{Y}$ holds, the remote state preparation theorem [23, Thm. 5] applies to t independent supersymbols. For every $\epsilon > 0$, with unlimited shared entanglement, Alice can remotely prepare t copies of Q at Bob's terminal with asymptotically vanishing trace-distance error and a classical message of size at most

$$2^{t(I(\mathbf{X};Q|\mathbf{Y})+\epsilon)}. \tag{78}$$

The required classical communication rate per supersymbol is asymptotically no greater than the original quantum-message cost, since

$$I(\mathbf{X}; Q \mid \mathbf{Y}) = H(Q \mid \mathbf{Y}) - H(Q \mid \mathbf{XY}) \tag{79}$$

$$\leq H(Q \mid \mathbf{Y}) \leq \log_2 \dim Q, \tag{80}$$

where $H(Q \mid \mathbf{XY})$ is nonnegative because $\mathbf{X}$ and $\mathbf{Y}$ are classical. After remote state preparation, Bob applies the original POVM to each prepared system, using the corresponding side-information sequence y^n, to reconstruct each block. The joint classical law approaches the t-fold product of the law induced by the original quantum-message code as $t \to \infty$. Since d is bounded, the average distortion converges to that of the original code. Thus a quantum-message code with rate $(\log_2 \dim Q)/n$ and average distortion D_n can be simulated with arbitrarily small asymptotic overhead using entanglement-assisted classical communication. This establishes the desired inclusion. □

D Proof of Theorem 2

By Theorem 1, we evaluate $\min I(X; Z \mid Y)$ over $\mathsf{P}_{Z|XY}$ that yields distortion at most D. We first argue that there is no loss of generality to restrict Z to one of the endpoints of Y. Define $f\colon \mathcal{Y} \times \mathcal{Z} \to \mathcal{Z}$ by $f(y, z) = z$ if z is an endpoint of y, and $f(y, z) = z_0(y)$ otherwise, where $z_0(y)$ is an arbitrarily fixed endpoint of y. Given any $\mathsf{P}_{Z|XY}$, define $\widetilde{Z} = f(Y, Z)$. Note that whenever $Z = X$, we have that Z is an endpoint of Y, so $\widetilde{Z} = X$ and thus replacing Z with $\widetilde{Z}$ cannot increase the distortion. Moreover, the Markov chain $X - (Y, Z) - \widetilde{Z}$ implies $I(X; \widetilde{Z} \mid Y) \leq I(X; Z \mid Y)$. This shows that $\widetilde{Z}$ cannot increase either the distortion or the conditional mutual information.

In the following, it suffices to consider that Z is always an endpoint of Y. Let $E = \mathbb{I}[X \neq Z]$ and $p_e = \Pr(E = 1) \leq D$. Consequently, X is determined by (Y, Z, E).

$$I(X; Z \mid Y) = H(X \mid Y) - H(X \mid Y, Z) \geq 1 - H(E \mid Y, Z) \tag{81}$$

$$\geq 1 - H(E) = 1 - h_2(p_e) \geq 1 - h_2(D), \tag{82}$$

where the last inequality uses $D \leq 1/2$. Conversely, conditioned on each (X, Y) such that X is an endpoint of Y, let $Z = X$ with probability $1 - D$ and let Z be the other endpoint of Y with probability D. Recall that conditioned on Y, X is uniformly distributed over the two endpoints. We have that $H(X \mid Y, Z) = h_2(D)$, attaining the lower bound. Finally, the expansion

$$h_2\left(\frac{1}{2} + t\right) = 1 - \frac{2}{\ln 2} t^2 + O(t^4) \tag{83}$$

gives the claimed asymptotics. □

E Proof of Theorem 3

Preliminary: Designing operators through vector geometry. Let us first prove the following lemma, which essentially comes from Tsirelson's representation of quantum correlations [13].

Lemma 1 (Clifford representation of Euclidean inner products)**.** *For every integer $r \geq 1$, there exist a positive even integer d and a real-linear map Γ from $\mathbb{R}^r$ to the real symmetric $d \times d$ matrices such that, for all $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^r$,*

$$\operatorname{Tr}[\Gamma(\boldsymbol{a})] = 0, \tag{84}$$

$$\Gamma(\boldsymbol{a})\Gamma(\boldsymbol{b}) + \Gamma(\boldsymbol{b})\Gamma(\boldsymbol{a}) = 2\langle \boldsymbol{a}, \boldsymbol{b} \rangle I_d, \tag{85}$$

where I_d is the $d \times d$ identity matrix. Consequently,

$$\Gamma(\boldsymbol{a})^2 = \|\boldsymbol{a}\|_2^2 I_d, \tag{86}$$

$$\frac{1}{d} \operatorname{Tr}[\Gamma(\boldsymbol{a})\Gamma(\boldsymbol{b})] = \langle \boldsymbol{a}, \boldsymbol{b} \rangle. \tag{87}$$

Proof. For our purpose and for simplicity, we choose $d = 2^r$. Other constructions can give smaller dimensions, but we do not need to minimize d here. Let

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{88}$$

For $i \in [r]$, define

$$G_i \triangleq Z^{\otimes(i-1)} \otimes X \otimes I^{\otimes(r-i)}. \tag{89}$$

To see the pattern more explicitly,

$$G_1 = X \otimes I \otimes I \otimes \cdots \otimes I \tag{90}$$

$$G_2 = Z \otimes X \otimes I \otimes \cdots \otimes I \tag{91}$$

$$\vdots \tag{92}$$

$$G_r = Z \otimes Z \otimes \cdots \otimes Z \otimes X \tag{93}$$

Each G_i is real symmetric with trace 0, and $G_i^2 = I_d$. For $k < \ell$, the factors of G_k and G_ℓ commute at every tensor position except the kth, where $XZ = -ZX$. Thus $G_k G_\ell = -G_\ell G_k$ whenever $k \neq \ell$.

For $\boldsymbol{a} = (a_1, \ldots, a_r) \in \mathbb{R}^r$, set

$$\Gamma(\boldsymbol{a}) \triangleq \sum_{i=1}^{r} a_i G_i. \tag{94}$$

This map is real-linear, takes values in the real symmetric matrices, and satisfies $\mathrm{Tr}[\Gamma(\boldsymbol{a})] = 0$. Moreover,

$$\begin{aligned} &\Gamma(\boldsymbol{a})\Gamma(\boldsymbol{b}) + \Gamma(\boldsymbol{b})\Gamma(\boldsymbol{a}) \\ &\quad = \sum_{k,\ell=1}^{r} a_k b_\ell (G_k G_\ell + G_\ell G_k) \qquad (95) \\ &\quad = 2\sum_{k=1}^{r} a_k b_k I_d = 2\langle \boldsymbol{a}, \boldsymbol{b} \rangle I_d. \qquad (96) \end{aligned}$$

Taking $\boldsymbol{b} = \boldsymbol{a}$ gives (86). Taking the trace of (85) and using $\mathrm{Tr}[AB] = \mathrm{Tr}[BA]$ gives (87). □

Step 1: Constructing the auxiliary states. We proceed to establish the achievability of certain rate–distortion pairs by classical-message entanglement-assisted codes using Lemma 1 and Proposition 2. Given the graph $\mathcal{G}$, remove any isolated vertices and label the remaining vertices by $[N]$, where $N \geq 2$. Choose real unit vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N$ forming a regular simplex centered at the origin, so that $\|\boldsymbol{v}_i - \boldsymbol{v}_j\|_2$ is the same positive constant for all distinct $i, j \in [N]$. For example, in $\mathbb{R}^N$, take $\boldsymbol{v}_i = \sqrt{N/(N-1)}(\boldsymbol{e}_i - N^{-1}\mathbf{1})$, where $\boldsymbol{e}_i$ is the ith standard basis vector and $\mathbf{1}$ is the all-ones vector. It can be verified that, for every pair $i \neq j$,

$$\left\| \frac{\boldsymbol{v}_i - \boldsymbol{v}_j}{2} \right\|_2 = \sqrt{\frac{N}{2(N-1)}} \triangleq \alpha, \qquad \left\| \frac{\boldsymbol{v}_i + \boldsymbol{v}_j}{2} \right\|_2 = \sqrt{\frac{N-2}{2(N-1)}} \triangleq \beta. \tag{97}$$

Applying the construction in Lemma 1 with $r = N$ and $d = 2^N$ gives us the mapping $\Gamma \colon \mathbb{R}^r \to \mathbb{R}^{d \times d}$. To establish any achievable rate–distortion pair, let us recall the region in Proposition 2. For $0 \leq \lambda \leq 1$ to be determined later, let the encoder's auxiliary state be

$$\rho_Q^i = \frac{I_d + \lambda \Gamma(\boldsymbol{v}_i)}{d}. \tag{98}$$

We claim that these are density operators. To see this, note that Lemma 1 guarantees that $\mathrm{Tr}[\Gamma(\boldsymbol{v}_i)] = 0$, and that $\Gamma(\boldsymbol{v}_i)$ is real symmetric and $\Gamma(\boldsymbol{v}_i)^2 = I_d$, so its eigenvalues are ± 1.

Step 2: Constructing the measurements and calculating the distortion. For each edge $y = \{i, j\}$, choose an ordering (i, j) and set

$$\boldsymbol{a}_{ij} = \frac{\boldsymbol{v}_i - \boldsymbol{v}_j}{\|\boldsymbol{v}_i - \boldsymbol{v}_j\|_2} = \frac{\boldsymbol{v}_i - \boldsymbol{v}_j}{2\alpha}, \qquad \Lambda_{i|y} = \frac{I_d + \Gamma(\boldsymbol{a}_{ij})}{2}, \qquad \Lambda_{j|y} = \frac{I_d - \Gamma(\boldsymbol{a}_{ij})}{2}, \tag{99}$$

with all other POVM elements zero. By definition, $\boldsymbol{a}_{ij}$ is a unit vector, and thus $\{\Lambda_{i|y}, \Lambda_{j|y}\}$ form a POVM. Moreover, it can be verified that

$$\langle \boldsymbol{a}_{ij}, \boldsymbol{v}_i \rangle = \alpha, \qquad \langle \boldsymbol{a}_{ij}, \boldsymbol{v}_j \rangle = -\alpha. \tag{100}$$

Lemma 1 then implies that

$$\operatorname{Tr}[\Lambda_{i|y}\rho_Q^i] = \frac{1}{2}\left(1 + \frac{\lambda}{d}\operatorname{Tr}[\Gamma(\boldsymbol{a}_{ij})\Gamma(\boldsymbol{v}_i)]\right) = \frac{1+\lambda\alpha}{2}, \tag{101}$$

and likewise $\operatorname{Tr}[\Lambda_{j|y}\rho_Q^j] = (1+\lambda\alpha)/2$. Thus the distortion is

$$D_\lambda = 1 - (1+\lambda\alpha)/2 = \frac{1}{2} - \frac{\lambda\alpha}{2}. \tag{102}$$

Step 3: Calculating the rate. To calculate the rate, note that for any non-zero vector $\boldsymbol{u} \in \mathbb{R}^r$, Lemma 1 guarantees that $\Gamma(\boldsymbol{u})$ has eigenvalues $\pm\|\boldsymbol{u}\|_2$, each with multiplicity $d/2$ because $\operatorname{Tr}[\Gamma(\boldsymbol{u})] = 0$. Consequently,

$$H\left(\frac{I_d + \lambda\Gamma(\boldsymbol{u})}{d}\right) = \log_2 d - 1 + h_2\left(\frac{1+\lambda\|\boldsymbol{u}\|_2}{2}\right). \tag{103}$$

Equation (103) also holds for $\boldsymbol{u} = \boldsymbol{0}$.

Also, since $X = i$ and $X = j$ are equiprobable given $Y = \{i, j\}$, the mean state is

$$\rho_{Q|Y=\{i,j\}} = \frac{2I_d + \lambda\Gamma(\boldsymbol{v}_i) + \lambda\Gamma(\boldsymbol{v}_j)}{2d} = \frac{I_d + \lambda\Gamma((\boldsymbol{v}_i+\boldsymbol{v}_j)/2)}{d}. \tag{104}$$

Therefore,

$$I(X;Q \mid Y) = \frac{1}{|\mathcal{Y}|}\sum_{\{i,j\}\in\mathcal{Y}}\left[H\left(\rho_{Q|Y=\{i,j\}}\right) - \frac{H(\rho_Q^i) + H(\rho_Q^j)}{2}\right] \tag{105}$$

$$= h_2\left(\frac{1+\lambda\|\frac{\boldsymbol{v}_i+\boldsymbol{v}_j}{2}\|_2}{2}\right) - h_2\left(\frac{1+\lambda}{2}\right) \tag{106}$$

$$= h_2\left(\frac{1+\lambda\beta}{2}\right) - h_2\left(\frac{1+\lambda}{2}\right). \tag{107}$$

where Equation (106) follows by applying (103) to the states in (98) and (104).

Step 4: Establishing the low-rate asymptotics. Using Proposition 2, to see the rate at $D_\lambda = 1/2 - \epsilon$, set $\lambda = 2\epsilon/\alpha$ by (102), which lies in $[0,1]$ for $0 \le \epsilon \le \alpha/2$. Equation (107) yields

$$R_{\mathcal{G}}^{\text{c-m,ent-a}}\left(\frac{1}{2} - \epsilon\right) \le h_2\left(\frac{1}{2} + \frac{\beta}{\alpha}\epsilon\right) - h_2\left(\frac{1}{2} + \frac{\epsilon}{\alpha}\right) \tag{108}$$

$$\overset{(83)}{=} \frac{2(1-\beta^2)}{\alpha^2 \ln 2}\epsilon^2 + o(\epsilon^2) = \frac{2}{\ln 2}\epsilon^2 + o(\epsilon^2). \tag{109}$$

The last equality follows because $\alpha^2 + \beta^2 = 1$ as can be verified from (97). A matching lower bound can be established from the non-signaling result of Theorem 2:

$$R_{\mathcal{G}}^{\text{c-m,ent-a}}\left(\frac{1}{2} - \epsilon\right) \ge R_{\mathcal{G}}^{\text{c-m,ns-a}}\left(\frac{1}{2} - \epsilon\right) = \frac{2}{\ln 2}\epsilon^2 + o(\epsilon^2). \tag{110}$$

Together, these bounds prove the theorem. □

F Proof of Theorem 4

Throughout this proof, the graph is $\mathcal{K}_m$ and its vertex set is $[m]$. In particular, X is uniform on $[m]$. Recall from Proposition 1 that

$$R^{\text{c-m,un-a}}_{\mathcal{K}_m}\left(\frac{1}{2}-\epsilon\right) = \min_{\substack{\mathsf{P}_{U|X},\, z:\mathcal{U}\times\mathcal{Y}\to[m] \\ \Pr(X\neq z(U,Y))\le 1/2-\epsilon}} I(X;U\mid Y), \tag{111}$$

where the joint distribution is $\mathsf{P}_{UXY} = \mathsf{P}_{XY}\mathsf{P}_{U|X}$. Our goal is to lower bound this minimum as $\epsilon \downarrow 0$. For each choice of $\mathsf{P}_{U|X}$, we should choose z that minimizes the error probability $\Pr(X \neq z(U,Y))$. We first express the mutual information and this error probability in terms of $\Pr(U=u)$ and $\Pr(X=i\mid U=u)$.

Step 1: Expressing the rate and distortion using conditional probabilities. For an auxiliary variable U satisfying $U - X - Y$, write $\alpha_u = \Pr(U=u)$ and $p_i^{(u)} = \Pr(X=i\mid U=u)$ for every u with $\alpha_u > 0$. We use bold letters for vectors, such as $\boldsymbol{p}$, and write their components as p_i, where $i\in[m]$. Let Δ_m be the set of probability distributions on $[m]$, i.e.,

$$\Delta_m = \left\{\boldsymbol{p}\in\mathbb{R}^m : p_i\ge 0,\ \sum_{i=1}^m p_i = 1\right\}. \tag{112}$$

Fix u and write $\boldsymbol{p} = \boldsymbol{p}^{(u)}$ to simplify notation. Given $X=i$, Y is uniform over the $m-1$ edges containing X. Thus,

$$\begin{aligned}
&\Pr(Y=\{i,j\}\mid U=u)\\
&= p_i\Pr(Y=\{i,j\}\mid U=u, X=i) + p_j\Pr(Y=\{i,j\}\mid U=u, X=j) &(113)\\
&= p_i\Pr(Y=\{i,j\}\mid X=i) + p_j\Pr(Y=\{i,j\}\mid X=j) &(114)\\
&= \frac{p_i+p_j}{m-1} &(115)
\end{aligned}$$

and, for $p_i+p_j>0$,

$$\begin{aligned}
&\Pr(X=i\mid Y=\{i,j\}, U=u)\\
&= \frac{\Pr(X=i, Y=\{i,j\}\mid U=u)}{\Pr(Y=\{i,j\}\mid U=u)} &(116)\\
&= \frac{p_i}{p_i+p_j} &(117)
\end{aligned}$$

Given $U=u$ and $Y=\{i,j\}$, the decoder minimizes the error probability by choosing $z(u,\{i,j\}) = i$ if $p_i\ge p_j$ and j otherwise. Thus, $\Pr(X\neq z(U,Y)\mid Y=\{i,j\}, U=u) = \min\{p_i,p_j\}/(p_i+p_j)$. Averaging over Y using (115), we have

$$\begin{aligned}
&\Pr(X\neq z(U,Y)\mid U=u)\\
&= \frac{1}{m-1}\sum_{i<j}\min\{p_i,p_j\} &(118)
\end{aligned}$$

$$= \frac{1}{2} - \frac{1}{2(m-1)} \sum_{i<j} |p_i - p_j|. \tag{119}$$

Here we used the identity $\min\{a,b\} = (a+b-|a-b|)/2$ and $\sum_{i<j}(p_i+p_j) = (m-1)\sum_i p_i = m-1$. Let D_{opt} denote the error probability of this decoder. Averaging over U gives

$$D_{\mathrm{opt}} = \Pr(X \neq z(U,Y)) \tag{120}$$

$$= \frac{1}{2} - \sum_u \alpha_u \frac{1}{2(m-1)} \sum_{i<j} |p_i^{(u)} - p_j^{(u)}| \tag{121}$$

$$= \sum_u \alpha_u d(\boldsymbol{p}^{(u)}), \tag{122}$$

where we define the function

$$d(\boldsymbol{p}) \triangleq \frac{1}{2} - \frac{1}{2(m-1)} \sum_{i<j} |p_i - p_j|. \tag{123}$$

It will also be convenient to define

$$\delta(\boldsymbol{p}) \triangleq \frac{1}{2} - d(\boldsymbol{p}) = \frac{1}{2(m-1)} \sum_{i<j} |p_i - p_j|. \tag{124}$$

Next, for a fixed u and $\boldsymbol{p} = \boldsymbol{p}^{(u)}$, the conditional probabilities in (115) and (117) give

$$H(X \mid Y, U=u)$$

$$= \sum_{i<j} \Pr(Y = \{i,j\} \mid U = u) H(X \mid Y = \{i,j\}, U = u) \tag{125}$$

$$= \frac{1}{m-1} \sum_{i<j} (p_i + p_j) h_2\left(\frac{p_i}{p_i + p_j}\right). \tag{126}$$

Since $H(X \mid Y) = 1$, we obtain

$$I(X;U \mid Y) = H(X \mid Y) - H(X \mid Y, U) \tag{127}$$

$$= 1 - \sum_u \alpha_u H(X \mid Y, U = u) \tag{128}$$

$$= 1 - \sum_u \alpha_u \frac{1}{m-1} \sum_{i<j} (p_i^{(u)} + p_j^{(u)}) h_2\left(\frac{p_i^{(u)}}{p_i^{(u)} + p_j^{(u)}}\right). \tag{129}$$

We therefore define

$$F(\boldsymbol{p}) \triangleq \frac{1}{m-1} \sum_{i<j} (p_i + p_j) h_2\left(\frac{p_i}{p_i + p_j}\right), \tag{130}$$

which is equal to $H(X \mid Y)$ when $X \sim \boldsymbol{p}$. We note that F is concave in $\boldsymbol{p}$ by the concavity of entropy with respect to the underlying probability distribution. Here and above, a summand

with $p_i + p_j = 0$ is defined to be zero. With these definitions, the mutual information and distortion can be written as

$$I(X;U \mid Y) = 1 - \sum_u \alpha_u F(\boldsymbol{p}^{(u)}), \qquad D_{\text{opt}} = \frac{1}{2} - \sum_u \alpha_u \delta(\boldsymbol{p}^{(u)}). \tag{131}$$

Because P_X is uniform, the distributions and probabilities must satisfy $\sum_u \alpha_u p_i^{(u)} = 1/m$ for each i. The distortion constraint requires $\sum_u \alpha_u \delta(\boldsymbol{p}^{(u)}) \geq \epsilon$. Note that both $I(X;U \mid Y)$ and D_{opt} depend on the chosen families of distributions $\{\boldsymbol{p}^{(u)}\}_u$.

Step 2: Bounding the rate using one probability distribution. Note that both F and δ are invariant under permutations of the coordinates. Sort each $\boldsymbol{p}^{(u)}$ in non-decreasing order and continue to denote the sorted vector as $\boldsymbol{p}^{(u)}$. Let $\bar{\boldsymbol{p}} = \sum_u \alpha_u \boldsymbol{p}^{(u)}$, and $\bar{\boldsymbol{p}}$ is non-decreasing. Note that $\bar{\boldsymbol{p}}$ now need not be the uniform distribution, but we will use it to derive a lower bound on $I(X;U \mid Y)$ and an equality of D_{opt} as follows.

Due to the concavity of F, we have

$$F(\bar{\boldsymbol{p}}) \geq \sum_u \alpha_u F(\boldsymbol{p}^{(u)}), \qquad \delta(\bar{\boldsymbol{p}}) = \sum_u \alpha_u \delta(\boldsymbol{p}^{(u)}). \tag{132}$$

Thus $\delta(\bar{\boldsymbol{p}}) \geq \epsilon$, and $1 - F(\bar{\boldsymbol{p}}) \leq I(X;U \mid Y)$. Therefore,

$$R_{\mathcal{K}_m}^{\text{c-m,un-a}}\left(\frac{1}{2} - \epsilon\right) \geq 1 - \max_{\substack{\boldsymbol{p} \in \Delta_m,\, p_1 \leq \cdots \leq p_m \\ \delta(\boldsymbol{p}) \geq \epsilon}} F(\boldsymbol{p}). \tag{133}$$

Step 3: Expansion around the uniform distribution. In the following consider any non-increasingly ordered distribution $\boldsymbol{p}$. We have $|p_i - p_j| = p_j - p_i$ whenever $i < j$. Thus, from (124),

$$\begin{aligned}
\delta(\boldsymbol{p}) &= \frac{1}{2(m-1)} \sum_{i<j} (p_j - p_i) \\
&= \frac{1}{2(m-1)} \sum_{i=1}^{m} \big[(i-1) - (m-i)\big] p_i \\
&= \frac{1}{2(m-1)} \sum_{i=1}^{m} (2i - m - 1) p_i.
\end{aligned} \tag{134}$$

The second equality follows because p_i appears with a positive sign in $i-1$ pairs and a negative sign in $m-i$ pairs.

Write $\boldsymbol{p} = 1/m + \boldsymbol{v}$, where $\sum_i v_i = 0$. Since

$$\frac{p_i}{p_i + p_j} = \frac{1}{2} + \frac{v_i - v_j}{2(p_i + p_j)}, \tag{135}$$

applying (83) gives

$$(p_i + p_j)\left[1 - h_2\left(\frac{p_i}{p_i + p_j}\right)\right]$$

$$= \frac{(v_i - v_j)^2}{2(p_i + p_j)\ln 2} + O(\|\boldsymbol{v}\|_2^4). \tag{136}$$

Also, $p_i + p_j = 2/m + v_i + v_j$, so $1/(p_i + p_j) = m/2 + O(\|\boldsymbol{v}\|_2)$. Using $\sum_{i<j}(p_i + p_j) = m - 1$, we obtain

$$1 - F(\boldsymbol{p}) \tag{137}$$
$$= \frac{1}{m-1}\sum_{i<j}(p_i + p_j) - \frac{1}{m-1}\sum_{i<j}(p_i + p_j)h_2\left(\frac{p_i}{p_i + p_j}\right) \tag{138}$$
$$= \frac{1}{m-1}\sum_{i<j}(p_i + p_j)\left[1 - h_2\left(\frac{p_i}{p_i + p_j}\right)\right] \tag{139}$$
$$= \frac{1}{2(m-1)\ln 2}\sum_{i<j}\frac{(v_i - v_j)^2}{p_i + p_j} + O(\|\boldsymbol{v}\|_2^4) \tag{140}$$
$$= \frac{1}{2(m-1)\ln 2}\Big(\frac{m}{2} + O(\|\boldsymbol{v}\|_2)\Big)\sum_{i<j}(v_i - v_j)^2 + O(\|\boldsymbol{v}\|_2^4) \tag{141}$$
$$= \frac{m^2}{4(m-1)\ln 2}\|\boldsymbol{v}\|_2^2 + O(\|\boldsymbol{v}\|_2^3) \tag{142}$$

where (142) follows from

$$\sum_{i<j}(v_i - v_j)^2 = \frac{1}{2}\sum_{i,j}(v_i^2 + v_j^2 - 2v_iv_j) \tag{143}$$
$$= \frac{1}{2}\sum_{i,j}v_i^2 + \frac{1}{2}\sum_{i,j}v_j^2 - \sum_{i,j}v_iv_j \tag{144}$$
$$= m\sum_i v_i^2 - (\underbrace{\sum_i v_i}_{=0})(\underbrace{\sum_j v_j}_{=0}) \tag{145}$$
$$= m\|\boldsymbol{v}\|_2^2. \tag{146}$$

To lower bound (142) under the constraint $\delta(\boldsymbol{p}) \geq \epsilon$, we now lower bound $\|\boldsymbol{v}\|_2^2$. First,

$$2(m-1)\epsilon \leq 2(m-1)\delta(\boldsymbol{p}) \tag{147}$$
$$\overset{(134)}{=} \sum_i (2i - m - 1)p_i \tag{148}$$
$$= \sum_i (2i - m - 1)v_i, \tag{149}$$

where the last equality follows from $\sum_{i=1}^m (2i - m - 1) = 0$. Squaring and applying Cauchy–Schwarz, we obtain

$$4(m-1)^2\epsilon^2 \leq \left(\sum_i (2i - m - 1)v_i\right)^2 \tag{150}$$
$$\leq \left(\sum_i (2i - m - 1)^2\right)\left(\sum_i v_i^2\right) \tag{151}$$

$$= \frac{m(m^2-1)}{3}\|\boldsymbol{v}\|_2^2. \tag{152}$$

Therefore,

$$\|\boldsymbol{v}\|_2^2 \geq \frac{12(m-1)}{m(m+1)}\epsilon^2. \tag{153}$$

Step 4: Controlling the remainder. It remains to show that the remainder $O(\|\boldsymbol{v}\|_2^3)$ in (142) is $o(\epsilon^2)$ at a maximizer $\boldsymbol{p}_\epsilon$ in (133). Write $\boldsymbol{v}_\epsilon = \boldsymbol{p}_\epsilon - 1/m$. To bound $1 - F(\boldsymbol{p}_\epsilon)$, consider the distribution $\widetilde{\boldsymbol{p}}_\epsilon$ with components

$$\widetilde{p}_{\epsilon,i} = \begin{cases} (1-2\epsilon)/m, & i < m, \\ (1-2\epsilon)/m + 2\epsilon, & i = m. \end{cases}$$

For $0 < \epsilon < 1/2$, these components are positive, nondecreasing, and sum to one. Only the $m-1$ pairs containing the last coordinate have a nonzero difference, each equal to 2ϵ. Thus

$$\delta(\widetilde{\boldsymbol{p}}_\epsilon) = \frac{(m-1)2\epsilon}{2(m-1)} = \epsilon,$$

so $\widetilde{\boldsymbol{p}}_\epsilon$ is feasible in (133). Its squared distance from the uniform distribution is

$$\begin{aligned} \|\widetilde{\boldsymbol{p}}_\epsilon - 1/m\|_2^2 &= (m-1)\left(\frac{2\epsilon}{m}\right)^2 + \left(\frac{2(m-1)\epsilon}{m}\right)^2 \\ &= \frac{4(m-1)}{m}\epsilon^2. \end{aligned}$$

Since $\boldsymbol{p}_\epsilon$ maximizes F over the feasible distributions, $F(\boldsymbol{p}_\epsilon) \geq F(\widetilde{\boldsymbol{p}}_\epsilon)$. Applying (142) to $\widetilde{\boldsymbol{p}}_\epsilon$ therefore gives

$$\begin{aligned} 0 \leq 1 - F(\boldsymbol{p}_\epsilon) &\leq 1 - F(\widetilde{\boldsymbol{p}}_\epsilon) \\ &= \frac{m^2}{4(m-1)\ln 2}\frac{4(m-1)}{m}\epsilon^2 + O(\epsilon^3) \\ &= \frac{m}{\ln 2}\epsilon^2 + O(\epsilon^3) = O(\epsilon^2). \end{aligned} \tag{154}$$

Thus $F(\boldsymbol{p}_\epsilon) \to 1$ as $\epsilon \downarrow 0$. Since $F(\boldsymbol{p}) \leq 1$, with equality only at $\boldsymbol{p} = 1/m$, we have $\|\boldsymbol{v}_\epsilon\|_2 \to 0$.

Together with (142), we have that

$$\frac{m^2}{4(m-1)\ln 2}\|\boldsymbol{v}_\epsilon\|_2^2 + O(\|\boldsymbol{v}_\epsilon\|_2^3) = 1 - F(\boldsymbol{p}_\epsilon) = O(\epsilon^2), \tag{155}$$

which implies that $\|\boldsymbol{v}_\epsilon\|_2 = O(\epsilon)$. Therefore $O(\|\boldsymbol{v}_\epsilon\|_2^3) = O(\epsilon^3) = o(\epsilon^2)$. Combining (142) and (153) with (133) now yields

$$R_{\mathcal{K}_m}^{\text{c-m,un-a}}\left(\frac{1}{2} - \epsilon\right) \geq 1 - F(\boldsymbol{p}_\epsilon) \geq \frac{3m}{(m+1)\ln 2}\epsilon^2 + o(\epsilon^2). \tag{156}$$

This proves the theorem. □

Acknowledgment

The authors acknowledge the use of ChatGPT in developing the proofs, conducting the literature review, and improving the clarity and presentation of the manuscript. In particular, ChatGPT contributed significantly to the development of the proofs of Theorem 3 via Lemma 1, and Theorem 4. It also assisted in deriving the converse proof of Theorem 1, the proofs of Proposition 2 and Proposition 3.

References

[1] T. S. Cubitt, D. Leung, W. Matthews, and A. Winter, "Zero-error channel capacity and simulation assisted by non-local correlations," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5509–5523, 2011.

[2] O. Fawzi and P. Fermé, "Multiple-access channel coding with non-signaling correlations," *IEEE Transactions on Information Theory*, vol. 70, no. 3, pp. 1693–1719, 2024.

[3] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.

[4] S. Gel'fand and M. Pinsker, "Coding for channels with random parameters," *Probl. Contr. Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[5] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on information Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[6] Y. Yao and S. A. Jafar, "Virtual signaling of CSIT via non-signaling assistance," *IEEE Transactions on Information Theory*, vol. 72, no. 10, pp. 7240–7257, 2026.

[7] R. Zamir, "The rate loss in the Wyner–Ziv problem," *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 2073–2084, Nov. 1996.

[8] Y. Quek and P. W. Shor, "Quantum and superquantum enhancements to two-sender, two-receiver channels," *Phys. Rev. A*, vol. 95, p. 052329, May 2017. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.95.052329

[9] F. Leditzky, M. A. Alhejji, J. Levin, and G. Smith, "Playing games with multiple access channels," *Nature Communications*, vol. 11, no. 1, p. 1497, 2020.

[10] Y. Yao and S. A. Jafar, "Unlocking exponential and unbounded robust gains in Shannon capacity of classical multiple access channels with causal CSIT via quantum entanglement assistance," Jun. 2026. [Online]. Available: https://arxiv.org/pdf/2606.05412

[11] E. Mizrahi and U. Pereg, "Can quantum entanglement increase the capacity of a classical relay channel?" 2026. [Online]. Available: https://arxiv.org/abs/2610.25334

[12] Y. Yao and S. A. Jafar, "Quantum entanglement assistance improves the capacity and activates the zero-error capacity of classical channels with causal CSIT," *arXiv:2603.20416*, 2026. [Online]. Available: https://arxiv.org/abs/2603.20416

[13] B. S. Tsirel'son, "Quantum analogues of the bell inequalities. the case of two spatially separated domains," *Journal of Soviet Mathematics*, vol. 36, no. 4, pp. 557–570, 1987.

[14] C. Bennett, P. Shor, J. Smolin, and A. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem," *IEEE Transactions on Information Theory*, vol. 48, no. 10, pp. 2637–2655, 2002.

[15] W. Matthews, "A linear program for the finite block length converse of Polyanskiy–Poor–Verdú via nonsignaling codes," *IEEE Transactions on Information Theory*, vol. 58, no. 12, pp. 7036–7044, 2012.

[16] U. Pereg, C. Deppe, and H. Boche, "Quantum broadcast channels with cooperating decoders: An information-theoretic perspective on quantum repeaters," *Journal of Mathematical Physics*, vol. 62, no. 6, 2021.

[17] O. Fawzi and P. Fermé, "Broadcast channel coding: Algorithmic aspects and non-signaling assistance," *IEEE Transactions on Information Theory*, vol. 70, no. 11, pp. 7563–7580, 2024.

[18] S.-H. Nam, H.-Y. Park, J. Yun, A. Rai, S.-H. Lee, and J. Bae, "Enhancing sum capacity via quantum and no-signaling cooperation between transmitters," *arXiv:2509.08219*, Jan. 2026. [Online]. Available: https://arxiv.org/abs/2509.08219

[19] A. Seshadri, F. Leditzky, V. Siddhu, and G. Smith, "On the separation of correlation-assisted sum capacities of multiple access channels," *IEEE Transactions on Information Theory*, vol. 69, no. 9, pp. 5805–5844, 2023.

[20] J. Yun, S.-H. Nam, H.-Y. Park, A. Rai, S.-H. Lee, and J. Bae, "Nonlocal and quantum advantages in network coding for multiple access channels," *Phys. Rev. A*, vol. 113, p. 042624, Apr 2026. [Online]. Available: https://link.aps.org/doi/10.1103/b4q8-bzcq

[21] Y. Yao and S. A. Jafar, "Can non-signaling assistance increase the degrees of freedom of a wireless network?" *IEEE Transactions on Information Theory*, vol. 72, no. 2, pp. 844–864, 2026.

[22] ——, "The capacity region of the multiple access channel with non-signaling assistance," 2026. [Online]. Available: https://arxiv.org/abs/2608.13860

[23] S. Beigi and A. Gohari, "On dimension bounds for auxiliary quantum systems," *IEEE Transactions on Information Theory*, vol. 60, no. 1, pp. 368–387, Jan. 2014.

[24] C. H. Bennett, P. Hayden, D. W. Leung, P. W. Shor, and A. Winter, "Remote preparation of quantum states," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 56–74, 2005.

[25] Z. Baghali Khanian and A. Winter, "A rate-distortion perspective on quantum state redistribution," *IEEE Transactions on Information Theory*, vol. 72, no. 4, pp. 2307–2318, Apr. 2026.

[26] H. Buhrman, R. Cleve, S. Massar, and R. De Wolf, "Nonlocality and communication complexity," *Reviews of modern physics*, vol. 82, no. 1, p. 665, 2010.

[27] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski, "Information causality as a physical principle," *Nature*, vol. 461, no. 7267, pp. 1101–1104, 2009.

[28] A. Tavakoli, J. Pauwels, E. Woodhead, and S. Pironio, "Correlations in entanglement-assisted prepare-and-measure scenarios," *PRX Quantum*, vol. 2, p. 040357, 2021.

[29] J. Pauwels, S. Pironio, E. Zambrini Cruzeiro, and A. Tavakoli, "Adaptive advantage in entanglement-assisted communications," *Physical Review Letters*, vol. 129, p. 120504, 2022.

[30] J. Nogueira, C. Vieira, L. E. A. Porto, L. Pollyceno, R. Rabelo, and O. Gühne, "Non-signaling assistance in prepare-and-measure scenarios with classical communication," arXiv:2607.01193, 2026. [Online]. Available: https://arxiv.org/abs/2607.01193

[31] A. Peres and D. R. Terno, "Quantum information and relativity theory," *Reviews of Modern Physics*, vol. 76, no. 1, p. 93, 2004.

[32] T. Linder, R. Zamir, and K. Zeger, "On source coding with side-information-dependent distortion measures," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2697–2704, 2000.

[33] M. M. Wilde, N. Datta, M.-H. Hsieh, and A. Winter, "Quantum rate distortion coding with auxiliary resources," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6755–6773, 2013.

[34] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[35] M. M. Wilde, *Quantum Information Theory*, 2nd ed. Cambridge University Press, 2017.

[36] T. M. Cover and J. A. Thomas, *Elements of Information Theory 2nd Edition*. USA: Wiley-Interscience, 2006.