A Rate-Distortion Perspective on Quantum State Redistribution
Zahra Baghali Khanian and Andreas Winter

Abstract

We consider a rate-distortion version of the quantum state redistribution task, where the error of the decoded state is judged via an additive distortion measure; it thus constitutes a quantum generalisation of the classical Wyner-Ziv problem. The quantum source is described by a tripartite pure state shared between Alice (A, encoder), Bob (B, decoder) and a reference (R). Both Alice and Bob are required to output a system (\(\tilde{A}\) and \(\tilde{B}\), respectively), and the distortion measure is encoded in an observable on \(\tilde{A}BR\).

It includes as special cases most quantum rate-distortion problems considered in the past, and in particular quantum data compression with the fidelity measured per copy; furthermore, it generalises the well-known state merging and quantum state redistribution tasks for a pure state source, with per-copy fidelity, and a variant recently considered by us, where the source is an ensemble of pure states [1], [2].

We derive a single-letter formula for the rate-distortion function of compression schemes assisted by free entanglement. A peculiarity of the formula is that in general it requires optimisation over an unbounded auxiliary register, so the rate-distortion function is not readily computable from our result, and there is a continuity issue at zero distortion. However, we show how to overcome these difficulties in certain situations.

I. Introduction and Setting

Source coding is for information theory as much a practical matter, as it is a fundamental paradigm to establish the amount of information in given data. Shannon’s original model of block coding [3], giving rise to the entropy, was subsequently generalized to situations with side information at the decoder [4], which gives an operational interpretation for the conditional entropy. In another direction, by considering more flexible error criteria, “distortions”, instead of the rigid block error probability [5], [6], leads to a rate-distortion tradeoff characterized by the mutual information. Many other variations of source compression have been conceived, but to conclude our rapid review of classical source coding, we highlight only one more, the Wyner-Ziv problem of rate-distortion of a source with correlated side information at the decoder [7].

Quantum Shannon theory has sought to emulate this approach by “quantizing” the preceding source coding problems, with the aim to gain both a fundamental and operationally grounded understanding of quantum information. The first and most important among these is Schumacher’s quantum source model and compression problem, whose optimal rate is given by the von Neumann entropy [8], [9], [10], [11], [12]. Quantum compression with side information turned out to have a surprisingly rich structure, compared to the classical case: [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22]. On the other hand, rate-distortion theory has received scarce quantum attention over the years, and the results are not as complete as the classical theory [23], [24], [25], [26].

Z. B. Khanian is with TUM—Technische Universität München, 85748 Garching, Germany. Email: zkhanian@gmail.com.
A. Winter is with ICREA—Institució Catalana de Recerca i Estudis Avançats, Pg. Lluis Companys, 23, 08010 Barcelona, Spain and Física Teòrica: Grup d’Informació Quàntica (GIQ), Departament de Física, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain, Email: andreas.winter@uab.cat.
A part of this work, focusing on the source coding aspects, has been presented at ISIT 2020 [1].
Here we present and solve a quantum version of the Wyner-Ziv problem, with unlimited entanglement, for a distributed source and relative to a convex, additive distortion measure. Concretely, we consider a pure state source $|\psi\rangle^{ABR}$, with $A$ Alice’s register, $B$ Bob’s and $R$ is a passive reference. Furthermore, let $\Delta : S(\vec{ABR}) \rightarrow \mathbb{R}$ be a convex continuous real function on the quantum states of the tripartite system $\vec{ABR}$. For block length $n$, the source is the i.i.d. extension $\psi^{A^nB^nR^n} = (|\psi^{ABR}\rangle)^{\otimes n}$, and the distortion measure is extended to $n$ systems as

$$\Delta^{(n)}(\rho^{\vec{A}^n\vec{B}^nR^n}) = \frac{1}{n} \sum_{i=1}^{n} \Delta(\rho^{\vec{A}_i\vec{B}_iR_i}),$$

where on the right hand side $\rho^{\vec{A}_i\vec{B}_iR_i} = \text{Tr}_{-i} \rho^{\vec{A}^n\vec{B}^nR^n}$ is the reduced state on the $i$-th systems $\vec{A}_i\vec{B}_iR_i$ (partial trace over all other systems).

An important special case, considered in earlier approaches to quantum rate-distortion [24], [25], [26], is that $\Delta(\rho) = \text{Tr} \rho \Delta$ for a selfadjoint distortion observable $\Delta$ and $\Delta^{(n)}(\rho) = \text{Tr} \rho \Delta^{(n)}$, where

$$\Delta^{(n)}(\rho) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}^{\otimes i-1} \otimes \Delta \otimes \mathbb{1}^{\otimes n-i}.$$  

This is enough to describe classical rate-distortion functions [5], [6], but the theory goes through in the generality of the above convex additive functions.

With these data, an entanglement-assisted compression scheme of block length $n$ and distortion $D \in \mathbb{R}$ consists of an entangled state $\Phi^{A_0B_0}$, w.l.o.g. pure, and a pair of CPTP maps $\mathcal{E} : A^nA_0 \rightarrow \vec{A}^nM$, $\mathcal{D} : MB_0B^n \rightarrow \vec{B}^n$ (see Fig. 1), such that the output state

$$\xi^{\vec{A}^n\vec{B}^nR^n} = (\mathcal{D} \otimes \text{id}_{\vec{A}^nR^n}) \circ (\mathcal{E} \otimes \text{id}_{B_0B^nR^n})(\psi^{\otimes n} \otimes \Phi^{A_0B_0})$$

satisfies

$$\Delta^{(n)}(\xi^{\vec{A}^n\vec{B}^nR^n}) = \frac{1}{n} \sum_{i=1}^{n} \Delta(\xi^{\vec{A}_i\vec{B}_iR_i}) \leq D. \quad (2)$$

The rate of the code is simply $\frac{1}{n} \log |M|$, i.e. the number of qubits sent per source system. An achievable asymptotic rate for distortion $D$ is a number $R$ such that there exists a sequence of codes with rates converging to $R$ and distortion converging to $D$. For a given $D$, the rate-distortion function $Q_{\text{ea}}(D)$ is the minimum achievable rate for distortion $D$ (the subscript ‘ea’ reminds that the codes are assisted by entanglement).

We stress that $A$, $B$ and $R$ are arbitrary quantum systems here, and so are $\vec{A}$ and $\vec{B}$: the latter need not bear any relation to $A$ and $B$, their names are chosen entirely as a reminder that ‘A’$'$s belong to Alice (compressor/sender) and ‘B’$'$s belong to Bob (receiver/decoder).

![Figure 1. Communication diagramme of the entanglement-assisted rate-distortion state redistribution task: $R$ is a passive reference, $A$ and $\vec{A}$ are Alice’s input and output systems, and $B$ and $\vec{B}$ are Bob’s input and output systems, respectively.](image-url)

Figure 1. Communication diagramme of the entanglement-assisted rate-distortion state redistribution task: $R$ is a passive reference, $A$ and $\vec{A}$ are Alice’s input and output systems, and $B$ and $\vec{B}$ are Bob’s input and output systems, respectively.
Note that for small enough $D$, there will be no codes with distortion $D$, and then $Q_{\text{eq}}(D) = +\infty$ by convention. Once codes exist, $Q_{\text{eq}}(D)$ is a non-negative real number, and for sufficiently large $D$, for example $D \geq \max_{\rho} \Delta(\rho)$, $Q_{\text{eq}}(D) = 0$ because every pair of maps is an eligible code.

**Remark 1**: In contrast to other previous work [26], which imposed a distinction between data to be compressed and side information, we think that our present model is both simpler and more natural, by applying a global distortion measure jointly to Alice’s and Bob’s parts of the output, as well as to the reference.

In the rest of the paper, we present our main result in Section II, which is a single-letter characterization of the rate-distortion function, for general sources and arbitrary distortion measures: we first use QSR to build a protocol giving us an achievable rate, and then show that it is essentially optimal. Then, in Section III, we discuss a number of special cases of the considered scenario, showing how the rate-distortion setting generalizes all sorts of conventional quantum source coding problems, some of which have appeared in the previous literature. The original source coding problems are recovered, after a fashion, in the limit of zero (per-copy) error. We conclude with a discussion of the result and open problems in Section IV.

**Notation and basic facts.** Quantum systems are associated with (in this paper: finite dimensional) Hilbert spaces $A$, $R$, . . . , whose dimensions are denoted by $|A|$, $|R|$, . . . , respectively. We identify states on a system $A$ with their density operators, $S(A)$, with is the set of all positive semidefinite matrices with unit trace. We use the notation $\phi = |\phi\rangle\langle\phi|$ as the density operator of the pure state vector $|\phi\rangle \in A$.

The von Neumann entropy is $S(\rho) = -\text{Tr} \rho \log \rho$, log by default being the binary logarithm. The conditional entropy and the conditional mutual information, $S(A|B)_\rho$ and $I(A : B|C)_\rho$, respectively, are defined in the same way as their classical counterparts:

$$S(A|B)_\rho = S(AB)_\rho - S(B)_\rho,$$

and

$$I(A : B|C)_\rho = S(A|C)_\rho - S(A|BC)_\rho = S(AC)_\rho + S(BC)_\rho - S(ABC)_\rho - S(C)_\rho.$$

The fidelity between two states $\rho$ and $\xi$ is defined as $F(\zeta, \xi) = \|\sqrt{\zeta}\sqrt{\xi}\|_1 = \text{Tr} \sqrt{\zeta^\dagger \xi}$, with the trace norm $\|X\|_1 = \text{Tr} |X| = \text{Tr} \sqrt{X^\dagger X}$. If one of the two states is pure, $F(\zeta, \xi)^2 = \text{Tr} \zeta \xi$. In general it relates to the trace distance in the following well-known way [27]:

$$1 - F(\zeta, \xi) \leq \frac{1}{2} \|\zeta - \xi\|_1 \leq \sqrt{1 - F(\zeta, \xi)^2}. \quad (3)$$

As we consider information theoretic limits, we have occasion to refer to many isomorphic copies of a single system, say $A$, which are always referred to by the same capital letter with a running index, i.e. $A_1$, $A_2$, . . . , $A_n$, . . . ; a block (tensor product) of the first $n$ of these systems is written $A^n = A_1 A_2 \cdots A_n = A_1 \otimes \cdots \otimes A_n$. More generally for a set $I \subset \mathbb{N}$ of indices, $A_I = \bigotimes_{i \in I} A_i$. We use the combinatorial shorthand $[n] = \{1, 2, \ldots , n\}$, so that $A^n = A_{[n]}$.

**II. SINGLE-LETTER CHARACTERIZATION OF THE RATE-DISTORTION FUNCTION**

In this section, we solve the quantum rate-distortion problem introduced above. First, we construct a protocol for a certain achievable rate, coming directly from quantum state redistribution (QSR); after that, we show the converse. QSR is a quantum compression protocol where both encoder and decoder have access to side information. We introduce this protocol more on subsection III-C.
A. An achievable rate from QSR

Assume that we have two CPTP maps $\mathcal{E}_0 : A \rightarrow \widetilde{A}Z$ and $\mathcal{D}_0 : BZ \rightarrow \widetilde{B}$ such that

$$\xi^{\widetilde{A}BR} = (\mathcal{D}_0 \otimes \text{id}_{\widetilde{A}R}) \circ (\mathcal{E}_0 \otimes \text{id}_{BR}) \psi^{ABR},$$

it holds $\Delta(\xi) \leq D$. Then, $R = \frac{1}{2} I(Z : R|B)_{\varphi}$ is an achievable asymptotic rate for distortion $D$, where $\varphi^{\widetilde{A}ZBR} = (\mathcal{E}_0 \otimes \text{id}_{BR}) \psi^{ABR}$ is the state after the action of $\mathcal{E}_0$.

**Proof.** Purify $\mathcal{E}_0$ to a Stinespring isometry $U : A \rightarrow \widetilde{A}ZW$ [28], so after applying it to the source we have the pure state $|\varphi\rangle^{\widetilde{A}WZBR} = (U \otimes \text{id}_{BR}) |\psi\rangle^{ABR}$. On block length $n$, use QSR, assisted by suitable entanglement, as a subroutine, to send $Z^n$ from Alice to Bob, with $\widetilde{A}^nW^n$ as Alice’s side information and $B^n$ as Bob’s. The block trace distance error of the QSR protocol goes to 0 as $n \rightarrow \infty$, so we get distortion $\leq D + o(1)$, using the continuity of $\Delta$. The rate, which is due to QSR, is $\frac{1}{2} I(Z : R|B)_{\varphi}$. ■

This coding theorem motivates the introduction of the following single-letter function,

$$Q'(D) := \inf_{\mathcal{E}_0, \mathcal{D}_0} \frac{1}{2} I(Z : R|B)_{\varphi}$$

where $\xi$ is defined in Eq. (4), and the conditional mutual information is with respect to the state $\varphi^{\widetilde{A}ZBR} = (\mathcal{E}_0 \otimes \text{id}_{BR}) \psi^{ABR}$. We have thus proved that

$$Q_{\text{ca}}(D) \leq \lim_{D' \rightarrow D_+} Q'(D').$$

Since $Q'$ is monotonically non-increasing with $D$, the latter limit from the right is also a supremum, equal to $\sup_{D' \geq D} Q'(D')$.

Before we go on, we analyze first some mathematical properties of the new function. Note that a major difficulty, both practically and for the theoretical development, is the unbounded nature of the auxiliary system $Z$. Define

$$D_0 := \inf D \text{ s.t. } Q'(D) < +\infty$$

$$= \inf D \text{ s.t. } \exists \mathcal{E}_0, \mathcal{D}_0 \Delta(\xi) \leq D.$$

By definition, $Q'(D) = +\infty$ for all $D < D_0$ and $Q'(D)$ is finite for all $D > D_0$. Because of the dimensionality issue, $Q'(D_0)$ may or may not be finite.

**Lemma 2:** On $[D_0, \infty)$, $Q'$ is a monotonically non-increasing, convex function of $D$. Consequently, on the open interval $(D_0, \infty)$ it is also continuous.

**Proof.** The monotonicity was already remarked to follow from the definition. For the convexity, we verify Jensen’s inequality, that is we start with maps $\mathcal{E}_1, \mathcal{D}_1$ eligible for distortion $D_1$, and $\mathcal{E}_2, \mathcal{D}_2$ eligible for distortion $D_2$, and $0 \leq p \leq 1$. By embedding into larger Hilbert spaces if necessary, we can w.l.o.g. assume that the maps act on the same systems for $i = 1, 2$. We define the following two maps:

$$\mathcal{E}(\rho) := p \mathcal{E}_1(\rho) \otimes |1\rangle\langle 1|^{Z'} + (1-p) \mathcal{E}_2(\rho) \otimes |2\rangle\langle 2|^{Z'},$$

$$\mathcal{D}(\rho) := \mathcal{D}_1(|1\rangle\langle 1|^{Z'}) + \mathcal{D}_2(|2\rangle\langle 2|^{Z'}).$$

They evidently realise the output state $\xi = p\xi_1 + (1-p)\xi_2$, hence by convexity the distortion is bounded as $\Delta(\xi) \leq p \Delta(\xi_1) + (1-p) \Delta(\xi_2) \leq pD_1 + (1-p)D_2$. Thus,

$$Q'(D) \leq \frac{1}{2} I(ZZ' : R|B)_{\xi} = \frac{p}{2} I(Z : R|B)_{\xi_1} + \frac{(1-p)}{2} I(Z : R|B)_{\xi_2},$$

and taking the infimum over maps $\mathcal{E}_i, \mathcal{D}_i$ shows convexity.
The continuity statement follows from a mathematical folklore fact, stating that any real-valued function that is
convex on an interval, is continuous on the interior of the interval, cf. [29, Prop. 2.17].

This lemma shows that the only possible discontinuity of $Q'$ is at $D_0$, and so we are motivated to define its
right-continuous extension, which differs from $Q'$ only possibly at $D_0$:

$$Q'(D) := \sup_{D' > D} Q'(D') = \begin{cases} +\infty & \text{if } D < D_0, \\ \sup_{D' > D_0} Q'(D') & \text{if } D = D_0, \\ Q'(D) & \text{if } D > D_0. \end{cases} \tag{8}$$

Our achievability result from the beginning of the present section can now be expressed more concisely as follows.

**Proposition 3:** For any source $\psi^{ABR}$ and convex distortion measure $\Delta$, it holds for all distortion values $D$,

$$Q_{ca}(D) \leq \overline{Q}(D).$$

(Note that this is trivially true for $D < D_0$, as then the right hand side, and as we shall see also the left hand side,
is $+\infty$.)

**B. Converse and main result**

Now we show the opposite inequality, $Q_{ca}(D) \geq \overline{Q}(D)$. Together with the achievability, this yields:

**Theorem 4:** For any source state $\psi^{ABR}$ and convex distortion measure $\Delta$, it holds for all distortion values $D$,

$$Q_{ca}(D) = \overline{Q}(D).$$

**Proof.** Consider a block length $n$ code of distortion $\Delta^{(n)}(\xi^{A^nB^nR^n}) \leq D + \delta$. The number of qubits, $\log |M|$, can be lower bounded as follows, with respect to the encoded state $\sigma^{MB_0\tilde{A}B^nR^n} = (\mathcal{E} \otimes \text{id}_{B_0B^nR^n}) (\psi^{\otimes n} \otimes \Phi^{R^nB_0})$:

$$2 \log |M| \geq 2S(M)$$

$$\geq I(M : R^n|B^nB_0)$$

$$= I(MB_0 : R^n|B^n) - I(B_0 : R^n|B^n)_{\rightarrow 0}$$

$$= I(Z : R^n|B^n) \quad [\text{with } Z \equiv MB_0]$$

$$= \sum_{i=1}^n I(Z_i : R^n_{[n]} ; R_{[n]}^{i-1}) + \sum_{i=1}^n I(R_{[n]}^{i-1} ; R_iB_i)_{\rightarrow 0}$$

$$= \sum_{i=1}^n I(R_{[n]}^{i-1} ; R_iB_i) \quad [\text{with } Z_i \equiv ZB_{[n]}^{i-1}]$$

where in the first two inequalities we use standard entropy inequalities; the equation in the third line is due to the
chain rule, and the second conditional information is 0 because $B_0$ is independent of $B^nR^n$; the fourth line introduces
a new register $Z$, noting that the encoding together with the entangled state defines a CPTP map $\mathcal{E}_0 : A^n \rightarrow \tilde{A}^{n}Z$, via $\mathcal{E}_0(\rho) = (\mathcal{E} \otimes \text{id}_{B_0}) (\rho \otimes \Phi^{A_0B_0})$; in the fifth we use the chain rule iteratively, and in the second term we introduce
each summand is 0 because for all $i$, $R_{[n]}^{i-1}$ is independent of $R_iB_i$; in the sixth line we use again the chain
rule for all $i$, and in the last line strong subadditivity (data processing).
For the $i$-th copy $\psi^{A_i,B_i;R_i}$, now define maps $\mathcal{E}_i : A_i \to \overline{A}_i Z_i$ and $\mathcal{D}_i : B_i Z_i \to \overline{B}_i$, as follows:

$\mathcal{E}_i$: Alice tensors her system $A_i$ with a dummy state $\psi^{|n\rangle}_i$ and with $\Phi^{A_0 B_0}$ (note that all systems are in her possession). Then she applies $\mathcal{E} : A^n A_0 \to \overline{A}_n M$, and sends $Z_i := M B_0 |n\rangle_i$ to Bob, while keeping $\overline{A}_i$.

Everything else, i.e. $R_0 |n\rangle_i \overline{A}_0 |n\rangle_i$, is trashed.

$\mathcal{D}_i$: Bob applies $\mathcal{D}$ to $Z_i B_i = M B_0 B^n$ and keeps $\overline{B}_i$, trashing the rest $\overline{B}_0 |n\rangle_i$.

By definition, the output state

$$\zeta^{\overline{A},\overline{B},R_i} = (\mathcal{D}_i \otimes \text{id}_{\overline{A}_i R_i}) \circ (\mathcal{E}_i \otimes \text{id}_{B_i R_i}) \psi^{A_i,B_i;R_i}$$

equals $\zeta^{\overline{A},\overline{B},R_i}$, and with $D_i := \Delta \left( \zeta^{\overline{A},\overline{B},R_i} \right)$ we have

$$D + \delta \geq \Delta^{(n)} \left( \zeta^{\overline{A}^{\otimes n},\overline{B}^{\otimes n},R^n} \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \Delta \left( \zeta^{\overline{A}_i,\overline{B}_i,R_i} \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \Delta \left( \zeta^{\overline{A}_i,\overline{B}_i,R_i} \right) = \frac{1}{n} \sum D_i.$$  

Thus, we obtain, with respect to the states $(\mathcal{E}_i \otimes \text{id}_{B_i R_i}) \psi^{A_i,B_i;R_i}$ for $i = 1, \ldots, n$,

$$\frac{1}{n} \log |M| \geq \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2} I(Z_i : R_i | B_i)$$

$$\geq \frac{1}{n} \sum_{i=1}^{n} Q'(D_i)$$

$$\geq Q' \left( \frac{1}{n} \sum_{i=1}^{n} D_i \right)$$

$$\geq Q'(D + \delta),$$

(10)

continuing from Eq. (9), then by definition of $Q'(D)$ since the pair $(\mathcal{E}, \mathcal{D})$ results in distortion $D_i$, in the next line by convexity and finally by monotonicity of $Q'$ (Lemma 2).

Since this has to hold for all $\delta > 0$ and in the limit $n \to \infty$, the claim follows. $\blacksquare$

**Remark 5:** The real problem with Theorem 4, and the formula (5), is that while the rate-distortion function on the face of it is single-letter, it is not still necessarily computable, because of the infimum over CPTP maps $\mathcal{E}_0 : A \to \overline{A} Z$ and $\mathcal{D}_0 : B Z \to \overline{B}$, with – crucially – unbounded quantum register $Z$.

With a bounded $|Z|$, the domain of optimization would become compact, and this would not only make $Q'(D)$ computable (in the sense that it can be approximated to arbitrary degree), and in fact a minimum, hence itself a continuous function, but also $\mathcal{D}_0$ would be computable, and we would get $\overline{Q} \equiv Q'$.

Without this information, and we have no evidence of finiteness or required infinity either way, in general, the rate-distortion function is only a formal expression, and shares the issue of computability or approximability with an astonishing number of other, similar capacity formulas in quantum Shannon theory: the entangling power of a bipartite unitary [30], the symmetric side-channel assisted quantum capacity [31] and the analogous private capacity [32], the squashed entanglement [33] and other entanglement measures such as the so-called CEMI [34], and the quantum information bottleneck function [35].

In the rest of the paper, we will show how this theorem permits a new view of various quantum source coding problems that have been considered in the literature. In all these cases, this rests on writing the pure state or the ensemble fidelity (per-copy) of a coding scheme as a distortion in the above sense.
III. ENTANGLEMENT-ASSISTED SOURCE CODING EMERGING IN THE LIMIT OF UNIT PER-COPY FIDELITY

In this section we are going to specialise the above general theory to the traditional setting of quantum source compression, where the distortion measure is the infidelity of decoding, i.e. one minus the fidelity (squared) between the decoded state and an ideal state.

A. Schumacher’s quantum data compression with an entanglement fidelity criterion

In [8], [9], quantum source coding is described with a pure state $|\psi\rangle^{AR}$ for the source, so that $B$ is trivial (one-dimensional) and so is $\tilde{A}$, while $\tilde{B} = \tilde{A} \approx A$. The use of the block fidelity as success criterion of the code corresponds to the distortion measure $1 - F(\xi, \psi)$, which would be eligible, being convex and continuous in the state; here, we will however consider $\Delta(\xi) = 1 - F(\xi, \psi)^2$, because it corresponds to a distortion operator, namely $\Delta = 1 - \psi\tilde{R}$, which will suit us better in the later developments. Note that for regular source coding, this is not an important change, since there anyway the focus is on $F(\xi, \psi) \approx 1$; to be precise, for the $n$-fold i.i.d. repetition $\psi^{\otimes n}$ and the $n$-system output state $\xi^{\tilde{A}^{\otimes n}}$, we demand $F(\xi, \psi^{\otimes n}) \approx 1$ [8], [9]. Under the present rate-distortion perspective, however, we consider the weaker (implied) criterion $\Delta(n)(\xi) = \text{Tr} \Delta^{(n)}(\xi) \approx 0$. Of course, rate-distortion theory makes good sense of all values of $D$, but we shall focus on the small ones to preserve the relation with source coding. Schumacher’s date compression implies that for all $D > 0$, $Q_{ca}(D) \leq Q_{ca}(0) \leq S(A)$. The latter bound is actually an equality, as it can be seen as follows (cf. [11]). Consider a $D > 0$, and consider pairs of CPTP maps $\mathcal{E}_0$ and $\mathcal{D}_0$ eligible for $\overline{Q}(D)$, then Theorem 4 implies the following converse bound considering per-copy fidelity:

$$\overline{Q}(D) = \inf \frac{1}{2} I(Z : R)_{\varphi}$$
$$\geq \inf \frac{1}{2} I(\tilde{A} : R)_{\xi}$$
$$\geq \frac{1}{2} I(A : R)_{\varphi} - 2\sqrt{D} \log |R| - g(\sqrt{D})$$

(11)

where the first line is by definition, the second invoking data processing, and the last one by first observing that by Eq. (3), $\frac{1}{2} \|\xi - \psi\|_1 \leq \sqrt{D}$ and then using the Alicki-Fannes continuity bound for the conditional entropy [36] in the form given in [37]: for two states with $\frac{1}{2} \| \rho^{UV} - \sigma^{UV} \|_1 \leq \delta$,

$$|S(U|V)_{\rho} - S(U|V)_{\sigma}| \leq 2\delta \log |U| + g(\delta),$$

(12)

with $g(\delta) = (1 + \delta) \log(1 + \delta) - \delta \log \delta$.

Thus, from the bound $Q_{ca}(0) \leq S(A)_{\varphi}$ and Eq. (11), letting $D \to 0$, we get $Q_{ca}(0) = \overline{Q}(0) = \frac{1}{2} I(A : R)_{\varphi} = S(A)_{\varphi}$.

B. Schumacher’s quantum data compression for an ensemble source

Schumacher [8] also introduced another model of the quantum source, as an ensemble $\{p(x), |\psi_x\rangle^A\}$, where $x$ ranges over a discrete set. One can of course describe this kind of source by a cq-state $\omega^{AR} = \sum_x p(x)|\psi_x\rangle^A \otimes |x\rangle|x\rangle^R$, but the rate-distortion setting allows to do it differently: not by changing the source state, which remains the pure state $|\psi\rangle^{AR} = \sum_x \sqrt{p(x)}|\psi_x\rangle^A \otimes |x\rangle|x\rangle^R$, but instead with a different distortion operator:

$$\Delta = 1 - \sum_x |\psi_x\rangle|\psi_x\rangle^\tilde{A} \otimes |x\rangle|x\rangle^R.$$

(13)
For this ensemble source, the output state of the composite system is
\[ \hat{\xi}^{R^n} = (D \otimes \text{id}_{R^n}) \circ (E \otimes \text{id}_{B_0 R^n})(\psi^{A B R^n} \otimes \Phi^{A_0 B_0}) \]
and the output state of the \( i \)-th system is
\[ \hat{\xi}^{A_i R_i} = \text{Tr}_{[n \backslash i]}(\hat{\xi}^{R^n}) = \sum_{x_i, x'_i} \sqrt{p(x_i)} p(x'_i) \hat{\xi}^{A_i R'_i} \otimes |x_i x'_i| X_i, \]
where \( \hat{\xi}^{A_i R'_i} = \sum_{i \in [n]} p(x_i) \hat{\xi}^{A_i R'_i} |x_i x'_i| x_i x'_i \). Measuring the distortion with the distortion operator of Eq. (13) is equivalent to measuring per-copy fidelity for the output state \( \hat{\xi}^{A_i R_i} \):
\[ \text{Tr}(\hat{\xi}^{A_i R_i} \Delta) = 1 - \sum_{x_i} p(x_i) \text{Tr}(\hat{\xi}^{A_i R'_i} \psi^{A_i R'_i}) = 1 - \sum_{x_i} p(x_i) F^2(\hat{\xi}^{A_i R'_i}, \psi^{A_i R'_i}). \]

The optimal entanglement-assisted compression rate for this ensemble source is found in [38] to be \( \frac{1}{2} (S(A)_\omega + S(A|Y)_\omega) \) where the decodability criterion is block fidelity. This rate is with respect to the following modified source defined as
\[ \omega^{A Y R} := \sum_x p(x) |\psi_x\rangle |y(x)\rangle^A \otimes |y(x)\rangle^Y \otimes |x\rangle^R, \quad (14) \]
where the register \( Y \) stores the corresponding orthogonal subspaces for the signals \( \{\psi_x^\omega\} \). For example, if signals \( \psi^A_1 \) and \( \psi^A_2 \) are orthogonal to signals \( \psi^A_3 \) and \( \psi^A_4 \), the variables \( y(1) = y(2) \) and \( y(3) = y(4) \) denote two underlying orthogonal subspaces. Since block fidelity implies per-copy fidelity, the rate \( \frac{1}{2} (S(A)_\omega + S(A|Y)_\omega) \) is achievable with per-copy fidelity as well. Notice that
\[ \frac{1}{2} (S(A)_\omega + S(A|Y)_\omega) = \frac{1}{2} I(A : R), \]
where the mutual information is with respect to the modified pure state \( |\psi\rangle^{A Y R} := \sum_x \sqrt{p(x)} |\psi_x\rangle^A \otimes |y(x)\rangle^A \otimes |x\rangle^R \).

The converse bound for the above rate considering per-copy fidelity is obtained in Corollary 11. Note that the lower bound in Eq. (11) is not valid here since in the last line we use the fact that the decoded state on systems \( \tilde{A}R \) is very close, in trace distance, to the pure source on systems \( AR \). However, for the ensemble source, the decoded state is close to the original ensemble and not the purification of the ensemble, therefore, the lower bound on Eq. (11) does not hold in this case.

C. Quantum state redistribution

To recover QSR itself, we replace \( A \) by the bipartite system \( AC, \tilde{A} = \tilde{C} \simeq C \) and \( \tilde{B} = \tilde{A}B \simeq AB \). The source is given by the pure state \( |\psi\rangle^{AC BR} \), where \( A \) and \( C \) are initially with Alice and \( B \) with Bob, and at the end \( A \) changes hands from Alice to Bob, while \( C \) and \( B \) remain in place. The distortion operator is \( \Delta = 1 - \psi^{AC BR} \), so that the distortion per letter is \( \Delta(\xi) = \text{Tr} \xi \Delta = 1 - \text{Tr} \xi \psi = 1 - F(\xi, \psi)^2 \).

Note that for a single system the criterion is the familiar fidelity (up to the square, that some authors put and others not), but for block length \( n \) the usual criterion considered for QSR [18], [19], [20] is not the per-copy but the block fidelity, which is a stronger requirement. Nevertheless, the well-known coding theorems for QSR [18], [19], [20] imply that \( \frac{1}{2} I(A : R|B) \) is an achievable rate for any distortion \( D \geq 0 \), since block fidelity implies per-copy fidelity, hence \( Q_{\text{ca}}(D) \leq \frac{1}{2} I(A : R|B) = \frac{1}{2} I(A : R|C) \) for all \( D \geq 0 \).
On the other hand, for $D \geq 0$, Theorem 4 implies the converse bound $Q_{ea}(D) \geq Q(D)$ considering per-copy fidelity. Namely, for $D \geq 0$ and pairs of CPTP maps $\mathcal{E}_0$ and $\mathcal{D}_0$ eligible for $Q(D)$, we obtain
\[ Q(D) \geq \inf \frac{1}{2} I(Z : R|B), \]
\[ \geq \inf \frac{1}{2} I(\hat{A} : R|B), \]
\[ \geq \frac{1}{2} I(A : R|B), \]
where the first line is by definition, the second invoking data processing, and the last one by first observing that by Eq. (3), $\frac{1}{2} \|\xi^{ACBR} - \psi^{ACBR}\|_1 \leq \sqrt{D}$ and then using the Alicki-Fannes continuity bound for the conditional entropy [36] in the form of Eq. (12) given in [37]. This lower bound together with the upper bound discussed above imply that in the limit of $D \to 0$, $Q_{ea}(0) = Q(0)$ converges to $\frac{1}{2} I(A : R|B)$. 

An important special case of QSR is state merging, which is recovered for trivial (one-dimensional) side-information system $C$, that is the source is given by the pure state $|\psi\rangle^{ABR}$, and $\hat{A} = \emptyset$ and $\hat{B} = \hat{A}\hat{B} = AB$ are respectively Alice and Bob’s decoded systems. As discussed above, we can conclude that for per-copy fidelity (distortion operator $\Delta = I - \psi^{ABR}$), the optimal rate is $\frac{1}{2} I(A : R|B)$. 

### D. Ensemble state quantum redistribution

Analogous to the discussion of Schumacher’s quantum source coding, if we have an ensemble source $\left\{ p(x), |\psi_x\rangle^{ACBR} \right\}$ denoted by the qqqc state $\omega^{ACBRX} = \sum_x p(x) |\psi_x\rangle^{ACBR} \otimes |x\rangle X$, we can define the pure state source $|\psi\rangle^{ACBR} := \sum_x \sqrt{p(x)} |\psi_x\rangle^{ACBR} |x\rangle X$ and a distortion operator such that measuring the distortion for the pure source $|\psi\rangle^{ACBR}$ is equivalent to measuring the ensemble infidelity for the source $\omega^{ACBRX}$. As before replace $A$ by the bipartite system $AC$, $\hat{A} = \hat{C} = C$, $\hat{B} = \hat{A}\hat{B} = AB$, and $R = R'X$. Then, the output state of the composite system is
\[ \xi^{\hat{A}\hat{C}\hat{B}R} = (\mathcal{D} \otimes \text{id}_{\hat{C}\hat{B}R}) \circ (\mathcal{E} \otimes \text{id}_{B\hat{B}R})((|\psi^{ACBR}\rangle \otimes \psi^{AB})^n \otimes \Phi^{A_0B_0}) \]
and the output state of the $i$-th system is
\[ \xi^{\hat{A}\hat{C}\hat{B}R_i} = \text{Tr}_{[n] \setminus i} \xi^{\hat{A}\hat{C}\hat{B}R^n} = \sum_{x_{x_{x_i}}} \sqrt{p(x_{x_{x_i}})} \xi^{\hat{A}\hat{C}\hat{B}R_i} \otimes |x_i\rangle X_i, \]
where $\xi_{x_{x_{x_i}}} = \sum_{x_{x_{x_i}}} p(x_{x_{x_i}}) |\xi_{x_{x_{x_i}}}, x_{x_{x_i}}\rangle^{\hat{A}\hat{C}\hat{B}R_i}$. 

Define the distortion operator (we consider the same distortion operator for all copies of the source, that is why in the following definition, we drop the index $i$)
\[ \Delta = \sum_{x} \left( \mathbb{1} - \psi^{AB}_{x} \right) \otimes |x\rangle X, \]
so that the distortion per letter for the output state $\xi^{\hat{A}\hat{C}\hat{B}R_i}$ is
\[ D = \text{Tr}(\xi^{\hat{A}\hat{C}\hat{B}R_i} \Delta) = 1 - \sum_{x} p(x_i) \text{Tr}(\xi^{\hat{A}\hat{C}\hat{B}R_i}, |x_i\rangle X_i), \]
\[ = 1 - \sum_{x} p(x_i) F(\xi^{\hat{A}\hat{C}\hat{B}R_i}, |x_i\rangle X_i). \]
Again, up to a square this is the average fidelity considered in [1], [2], and it extends to the average-squared of per-copy fidelity when the extended distortion operator of Eq. (16) is considered. This implies that in the limit of $D \to 0$, the optimal compression rate of the ensemble source considering per-copy fidelity converges to $Q_{ea}(0)$. Therefore, by Theorem 4 (as well as the results of [1], [2]) we obtain that $Q_{ea}(0) = Q(0)$. 


Now, we define a new single-letter function which then we use to obtain simplified rates.

**Definition 6:** For a state \( \omega^{ACBR'X} = \sum_x p(x)|\psi_x^A\rangle\langle\psi_x^A|^{ACBR'} \otimes |x|^X \) and \( \epsilon \geq 0 \) define:

\[
K(D) := \sup \frac{1}{2} I(W : X|\hat{C})_\sigma \text{ over isometries } \ U : AC \to Z\hat{C}W \text{ and } \bar{U} : ZB \to \hat{A}\hat{B}V \text{ s.t.} \\
\sum_x p(x)F^2(\psi_x^{ACBR'}, \tau_x^{\hat{A}\hat{B}R'}) \geq 1 - D,
\]

where

\[
\sigma^{Z\hat{C}WBR'X} := (U \otimes \mathbb{1}_{BR'X}) \omega^{ACBR'X} (U \otimes \mathbb{1}_{BR'X})^\dagger \\
= \sum_x p(x)[\sigma_x |\sigma_x|]^{Z\hat{C}WBR'} \otimes |x|^X,
\]

\[
\tau^{\hat{A}\hat{C}BWVR'X} := (\bar{U} \otimes \mathbb{1}_{CWBR'X}) \sigma^{Z\hat{C}WBR'X} (\bar{U} \otimes \mathbb{1}_{CWBR'X})^\dagger \\
= \sum_x p(x)[\tau_x |\tau_x|]^{\hat{A}\hat{C}BWVR'} \otimes |x|^X,
\]

\[
\tau^{\hat{A}\hat{C}BVR'X} := \text{Tr}_{VW}(\tau^{\hat{A}\hat{C}BWVR'X}) \\
x_{x}^{\hat{A}\hat{C}BVR'} := \text{Tr}_{VW}(\tau_x^{\hat{A}\hat{C}BWVR'}). \notag
\]

Moreover, define \( \overline{K}(0) := \lim_{D \to 0^+} K(D) \).

**Remark 7:** Definition 6 directly implies that \( K(0) \leq \overline{K}(0) \) because \( K(D) \) is a non-decreasing function of \( \epsilon \). Furthermore, \( K(0) \) can be strictly positive, for example, for a source with trivial system \( C \) where \( \psi_x^A \psi_x^A = 0 \) holds for \( x \neq x' \), we obtain \( K(0) = S(X) \). This follows because Alice can measure her system and obtain the value of \( X \) and then copy this classical information to the register \( W \).

**Lemma 8:** For the source \( |\psi\rangle^{ACBR} = \sum_x \sqrt{p(x)}|\psi_x^A\rangle^{ACBR'} |x|^X \) and the distortion operator of Eq. (16), the rate \( \overline{Q}(0) \) is lower bounded as:

\[
\overline{Q}(0) \geq \frac{1}{2} \left( S(A|B)_\psi + S(A|C)_\psi \right) - \overline{K}(0) \\
= \frac{1}{2} I(A : R|B)_\psi - \overline{K}(0),
\]

where the above conditional mutual information is precisely the communication rate of QSR for the pure source \( |\psi\rangle^{ACBR} \). Moreover, if system \( C \) is trivial, then \( \overline{Q}(0) = \frac{1}{2} I(A : R|B)_\psi - \overline{K}(0) \).

We use the above lemma to obtain simplified rates.

**Definition 9 (Barnum et al. [39]):** An ensemble of pure states \( \mathcal{E} = \{p(x), |\psi_x^A\rangle|\psi_x^{ACBR'}\}_{x \in X} \) is called reducible if its states fall into two or more orthogonal subspaces. Otherwise the ensemble \( \mathcal{E} \) is called irreducible. We apply the same terminology to the source state \( \omega^{ACBR'X} \).

**Proposition 10:** For irreducible sources \( \overline{K}(0) = K(0) = 0 \). Hence, the optimal compression rate considering per-copy fidelity is

\[
\overline{Q}(0) = \frac{1}{2} I(A : R'|XX'|B)_\omega = \frac{1}{2} I(A : R|B)_\psi.
\]

**Proof.** Consider the following mutual information

\[
\sup I(E : X|\hat{C})_\nu \text{ over isometries } U : ACB \to \hat{A}\hat{C}BE \text{ s.t.} \\
\sum_x p(x)F^2(\psi_x^{ACBR'}, \nu_x^{\hat{A}\hat{C}BVR'}) \geq 1 - D,
\]
where the state $\nu^{\hat{A}\hat{C}\hat{B}}$ is the output state after applying the isometry $U$ on the input systems. In fact the isometries and the environments in Definition 6 are respectively special cases of the above isometry and the environment $E$ in the above optimization. Therefore, the mutual information of Definition 6 is bounded as

$$I(W : X|\hat{C})_\nu \leq I(WV : X|\hat{C})_\nu \leq I(E : X|\hat{C})_\nu. \quad (17)$$

Furthermore, for $D = 0$ we obtain

$$I(E : X|\hat{C})_\nu \leq I(E : X\hat{C})_\nu = I(E : X)_\nu + I(E : \hat{C}|X)_\nu = I(E : X)_\nu,$$

where the last equality follows because for $D = 0$ the environment $E$ and decoded system $\hat{C}$ are decoupled given $X$ (see Appendix B). For irreducible sources the mutual information $I(E : X)_\nu$ is zero which follows from the detailed discussion on p. 2028 of [39]. In the limit $D \to 0$, the value of the optimization converges to its value at $D = 0$ which follows from the fact that the fidelity and the conditional mutual information are continuous functions of CPTP maps, and the domain of the optimization is a compact set. Therefore, from Eq. (17) we conclude that

$I(W : X|\hat{C}) = 0$.

The above argument proves that $Q'(0) \geq \overline{Q}(0) \geq \frac{1}{2}I(A : R'X|B)_\omega$. Also, by definition we have $Q'(0) \leq \frac{1}{2}I(A : R'X|B)_\omega$. Therefore, $Q'(0) = \overline{Q}(0) = \frac{1}{2}I(A : R'X|B)_\omega$.

**Corollary 11:** The compression rate of the modified source defined in Eq. (14) is bounded as follows

$$\overline{Q}(0) \geq \frac{1}{2}(S(A)_\omega + S(A|Y)_\omega).$$

**Proof.** By Lemma 8, the first inequality below holds:

$$\overline{Q}(0) \geq S(A)_\omega - \frac{1}{2}I(W : X)_\sigma$$

$$= \frac{1}{2}(S(A)_\omega + S(AY)_\omega) - \frac{1}{2}I(W : X)_\sigma$$

$$= \frac{1}{2}(S(A)_\omega + S(AY)_\omega) - \frac{1}{2}I(WY : X)_\sigma$$

$$= \frac{1}{2}(S(A)_\omega + S(AY)_\omega) - \frac{1}{2}I(Y : X)_\omega - \frac{1}{2}I(W : X)_\sigma$$

$$= \frac{1}{2}(S(A)_\omega + S(AY)_\omega) - \frac{1}{2}I(Y : X)_\omega$$

$$= \frac{1}{2}(S(A)_\omega + S(A|Y)_\omega),$$

where the second line follows because the information of the orthogonal subspaces can obtained by an isometry on system $A$. The third line holds since $Y$ can be copied to the environment system. The penultimate line follows from Proposition 10 because conditioned on $Y$, the source is irreducible. The last line follows because $S(Y|X) = 0$. ■

**Definition 12:** An ensemble of pure states $\mathcal{E} = \{p(x), |\psi_x\rangle |\psi_x\rangle^{ACBR'_x}\}_{x \in X}$ is called a generic source if there is at least one $x$ for which the reduced state $\psi_x^{ACB} = \text{Tr}_{R'}|\psi_x\rangle |\psi_x\rangle^{ACBR'_x}$ has full support on $ACB$.

**Proposition 13:** For generic sources, $\overline{R}(0) = K(0) = 0$. Hence, the optimal compression rate considering per-copy fidelity is

$$\overline{Q}(0) = \frac{1}{2}I(A : R'X|B)_\omega = \frac{1}{2}I(A : R|B)_\psi.$$

We leave the proof of the above proposition to Appendix C.
IV. Discussion

We consider an entanglement-assisted rate-distortion problem with side information systems at the encoder and decoder side where the distortion measure is a general convex and continuous function of source states. We show that the optimal rate-distortion function is equal to the single-letter function $Q'(D)$ for $D > D_0$ and $\lim_{D \to D_0} Q'(D)$ for $D = D_0$, where $D_0$ is minimal distortion. Furthermore, we show that this is a convex and continuous function for $D > D_0$. Despite being single letter, computing $Q'(D)$ potentially involves unbounded optimisation since a priori there is not a dimension bound on system $Z$. Therefore, we cannot apply compactness arguments to show that if it is continuous at $D = D_0$. We later use this general theory with specific distortion operators to study various source coding problems with per-copy fidelity criteria. We consider both pure and ensemble source models of Schumacher and quantum state redistribution, and argue that we can always define quantum sources as pure states and adjust the distortion operator accordingly to impose entanglement fidelity or ensemble fidelity as the decodability criterion. Therefore, we derive the optimal entanglement-assisted compression rates for Schumacher and QSR sources with entanglement and ensemble fidelity. For both Schumacher models and also pure QSR these rates are equal to the rates considering block fidelity. The ensemble QSR with block fidelity is studied in [1], [2] where the converse is equal to $\bar{Q}(0) = \lim_{D \to D_0} Q'(D)$. The rate $Q'(0)$ is shown to be achievable, and it would only match with the converse if the function $Q'(D)$ is continuous at $D = 0$.

To analyse the distortion measure for vanishing $D$, we find a lower bound on $\bar{Q}(0)$ in terms of the limit of another function at $D = 0$, i.e. $\bar{K}(0)$. Despite the fact that computing $\bar{K}(0)$ might involve unbounded optimization as well, it is sometimes easier to analyse. In particular, we show that $\bar{K}(0) = 0$ for irreducible and generic sources. This implies that for these sources both ensemble and entanglement fidelity lead to the same compression rate, i.e. the rate of pure QSR source.

Finally, recall that in our definition of the rate-distortion task we have assumed that the encoder and decoder share free entanglement. This was motivated so as to make a smoother connection to QSR. However, it is not known whether the pre-shared entanglement is always necessary to achieve the corresponding quantum rates. There are certainly cases where QSR does not require prior entanglement, such as when Alice’s side information $C$ is trivial, which would carry over to our setting whenever $\bar{K}(0) = K(0) = 0$, for instance for an irreducible ensemble. More generally, in future work we plan to consider the trade-off between the quantum rates and the entanglement rate.

Acknowledgments. ZBK was supported by the DFG cluster of excellence 2111 (Munich Center for Quantum Science and Technology). AW was supported by the Spanish MINECO (project PID2019-107609GB-I00) with the support of FEDER funds, and by the Generalitat de Catalunya (project 2017-SGR-1127).

APPENDIX

A. Proof of Lemma 8: For the pure source $|\psi\rangle^{ACBR} = \sum_x \sqrt{p(x)} |\psi_x\rangle^{ACBR} |x\rangle^X$ and the distortion operator of Eq. (16), let $E_0$ and $D_0$ be the CPTP maps realizing the infimum of $\frac{1}{2} I(Z:R|B)_{\varphi}$, in the definition of Eq. (5). Moreover, let $U_{E_0} : AC \to ZCBW$ and $U_{D_0} : ZB \to A\tilde{B}V$ denote respectively the Stinespring isometries of $E_0$ and $D_0$. Then, the states after applying the isometries are

$$|\varphi\rangle^{ZCBWR} = (U_{E_0} \otimes 1_{BR}) |\psi\rangle^{ACBR}$$

$$|\xi\rangle^{\tilde{A}\tilde{C}B\tilde{W}VR} = (U_{D_0} \otimes 1_{\tilde{C}WR}) |\varphi\rangle^{Z\tilde{C}BWR}.$$
Now, let \( |\omega\rangle^{ABCBR'XX'} = \sum_x \sqrt{p(x)} |\psi_x\rangle^{ABCBR} \otimes |x\rangle^X \otimes |x\rangle^{X'}\) be the purification of the state \( \omega^{ABCBR'XX'} \) in Definition 6 and define the following states:

\[
\begin{align*}
|\sigma\rangle^{\bar{Z}\bar{C}\bar{BWR}'XX'} &= (U_{E\sigma} \otimes 1_{BWR'XX'}) |\omega\rangle^{ABCBR'XX'} \\
|\tau\rangle^{\bar{X}X'\bar{C}\bar{BWR}'XX'} &= (U_{D\sigma} \otimes 1_{\bar{C}BWR'XX'}) |\sigma\rangle^{\bar{Z}\bar{C}\bar{BWR}'XX'}.
\end{align*}
\]

(18)

Notice that \( \frac{1}{2} I(Z : R|B)_\psi = \frac{1}{2} I(Z : R'X'|B)_\sigma \). In what follows, we establish lower bounds on \( I(Z : R'X'|B)_\sigma \).

\[
I(Z : R'X'|B)_\sigma = S(ZB)_\sigma - S(B)_\sigma - S(ZBR'XX')_\sigma + S(BR'XX')_\sigma \\
= S(\bar{C}WR'XX')_\sigma - S(B)_\sigma - S(\bar{C}W)_\sigma + S(AC)_\sigma \\
= S(AB)_\omega - S(AB)_\omega + S(C)_\omega - S(AC)_\omega + S(\bar{C}WR'XX')_\sigma - S(B)_\sigma - S(\bar{C}W)_\sigma + S(AC)_\sigma \\
= I(A : R'XX'|B)_\omega - S(AB)_\omega + S(C)_\omega + S(\bar{C}WR'XX')_\sigma - S(\bar{C}W)_\sigma,
\]

(19)

where the first line follows by the definition of the conditional mutual information. The second line follows because the Fannes-Audenaert inequality. Eq. (24) follows from the subadditivity of the entropy. Eq. (25) follows from the fidelity criterion in Definition 6: the output state on the quantum conditional entropy. Eq. (23) follows from the fidelity criterion in Definition 6: the output state on the system \( \bar{A} \bar{B} \) is \( 2\sqrt{D} \)-close to the original state \( AB \). The last line follows since the state \( |\omega\rangle^{ABCBR'XX'} \) is pure. Also, notice that \( I(A : R'XX'|B)_\omega = I(A : R|B)_\psi \). Now, we lower bound the following term in the last line of the above chain of equations

\[
-S(AB)_\omega + S(C)_\omega + S(\bar{C}WR'XX')_\sigma - S(\bar{C}W)_\sigma
\]

\[
\geq -S(AB)_\omega + S(C)_\omega + S(\bar{A} \bar{B} V)_\omega - S(W|\bar{C})_\sigma - 2\sqrt{D} \log(|C|) - h(\sqrt{2D})
\]

(20)

\[
= -S(\bar{A} \bar{B} V)_\omega + S(W|\bar{C})_\sigma - 2\sqrt{D} \log(|C|) - h(\sqrt{2D})
\]

(21)

\[
\geq -S(\bar{A} \bar{B} V)_\omega + S(W|\bar{C})_\sigma - 2\sqrt{D} \log(|A|) - h(\sqrt{2D}) - 2h(\sqrt{2D})
\]

(22)

\[
= S(W|\bar{C})_\sigma - S(V|\bar{A} \bar{B} V)_\sigma - 2\sqrt{D} \log(|A|) - h(\sqrt{2D}) - 2h(\sqrt{2D})
\]

(23)

where Eq. (20) follows because the state \( |\sigma\rangle^{\bar{Z}\bar{C}\bar{BWR}'XX'} \) is pure. Eq. (21) follows from the fidelity criterion in Definition 6: the output state on the system \( \bar{C} \) is \( 2\sqrt{D} \)-close to the original state \( C \) in trace norm; then the inequality follows by applying the Fannes-Audenaert inequality [40], [41]. Eq. (22) follows due to definition of the quantum conditional entropy. Eq. (23) follows from the fidelity criterion in Definition 6: the output state on the system \( \bar{A} \bar{B} \) is \( 2\sqrt{D} \)-close to the original state \( AB \) in trace norm; then the inequality follows by applying the Fannes-Audenaert inequality. Eq. (24) follows from the subadditivity of the entropy. Eq. (25) follows from the
fidelity criterion in Definition 6: the output state on the system $\overline{A'B}X$ is $2\sqrt{2D}$-close to the original state $ABX$ in trace norm; then the inequality follows by applying the Fannes-Audenaert inequality. Eq. (26) follows because the state $|\omega\rangle^{ACBR'XX'}$ is pure. Eq. (27) follows from the fidelity criterion in Definition 6: the output state on the system $\overline{C'R'}X$ is $2\sqrt{2D}$-close to the original state $CR'X$ in trace norm; then the inequality follows by applying the Fannes-Audenaert inequality. Eq. (28) follows because the state $|\tau\rangle^{\overline{AC'B}WV'R'XX'}$ is pure. Eq. (29) follows from the chain rule. Eq. (30) follows from the definition of $K(D)$.

From Eq. (19) and Eq. (30), we obtain

$$\frac{1}{2}I(Z : R|B)_\varphi = \frac{1}{2}I(Z : R'XX'|B)_\sigma \geq \frac{1}{2}I(A : R'XX'|B)_\omega - K(D) - \frac{1}{2}I(W : R'\overline{C}X)_\sigma - \sqrt{\frac{D}{2}} \log(|A|^2|C|^2|B|^2|X|^2|R'|) - 2h(\sqrt{2D}).$$

In section B of the appendix we prove that $\lim_{D\to 0} I(W : R'|\overline{C}X)_\sigma = \lim_{D\to 0} I(W : R'|X)_\sigma = 0$ (decoupling condition). Therefore, in the limit of $D \to 0$ the inequality of the lemma follows

$$\frac{1}{2}I(Z : R|B)_\varphi \geq \frac{1}{2}I(A : R'XX'|B)_\omega - K(0).$$

$$= \frac{1}{2}I(A : R|B)_\psi - K(0).$$

Finally, we prove the last statement of the lemma, that is if $C$ and $\overline{C}$ are trivial systems, then $\frac{1}{2}I(Z : R|B)_\varphi = \frac{1}{2}I(A : R|B)_\psi - K(0)$. From Eq. (19), we have

$$I(Z : R'XX'|B)_\sigma = I(A : R'XX'|B)_\omega - S(AB)_\omega + S(WR'XX')_\sigma - S(W)_\sigma$$

$$= I(A : R'XX'|B)_\omega - S(R'XX')_\omega + S(WR'XX')_\sigma - S(W)_\sigma$$

$$= I(A : R'XX'|B)_\omega - I(W : R'XX')_\sigma$$

$$\leq I(A : R'XX'|B)_\omega - I(W : R'XX)_\sigma,$$

(31)

where the second line follows because for trivial $C$ the state on systems $ABR'XX'$ is pure. The last line follows from data processing inequality. Also, from Eq. (29) we have the following:

$$I(Z : R'XX'|B)_\sigma \geq I(A : R'XX'|B)_\omega - I(W : R'X)_\sigma - \sqrt{2D} \log(|A|^2|C|^2|B|^2|X|^2|R'|) - 4h(\sqrt{2D})$$

(32)

The decoupling condition (section B of the appendix), Eq. (31) and Eq. (32) imply that in the limit of $D \to 0$

$$\frac{1}{2}I(Z : R|B)_\varphi = \frac{1}{2}I(Z : R'XX'|B)_\sigma$$

$$= \frac{1}{2}I(A : R'XX'|B)_\omega - \frac{1}{2}I(W : X)_\sigma$$

$$= \frac{1}{2}I(A : R|B)_\psi - \frac{1}{2}I(W : X)_\sigma.$$

Notice that since the term $I(A : R|B)_\psi$ is constant, taking the infimum of $\frac{1}{2}I(Z : R|B)_\varphi$ is equivalent to taking the supremum of $I(W : X)_\sigma$, therefore, the lemma follows.

B. Decoupling condition: Here we prove that the conditional mutual information of Eq. (30) $\lim_{D\to 0} I(W : R'|\overline{C}X)_\sigma = \lim_{D\to 0} I(W : R'|X)_\sigma = 0$.

Consider the following reduced states of the states defined in Eq. (18)
\[
\sigma^x_{\mathcal{BC}W'V'} = \sum_x p(x) |\sigma_x\rangle\langle \sigma_x| \otimes |x\rangle\langle x|_{X}
\]
\[
\tau_{\mathcal{AB}W'V'} = \sum_x p(x) |\tau_x\rangle\langle \tau_x| \otimes |x\rangle\langle x|_{X}.
\]

The fidelity criterion of Definition 6 implies the following
\[
1 - D \leq \sum_x p(x) F^2(\psi_x^{\mathcal{AB}W'V'}, \tau_x^{\mathcal{AB}W'V'})
\]
\[
= \sum_x p(x) (\psi_x^{\mathcal{AB}W'V'} | \psi_x)
\]
\[
\leq \sum_x p(x) \| \tau_x^{\mathcal{AB}W'V'} \|,
\]
\[
\text{(33)}
\]

where \(\| \cdot \|\) denotes the operator norm, which in this case of a positive semidefinite operator is the maximum eigenvalue of \(\tau_x^{\mathcal{AB}W'V'}\). Now, consider the Schmidt decomposition of the state \(|\tau_x\rangle^{\mathcal{AB}W'V'}\) with respect to the partition \(\mathcal{AB}|W'V'\), i.e.
\[
|\tau_x\rangle^{\mathcal{AB}W'V'} = \sum_i \sqrt{\lambda_x(i)} |v_x(i)\rangle^{\mathcal{AB}W'} |w_x(i)\rangle^{W'}.
\]

The fidelity of Eq. (33) implies that the subsystems of the partition are almost decoupled on average:
\[
\sum_x p(x) F^2(\tau_x^{\mathcal{AB}W'V'}, \tau_x^{\mathcal{AB}W'V'} \otimes x^{WV})
\]
\[
= \sum_x p(x) \sum_i \lambda_x(i)^3
\]
\[
\geq \sum_x p(x) \| \tau_x^{\mathcal{AB}W'V'} \|^3
\]
\[
\geq \left( \sum_x p(x) \| \tau_x^{\mathcal{AB}W'V'} \| \right)^3
\]
\[
\geq (1 - D)^3,
\]
\[
\geq 1 - 3D,
\]
\[
\text{(34)}
\]

where the penultimate line follows from the convexity of the function \(x^3\). The last line is due to Eq. (33). By the Alicki-Fannes inequality, this implies
\[
I(WV : R'|\bar{C}X)_\tau = S(R'|\bar{C}X)_\tau - S(R'|WV\bar{C}X)_\tau
\]
\[
\leq S(R'|\bar{C}X)_\tau - S(R'|WV\bar{C}X)_\tau + 4\sqrt{6D} \log(|R'|) + g(2\sqrt{6D})
\]
\[
= 4\sqrt{6D} \log(|R'|) + g(2\sqrt{6D}),
\]

the second line follows because \(\bar{\tau} = 2\sqrt{6D}\)-close to state \(\tau\) in trace norm where \(\bar{\tau}^{\mathcal{AB}W'VW} = \sum_x p(x) \tau_x^{\mathcal{AB}W'V} \otimes |x\rangle\langle x|_{X}\). The last line follows because \(S(R'|WV\bar{C}X)_\tau = S(R'|\bar{C}X)_\tau\). Then, the decoupling follows in the limit of \(D \to 0\) since by data processing \(I(W : R'|\bar{C}X)_\tau \leq I(WV : R'|\bar{C}X)_\tau\). We can similarly prove that \(\lim_{D \to 0} I(W : R'|X)_\tau = 0\).

C. Proof of Proposition 13: We denote the Stinespring isometries of CPTP maps \(\mathcal{E}_0\) and \(\mathcal{D}_0\) defined in Definition 6 respectively by \(U_{\mathcal{E}_0} : AC \to Z\bar{C}W\) and \(U_{\mathcal{D}_0} : ZB \to \bar{AB}V\). For generic sources, we show that the environment systems \(W\) and \(V\) satisfy \(\lim_{D \to 0} I(WV : X|\bar{C})_\tau = 0\). Thus, we obtain
\[
K(0) = \lim_{D \to 0} I(W : X|\bar{C})_\tau \leq \lim_{D \to 0} I(WV : X|\bar{C})_\tau = 0.
\]
First, we note that the fidelity in Eq. (34) implies the following
\[ \sum_x p(x) \| \tau_x \hat{A} \hat{C} \hat{BWV} \hat{R} - \tau_x \hat{A} \hat{CB} \hat{R} \otimes \tau_x \hat{WV} \|_1 \leq 2\sqrt{\lambda_D}. \]

We also obtain the following bound by the definition of the state \( \tau \) (Definition 6):
\[ \sum_x p(x) \| \tau_x \hat{ACBR} \otimes \tau_x \hat{WV} - \psi_x^{ACBR} \otimes \tau_x \hat{WV} \|_1 \leq 2\sqrt{\lambda_D}. \]

By applying triangle inequality to the above equations, we obtain
\[ \sum_x p(x) \| \tau_x \hat{ACBWV} \hat{R} - \psi_x^{ACBR} \otimes \tau_x \hat{WV} \|_1 \leq 2(\sqrt{\lambda_D} + \sqrt{2\lambda_D}). \]  
(35)

Since the source is generic, there is an \( x \), say \( x = 0 \), for which \( \psi_0^{ABC} \) has full support on \( \mathcal{L}(\mathcal{H}_{ACB}) \), i.e. \( \lambda_0 := \lambda_{\text{min}}(\psi_0^{ABC}) > 0 \). Therefore, for any \( \psi_x^{ACBR} \) there is an operator \( T_x \) acting on the reference system \( R' \) such that
\[
|\psi_x^{ACBR} = (\mathbf{1}_{ACB} \otimes T_x)|\psi_0^{ACBR},
\]
and \( \| T_x \| \leq \frac{1}{\sqrt{\lambda_0}} \) [21]. We can also rewrite the output state as follows:
\[
\tau_x^{\hat{ACBWV} R'} = (U_{Dx} U_{\xi_0} \otimes \mathbf{1}_{R'}) \psi_x^{ACBR} (U_{Dx} U_{\xi_0} \otimes \mathbf{1}_{R'})^\dagger
= (\mathbf{1}_{ACB} \otimes T_x) \psi_0^{ACBR} (\mathbf{1}_{ACB} \otimes T_x)^\dagger (U_{Dx} U_{\xi_0} \otimes \mathbf{1}_{R'})^\dagger
= (\mathbf{1}_{ACBWV} \otimes T_x) (U_{Dx} U_{\xi_0} \otimes \mathbf{1}_{R'}) \psi_0^{ACBR} (U_{Dx} U_{\xi_0} \otimes \mathbf{1}_{R'})^\dagger (\hat{ACBWV} \otimes T_x)^\dagger
= (\hat{ACBWV} \otimes T_x) \tau_0^{\hat{ACBWV} R'} (\hat{ACBWV} \otimes T_x)^\dagger
\]

We now replace \( \tau_x \) and \( \tau_x \) with the above states to obtain the following:
\[
\sum_x p(x) \| \tau_x^{\hat{ACBWV} R'} - \psi_x^{ACBR} \otimes \tau_0^{WV} \|_1
= \sum_x p(x) \| (\mathbf{1} \otimes T_x) \tau_0^{\hat{ACBWV} R'} (\mathbf{1} \otimes T_x)^\dagger - (\mathbf{1} \otimes T_x) \psi_0^{ACBR} (\mathbf{1} \otimes T_x)^\dagger \otimes \tau_0^{WV} \|_1
= \sum_x p(x) \| T_x \| \| \tau_0^{\hat{ACBWV} R'} - \psi_0^{ACBR} \otimes \tau_0^{WV} \|_1
\leq 2(\sqrt{\lambda_D} + \sqrt{2\lambda_D}) \sum_x p(x) \| T_x \|
\leq \frac{2(\sqrt{\lambda_D} + \sqrt{2\lambda_D})}{\sqrt{\lambda_0}}.
\]

where the third line follows from Eq. (35). We use the above upper bound on the average distance between the reduced states \( \tau_x^{\hat{CWV}} \) and \( \psi_x^{C} \otimes \tau_0^{WV} \) to conclude that the environment systems \( W V \) are decoupled from systems \( \hat{C} X \):
\[
\frac{1}{2} \sum_x p(x) \| \tau_x^{\hat{CWV}} \otimes |x\rangle^X \langle x| - \sum_x p(x) \omega_x^{C} \otimes \tau_0^{WV} \|_1 \leq \frac{1}{2} \sum_x p(x) \| \tau_x^{\hat{CWV}} - \omega_x^{C} \otimes \tau_0^{WV} \|_1
\leq \frac{(\sqrt{\lambda_D} + \sqrt{2\lambda_D})}{\sqrt{\lambda_0}} =: \delta_D.
\]

By applying the Alicki-Fannes inequality in the form of Eq. (12) to the above states, we obtain
\[
I(WV : \hat{C}X)_\tau = S(\hat{C}X)_\tau - S(CX|WV)_\zeta + S(CX|WV)_\zeta - S(\hat{C}X|WV)_\tau
= S(\hat{C}X)_\tau - S(CX)_\zeta + S(CX|WV)_\zeta - S(\hat{C}X|WV)_\tau
\leq S(\hat{C}X|WV)_\zeta - S(\hat{C}X|WV)_\tau + 2\sqrt{\lambda_D} \log(|C| \cdot |X|) + h(\sqrt{2\lambda_D})
\leq 2\delta_D \log(|C| \cdot |X|) + h(\delta_D) + 2\sqrt{\lambda_D} \log(|C| \cdot |X|) + h(\sqrt{2\lambda_D}),
\]
where the third line follows from the fidelity criterion in Definition 6: the output state on the system $\tilde{C}X$ is $2\sqrt{2D}$-close to the original state $CX$ in trace norm; then the inequality follows by applying the Fannes-Audenaert inequality. Therefore, we conclude that $I(W: X|\tilde{C})_\tau \leq I(WV: X|\tilde{C})_\tau \leq I(WV: X\tilde{C})_\tau$ which the latter vanishes for $D \to 0$.

REFERENCES

[1] Z. B. Khanian and A. Winter, “Quantum State Redistribution for Ensemble Sources”, in: Proc. 2020 IEEE International Symposium on Information Theory (ISIT), Los Angeles, CA, 21-26 June 2020, pp. 1858-1863, IEEE, 2020.
[2] Z. B. Khanian, From Quantum Source Compression to Quantum Thermodynamics, Ph.D. dissertation, Universitat Autònoma de Barcelona, Department of Physics, Spain, Nov 2020, arXiv:quant-ph/2012.14143.
[3] C. E. Shannon, “A mathematical theory of communication”, Bell Syst. Tech. J., vol. 27, pp. 379-423 & 623-656, 1948.
[4] D. B. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources”, IEEE Trans. Inf. Theory, vol. 19, no. 4, pp. 461-480, 1973.
[5] C. E. Shannon, “Coding theorems for a discrete source with a fidelity criterion”, IRE Int. Conv. Records, vol. 7, pp. 142-163, 1959.
[6] T. Berger, Rate-Distortion Theory: A Mathematical Basis for Data Compression, Prentice-Hall, Englewood Cliffs, N.J., 1971.
[7] A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder”, IEEE Trans. Inf. Theory, vol. 22, no. 1, pp. 1–10, 1976.
[8] B. Schumacher, “Quantum coding,” Phys. Rev. A, vol. 51, no. 4, pp. 2738-2747, Apr 1995.
[9] R. Jozsa and B. Schumacher, “A new proof of the quantum noiseless coding theorem”, J. Mod. Optics, vol. 41, no. 12, pp. 2343-2349, 1994.
[10] H. Barnum, C. A. Fuchs, R. Jozsa and B. Schumacher, “General fidelity limit for quantum channels,” Physical Review A, vol. 54, no. 6, pp. 4707-4711, Dec 1996.
[11] M. Horodeckii,”Limits for compression of quantum information carried by ensembles of mixed states,” Phys. Rev. A, vol. 57, no. 5, pp. 3364-3369, May 1998.
[12] A. Winter, Coding Theorems of Quantum Information Theory, Ph.D. dissertation, Universität Bielefeld, Department of Mathematics, Germany, July 1999, arXiv:quant-ph/9907077.
[13] I. Devetak and A. Winter, “Classical data compression with quantum side information,” Phys. Rev. A, vol. 68, 042301, Oct 2003.
[14] C. A. Fuchs and J. van de Graaf, “Cryptographic distinguishability measures for quantum-mechanical states,” IEEE Trans. Inf. Theory, vol. 59, no. 10, pp. 6755-6773, July 2013.
[15] G. Smith, J. A. Smolin, and A. Winter, “The quantum capacity with symmetric side channels”, IEEE Trans. Inf. Theory, vol. 54, no. 9, pp. 4208-4217, 2008.

[32] G. Smith, “The private classical capacity with a symmetric side channel and its application to quantum cryptography”, Phys. Rev. A, vol. 78, 022306, 2008.

[33] M. Christandl and A. Winter, “‘Squashed Entanglement’ – An Additive Entanglement Measure”, J. Math. Phys., vol. 45, no. 3, pp. 829-840, 2004.

[34] D. Yang, M. Horodecki, and Z. D. Wang, “An Additive and Operational Entanglement Measure: Conditional Entanglement of Mutual Information”, Phys. Rev. Lett., vol. 101, 140501, 2008.

[35] N. Datta, C. Hirche, and A. Winter, “Convexity and Operational Interpretation of the Quantum Information Bottleneck Function”, in: Proc. 2019 IEEE Int. Symp. Inf. Theory (ISIT), 7-12 July 2019, Paris, pp. 1157-1161, IEEE, 2019; arXiv[quant-ph]:1810.03644v3.

[36] R. Alicki and M. Fannes, “Continuity of quantum conditional information,” J. Phys. A: Math. Gen., vol. 37, no. 5, pp. L55-L57, Jan 2004.

[37] A. Winter, “Tight Uniform Continuity Bounds for Quantum Entropies: Conditional Entropy, Relative Entropy Distance and Energy Constraints,” Commun. Math. Phys., vol. 347, no. 1, pp. 291-313, Oct 2016.

[38] Z. B. Khanian and A. Winter, “Entanglement-Assisted Quantum Data Compression,” in: Proc. 2019 IEEE Int. Symp. Inf. Theory (ISIT), Paris, France, pp. 1147-1151, July 2019.

[39] H. Barnum, P. Hayden, R. Jozsa, and A. Winter, “On the reversible extraction of classical information from a quantum source,” Proc. Roy. Soc. London Ser. A, vol. 457, no. 2012, pp. 2019-2039, Aug. 2001.

[40] M. Fannes, “A continuity property of the entropy density for spin lattice systems,” Commun. Math. Phys., vol. 31, no. 4, pp. 291-294, Dec 1973.

[41] K. M. R. Audenaert, “A sharp continuity estimate for the von Neumann entropy,” J. Phys. A: Math. Theor., vol. 40, no. 28, pp. 8127-8136, Jun 2007.

[42] D. Leung and G. Smith, “Continuity of Quantum Channel Capacities,” Commun. Math. Phys., vol. 292, no. 1, pp. 201-215, 2009.

[43] A. Anshu, M.-H. Hsieh, and R. Jain, “Noisy quantum state redistribution with promise and the Alpha-bit,” arXiv[quant-ph]:1803.03414, Mar 2018.

[44] M.-H. Hsieh and S. Watanabe, “Channel Simulation and Coded Source Compression”, IEEE Trans. Inf. Theory, vol. 62, no. 11, pp. 6609-6619, 2016.