A New Formulation of Lossy Quantum-Classical and Classical Source Coding based on a Posterior Channel

Mohammad Aamir Sohail
University of Michigan, USA
Email: mdaamir@umich.edu

Touheed Anwar Atif
University of Michigan, USA
Email: touheed@umich.edu

S. Sandeep Pradhan
University of Michigan, USA
Email: pradhanv@umich.edu

2023 IEEE International Symposium on Information Theory (ISIT) | 978-1-6654-7554-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISIT54713.2023.10206859

Abstract—In this work, we address the lossy quantum-classical (QC) source coding problem, where the task is to compress the classical information about a quantum source, obtained after performing a measurement, below the Shannon entropy of the measurement outcomes, while incurring a bounded reconstruction error. We propose a new formulation, namely, "rate-channel theory", for the lossy QC source coding problem based on the notion of a backward (posterior) channel. We employ a single-letter posterior channel to capture the reconstruction error in place of the single-letter distortion observable. The formulation requires the reconstruction of the compressed quantum source to satisfy a block error constraint as opposed to the average single-letter distortion criterion in the rate-distortion setting. We also develop an analogous formulation for the classical variant with respect to a corresponding posterior channel. Furthermore, we characterize the asymptotic performance limit of the lossy QC and classical source coding problems in terms of single-letter quantum mutual information and mutual information quantities of the given posterior channel, respectively. We provide examples for the above formulations.

I. Introduction

We consider a fundamental task of compressing the classical information about a quantum source while suffering a bounded error between the source and its reconstruction, namely, the lossy quantum-classical (QC) source coding problem [1]. In this setting, the sender performs a collective measurement on several copies of the quantum source, creating a classical sequence that is compressed and sent to the receiver over a noiseless classical channel. The receiver then outputs a classical sequence, stored in a quantum register, using a decoding map while incurring a bounded reconstruction error. The authors in [1] considered the additive single-letter distortion observable as a reconstruction error criterion and obtained a single-letter expression for the asymptotic QC rate-distortion function in terms of minimal quantum mutual information. The minimization is done over all POVMs that satisfy the distortion constraint.

In QC setup, considering the average single-letter distortion criterion draws inspiration from Shannon's formulation of lossy classical source coding problem [2], where a single-letter characterization is available for the additive single-letter distortion criterion. In Shannon's formulation, the motivation for an average single-letter error criterion emanates from the strong converse of the lossless classical source coding theorem. The theorem states that the entropy bound cannot be breached even when the asymptotic probability of block error is relaxed to any number in $(0, 1)$ [3, Theorem 1.1].

In addition to Shannon's work on the rate-distortion problem [2], there have been several works discussing the lossy source compression problem [4]. A concept that has received significant attention in the lossy regime is the notion of a backward channel [3, Problem 8.3], which characterizes the posterior distribution of the source given the reconstruction. The structure of this channel has been studied in [5]–[7]. Moreover, the rate-distortion achievability result in [3, Theorem 7.3] is shown by constructing a channel code for a backward channel with a large probability of error and by using the encoder of the latter as a decoder of the former and vice versa. For further developments on this concept, see [8]–[13]. Highlighting this further, the posterior or backward channel (a posteriori probability) can be defined for any information processing system, and it has been widely used in Bayesian decision theory [14], detection and estimation theories such as MAP estimation. [15]. On the other hand, in many applications such as speech processing and inference in machine learning, defining a suitable and well-behaved distortion function that can capture the loss incurred in compression is a difficult task [16]. Therefore, to encompass a broader set of applications, our motivation is to bring the Bayesian perspective to the lossy compression problem without using a distortion function. Furthermore, we aim to use a more information-theoretic object, such as a channel, to capture the loss in lossy source coding problems. The lossy compression scheme can now be interpreted as providing a rate-limited representation of the source with a pre-specified a posteriori probability distribution of the source given its reconstruction.

In light of this, we develop a new formulation of the lossy QC and classical source coding problems based on the notion of a posterior channel that produces the source from its reconstruction, which we called as "rate-channel theory". We use a single-letter posterior channel to characterize the nature of the loss incurred in the encoding and decoding operations, instead of a single-letter distortion function. In the QC setup, we use posterior classical-quantum (CQ) channel to describe the relationship between the source's reference and its reconstruction. Furthermore, motivated by the probability of error (block error) constraint in Shannon's lossless compression, we consider a block error constraint instead of the average symbol-wise error criterion. We want to construct an

This work was supported by NSF grants CCF 2007878 and CCF 2132815.

encoder and a decoder such that the joint effect of producing a reconstruction sequence from the source sequence is close to the effect of the n-product posterior channel acting on the non-product reconstruction sequence, manifesting as a block error constraint. The closeness is measured using the trace distance the total variation in the QC and the classical case, respectively.

As the main contribution of our work, we provide a single-letter characterization of the asymptotic performance limit using the minimal quantum mutual information and the minimal mutual information of the posterior channel in the QC and classical setup, respectively, where the minimization is over all reconstruction distributions (see Theorem 1 and 2). As for the achievability of Theorems 1 and 2, we use Winter's measurement compression protocol [17] to construct the encoding POVM, and the likelihood encoder as discussed in [18], [19] to construct the randomized encoder, respectively. Proof of the converse of Theorem 1 uses inequalities such as the concavity of conditional quantum entropy, and the continuity of quantum mutual information. Similar tools are used for the converse of Theorem 2. Moreover, this work opens up an opportunity to investigate the application of the notion of the posterior channel in other quantum source coding problems where a single-letter characterization is not available.

II. Preliminaries and Notations

The set of density operators on H_B is denoted by $D(H_B)$. We denote H_{B_R} as the Hilbert space associated with the reference space of H_B, with $\dim H_{B_R} = \dim H_B$. We denote the finite alphabet of a source as X, and the set of probability distributions on the finite alphabet X as $P(X)$. Let $[\Theta] \triangleq \{1, 2, \cdots, \Theta\}$.

Definition 1: (Classical-Quantum (CQ) Channel) Given a finite set X and a Hilbert space H, a CQ channel W is specified by a collection $(W_x \in D(H) : x \in X)$ of density operators.

III. Main Results

A. Lossy Quantum-Classical Source Coding

Consider a memoryless quantum source $\rho_B \in D(H_B)$.

Definition 2: (QC Source Coding Setup) A QC source coding setup is characterized by a triple (ρ_B, X, W) where ρ_B is the source density operator acting on H_B, X is the reconstruction alphabet, and $W : X \to D(H_B)$ is a single-letter posterior classical-quantum (CQ) channel.

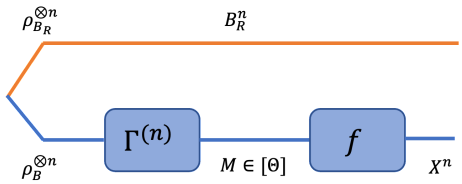

Fig. 1. Illustration of Lossy QC Source Compression Protocol.

Definition 3: (Lossy QC Compression Protocol) For a given source density operator ρ_B and a reconstruction alphabet X, an (n, Θ) lossy QC compression protocol is characterized by (i) a POVM $\Gamma^{(n)} \triangleq \{A_m\}_{m=1}^{\Theta}$ and (ii) a decoding map $f : \{1, 2, \cdots, \Theta\} \to X^n$, as shown in Fig. 1.

Definition 4: (Achievability) For a given QC source coding setup (ρ_B, X, W), a rate R is said to be achievable if for all $\epsilon > 0$ and all sufficiently large n, there exists an (n, Θ) QC lossy compression protocol such that $\log \Theta / n \le R + \epsilon$, and $\Xi(\Gamma^{(n)}, f) \le \epsilon$, where $\Xi(\Gamma^{(n)}, f) \triangleq$

$$\sum_{x^n} M_{x^n} \otimes |x^n\rangle\langle x^n| - \sum_{x^n} P_{X^n}(x^n) W_{x^n} \otimes |x^n\rangle\langle x^n|,$$

for all $x^n \in X^n$, $M_{x^n} \otimes |x^n\rangle\langle x^n| \triangleq \sqrt{\rho_B^{\otimes n}}^{-1} A_{f^{-1}(x^n)} \sqrt{\rho_B^{\otimes n}}^{-1} \otimes |x^n\rangle\langle x^n|$, is the unnormalized system-induced density operator on systems $B_R B$, $P_{X^n}(x^n) \triangleq \mathrm{Tr}(A_{f^{-1}(x^n)} \rho_B^{\otimes n})$ is the probability of observing the reconstruction sequence x^n, $W_{x^n} \otimes |x^n\rangle\langle x^n|$ is the approximating density operator on systems $B_R B$, and $W_{x^n} \triangleq \bigotimes_{i=1}^{n} W_{x_i}$.

Theorem 1: (Lossy QC Source Compression Theorem) For a (ρ_B, X, W) QC source coding setup, a rate R is achievable if and only if $A(\rho_B, W)$ is non-empty, and

$$R \ge \min_{P_X \in A(\rho_B, W)} I(X; B_R)_\sigma,$$

where the quantum mutual information is computed with respect to the classical-quantum state, $\sigma^{X B_R} \triangleq \sum_x P_X(x) |x\rangle\langle x|_X \otimes W_x$, $\{|x\rangle\}_{x \in X}$ is an orthonormal basis for the Hilbert space H_X with $\dim(H_X) = |X|$, and A is the set of reconstruction distributions defined as $A(\rho_B, W) \triangleq \{P_X \in P(X) : \sum_x P_X(x) W_x = \rho_B\}$.

Proof. A proof of the achievability is provided in Section V-A, and a converse proof is provided in Section V-B. □

B. Lossy Classical Source Coding

Consider a discrete memoryless source (DMS) X characterized by a source distribution P_X over a finite alphabet X.

Definition 5: (Source Coding Setup) A source coding setup is characterized by a triple $(P_X, \hat{X}, W_{X|\hat{X}})$ where P_X is the source distribution over a finite alphabet X, $\hat{X}$ is the reconstruction alphabet, and $W_{X|\hat{X}} : \hat{X} \to X$ is the single-letter posterior (backward) channel, i.e., the conditional distribution of source given the reconstruction.

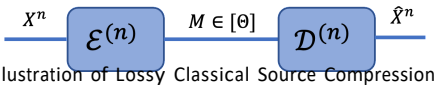

Fig. 2. Illustration of Lossy Classical Source Compression Protocol.

Definition 6: (Lossy Source Compression Protocol) For a given source distribution P_X and a reconstruction alphabet $\hat{X}$, an (n, Θ) lossy source compression protocol consists of (i) a randomized encoding map $E : X^n \to [\Theta]$ and (ii) a randomized decoding map $D : [\Theta] \to \hat{X}^n$, as shown in Fig. 2.

Definition 7: (Achievability) Given a source coding setup $(P_X, \hat{X}, W_{X|\hat{X}})$, a rate R is said to be achievable if for all $\epsilon > 0$ and all sufficiently large n, there exists an (n, Θ) lossy source compression protocol such that $\frac{1}{n} \log \Theta \le R + \epsilon$, and $\Xi(E, D) \le \epsilon$, where $\Xi(E, D) \triangleq$

$$\frac{1}{2} \sum_{x^n \hat{x}^n} P_{X^n \hat{X}^n}(x^n, \hat{x}^n) - P_{X^n}(\hat{x}^n) W^n_{X|\hat{X}}(x^n|\hat{x}^n),$$

$P_{X^n\hat{X}^n}(x^n,\hat{x}^n) = P_X^n(x^n)\sum_{m\in[\Theta]} E(m|x^n)D(\hat{x}^n|m)$, for all $(x^n,\hat{x}^n)\in X^n\times\hat{X}^n$, is the system-induced joint distribution, $P_{X}\,W^n_{X|\hat{X}}$ is the approximating joint distribution, i.e., the n−product posterior channel acting on the non-product reconstruction and $W^n_{X|\hat{X}}(x^n|\hat{x}^n)\triangleq\prod_{i=1}^n W_{X|\hat{X}}(x_i|\hat{x}_i)$.

Theorem 2: (Lossy Source Compression Theorem) For a $(P_X,\hat{X},W_{X|\hat{X}})$ source coding setup, a rate R is said to be achievable if and only if $A(P_X,W_{X|\hat{X}})$ is non-empty, and

$$R \geq \min_{P_{\hat{X}}\in A(P_X,W_{X|\hat{X}})} I(X;\hat{X}), \tag{1}$$

where $A(P_X,W_{X|\hat{X}})\triangleq\{P_{\hat{X}}\in P(\hat{X}) : \text{for all } x\in X, \sum_{\hat{x}} P_{\hat{X}}(\hat{x})W_{X|\hat{X}}(x|\hat{x}) = P_X(x)\}$, is the set of reconstruction distributions.

Proof. A proof of the achievability is provided in Section VI-A, and a converse proof is provided in Section VI-B. □

Remark 1 (Comparison with Shannon's noiseless source compression): Noiseless source compression requires $\lim_{n\to\infty} P(X^n=\hat{X}^n)=0$. In the current formulation, if one chooses the identity posterior channel, i.e., $W_{X|\hat{X}}(x|\hat{x})=1_{\{x=\hat{x}\}}$, for all $(x,\hat{x})\in X\times\hat{X}$, we require $\lim_{n\to\infty}\|P_{X^n\hat{X}^n} - P_{\hat{X}^n}W^n_{X|\hat{X}}\|_{TV}=0$. We can easily see that the two conditions are equivalent, and both formulations yield the same asymptotic performance limit of Shannon's entropy. However, the standard source coding formulation using the average single-letter distortion criterion at zero distortion level is not equivalent to noiseless source compression.

IV. Illustrative Examples

Example 1: (Lossy QC Source Coding for Binary Quantum Source with Binary Symmetric Posterior CQ Channel) We develop an example similar to that studied in [1]. Consider a quantum source ρ_B that generates the state $|+\rangle$ and $|0\rangle$ with probability p and $(1-p)$, respectively, where $p\in[0,1/2]$, i.e., $\rho_B = p|+\rangle\langle+| + (1-p)|0\rangle\langle 0|$, the reconstruction set $\hat{X}=\{0,1\}$, and the posterior CQ channel $W_x = (1-q)\,\omega_x + q\,\omega_{\bar{x}}$, where $q\in[0,1/2]$, $\omega_0 = (1/4)|+\rangle\langle+| + (3/4)|0\rangle\langle 0|$, $\omega_1 = (3/4)|+\rangle\langle+| + (1/4)|0\rangle\langle 0|$, and $\bar{x}\triangleq x\oplus 1$. Toward identifying the set A, we assume $P_X(0)=r$, which characterizes the set A, and solve the following: $\rho_B = rW_0+(1-r)W_1$, $0\leq r\leq 1$. This gives, if $0\leq q\leq 2\min\{(3/4-p),(p-1/4)\}$, $q<1/2$ then $A(\rho_B,W)=\{1/2+(1-2p)/(1-2q)\}$, if $q=p=1/2$ then $A(\rho_B,W)=[0,1]$, otherwise $A(\rho_B,W)=\varphi$, where φ denotes the empty set. We now compute the asymptotic performance described in Theorem 1. We have,

$$I(X;B_R)_\sigma = S(\rho_B) - rS(W_0) - (1-r)S(W_1), \tag{2}$$

where $\sigma^{XB_R}\triangleq r|0\rangle\langle 0|\otimes W_0 + (1-r)|1\rangle\langle 1|\otimes W_1$. Figure 3 shows the lossy QC source compression rate curve for source ρ_B with $p=0.4$ and 0.5. Note that the curve decreases monotonically with q, as expected.

Example 2: (Lossy Classical Source Coding for Binary Source with Binary Symmetric Channel (BSC) as Posterior Channel) Consider a source $P_X\sim$ Bernoulli(p), $\hat{X}=\{0,1\}$, $W_{X|\hat{X}}\sim$ BSC(q), and $p,q\in[0,1/2]$. Toward identifying the set A, we assume $P_{\hat{X}}(0)=r$, which characterizes the set A, and solve the following system of linear equations: $p = r(1-q)+(1-r)q$, $(1-p) = rq+(1-r)(1-q)$. This gives, if $0\leq q\leq\min\{p,(1-p)\}$, $q<1/2$ then $A(P_X,W_{X|\hat{X}})=\{(p-q)/(1-2q)\}$, if $q=p=1/2$ then $A(P_X,W_{X|\hat{X}})=[0,1]$, otherwise $A(P_X,W_{X|\hat{X}})=\varphi$, where φ denotes the empty set. We now compute the asymptotic performance described in Theorem 2. We have,

$$I(X;\hat{X}) = H(X) - H(X|\hat{X}) = h_b(p) - h_b(q). \tag{3}$$

Observe that the rate in (3) is identical to the rate-distortion function for a Bernoulli(p) source with Hamming distortion criterion for $D<p$ [20, Theorem 10.3.1].

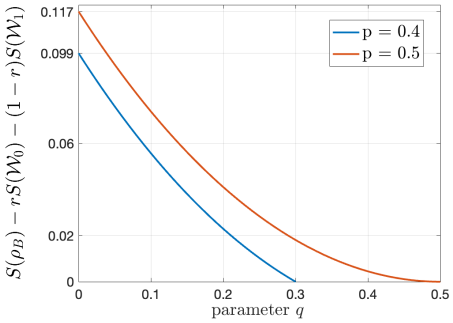

Fig. 3. Example for Lossy QC Source Coding for Binary Quantum Source with Binary Symmetric Posterior CQ Channel.

V. Proof of Theorem 1

A. Proof of Achievability

Let $\Omega=\rho_B^{\otimes n}$. For a given (ρ_B,X,W) QC source coding setup, we choose a reconstruction distribution $P_X\in A(\rho_B,W)$. Codebook Design: We generate a codebook C consisting of n-length codewords by randomly and independently selecting 2^{nR} sequences $\{X^n(m)\}_{m\in[2^{nR}]}$ according to the following pruned distribution:

$$P(X^n(m)=x^n) = \begin{cases}\frac{P_X^n(x^n)}{(1-\varepsilon)} & \text{for } x^n\in T_\delta^{(n)}(X)\\ 0 & \text{otherwise,}\end{cases} \tag{4}$$

where $P_X^n(x^n)=\prod_{i=1}^n P_X(x_i)$, $T_\delta^{(n)}(X)$ is the δ-typical set corresponding to the distribution P_X on the set X, and $\varepsilon(\delta,n)\triangleq\sum_{x^n\notin T_\delta^{(n)}(X)} P_X^n(x^n)$. Note that $\varepsilon(\delta,n)\searrow 0$ as $n\to\infty$ and for all sufficiently small $\delta>0$.

Construction of POVM: We use Winter's POVM construction [17]. Let π_{ρ_B} and π_{x^n} denote the δ-typical and conditional δ-typical projectors defined as in [21, Def. 15.1.3] and [21, Def. 15.2.4], with respect to ρ_B and W, respectively. Let $\hat{\pi}$ be the cut-off projector onto the subspaces spanned by the eigenstates of ξ with eigenvalues greater than ϵd, where $d\triangleq 2^{-n(H(\rho_B)+\delta_1)}$ and δ_1 will be specified later. Consider the following positive operators with a trace of less than one, and we exploit the random selection of these operators to construct the sub-POVM $\{A_{x^n}\}$. For all $x^n\in T_{\hat{\delta}}^{(n)}(X)$, define

$$\tilde{\rho}_{x^n}\triangleq\hat{\pi}\pi_{\rho_B}\pi_{x^n}W_{x^n}\pi_{x^n}\pi_{\rho_B}\hat{\pi} \quad\text{and}\quad \tilde{\rho}\triangleq E_P[\tilde{\rho}_{x^n}], \tag{5}$$

and $\tilde{\rho}_{x^n} = 0$ for $x^n \notin T_\delta^{(n)}(X)$. Using the Average Gentle Measurement Lemma [21, Lemma 9.4.3], for any given $\epsilon \in (0,1)$, and all sufficiently large n and all sufficiently small δ, we have

$$\sum_{x^n \in X^n} \frac{P_X^n(x^n)}{1-\varepsilon} \left\| \tilde{\rho}_{x^n} - W_{x^n} \right\|_1 \leq \epsilon, \tag{6}$$

Detailed proof of the above statement can be found in [22, Eq. 35]. Using the above definitions, for all $x^n \in X^n$, we construct the operators, $A_{x^n} \triangleq \gamma_{x^n} \Omega^{-1/2} \tilde{\rho}_{x^n} \Omega^{-1/2}$, where $\gamma_{x^n} \triangleq 2^{-nR\frac{(1-\varepsilon)}{(1+\eta)}} |\{m : X^n(m) = x^n\}|$ and $\eta \in (0,1)$. Let $1_{\{sP\}}$ denote the indicator random variable corresponding to the event that $\{A_{x^n} : x^n \in T_\delta^{(n)}(X)\}$ forms a sub-POVM. If $1_{\{sP\}} = 1$, then construct sub-POVM $\Gamma^{(n)}$ as follows: $\Gamma^{(n)} = \{A_{x^n} : x^n \in T_\delta^{(n)}(X)\}$. We then add an extra operator $A_{x_0^n} \triangleq I - \sum_{x^n \in T_\delta^{(n)}(X)} A_{x^n}$, associated with an arbitrary sequence $x_0^n \in X^n \backslash T_\delta^{(n)}(X)$, to form a valid POVM $\Gamma^{(n)}$ with at most $(2^{nR}+1)$ elements. If $1_{\{sP\}} = 0$, then we define $\Gamma^{(n)} = \{I\}$ and associate it with x_0^n. This defines the POVM and the associated decoder. We now provide a proposition from [17], which will be helpful later in the analysis.

Proposition 1: For all $\epsilon, \eta \in (0,1)$, for all sufficiently small $\delta > 0$, and for all sufficiently large n, we have $E\, 1_{\{sP\}} \geq 1 - \epsilon$, if $R > I(X; B_R)_\sigma$, where the quantum mutual information is computed with respect to the CQ state, $\sigma^{XB_R} \triangleq \sum_x P_X(x)|x\rangle\langle x|_X \otimes W_x$, and $\{|x\rangle\}$ is an orthonormal basis for the Hilbert space H_X with $\dim(H_X) = |X|$.

Error Analysis: We begin by splitting the error $\Xi(\Gamma^{(n)})$ into two terms using the indicator function $1_{\{sP\}}$ as

$$\begin{aligned}\Xi(\Gamma^{(n)}) &= 1_{\{sP\}}\Xi(\Gamma^{(n)}) + (1 - 1_{\{sP\}})\,\Xi(\Gamma^{(n)}),\\ &\leq 1_{\{sP\}}\Xi(\Gamma^{(n)}) + 2\,(1 - 1_{\{sP\}}),\end{aligned} \tag{7}$$

where (7) follows from upper bounding the trace distance between two density operators by its maximum value of two. Using the triangle inequality, we now expand $\Xi(\Gamma^{(n)})$. Under the condition $1_{\{sP\}} = 1$, $\Xi(\Gamma^{(n)}) \leq \zeta + 2\tilde{\zeta}$, where $\zeta \triangleq \sum_{x^n \in T_\delta^{(n)}(X)} \left\| \sqrt{\Omega} A_{x^n} \sqrt{\Omega} - \mathrm{Tr}(A_{x^n}\Omega) W_{x^n} \right\|_1$, $\tilde{\zeta} \triangleq \mathrm{Tr}(A_{x_0^n}\Omega)$. The error terms ζ and $\tilde{\zeta}$ capture the error induced by covering and by not covering the n-tensored posterior reference state, respectively. The analysis of these error terms is provided in [23]. Below, we summarize the results obtained from bounding these error terms. For all $\epsilon \in (0,1)$, for all sufficiently small $\delta > 0$, and sufficiently large n, we have $E[1_{\{sP\}}\zeta] \leq \epsilon$, for all sufficiently small $\eta > 0$, and $E[1_{\{sP\}}\tilde{\zeta}] \leq 2\epsilon$, for all $\eta \in (0,1)$. Now, we bound $E\,\Xi(\Gamma^{(n)}$, for all $\epsilon \in (0,1)$, $E[\Xi(\Gamma^{(n)})] \leq E[1_{\{sP\}}\Xi(\Gamma^{(n)})] + 2\epsilon \leq 6\epsilon$. Since $E\,\Xi(\Gamma^{(n)}) \leq 6\epsilon$, there exists a codebook C and the associated POVM $\Gamma^{(n)}$ such that $\Xi(\Gamma^{(n)}) \leq 6\epsilon$. This completes the achievability proof.

B. Proof of Converse

Let R be an achievable rate. Then from Definition 4, given a triple (ρ_B, X, W), for all $\epsilon > 0$, and all sufficiently large n, there exists (n, Θ) QC lossy compression protocol with a POVM $\Gamma^{(n)} = \{A_m\}_{m \in [\Theta]}$ and a decoding map f that satisfies the following constraint: $\sum_{x^n} \left\| \sqrt{\Omega} A_{f^{-1}(m)} \sqrt{\Omega} - \mathrm{Tr}(A_{f^{-1}(x^n)}\Omega) W_{x^n} \right\|_1 \leq \epsilon$, and $\frac{1}{n}\log\Theta \leq R + \epsilon$. Let M denote the transmitted message, and define the following classical-quantum state: $\omega^{X^n B_R^n} \triangleq \sum_{x^n} |x^n\rangle\langle x^n| \otimes \sqrt{\Omega} A_{f^{-1}(m)} \sqrt{\Omega}$ and $\tau^{X^n B_R^n} = \sum_{m,x^n} \frac{1}{n} \mathrm{Tr}(A_{f^{-1}(x^n)}\Omega)|x^n\rangle\langle x^n| \otimes W_{x^n}$. Here $\omega^{X^n B_R^n}$ and $\tau^{X^n B_R^n}$ are the resulting CQ-states of the QC lossy compression protocol and the ideal QC lossy compression protocol according to Definition 4, respectively. By triangle equality, we have $\left\| \omega^{X^n B_R^n} - \tau^{X^n B_R^n} \right\|_1 \leq \epsilon$. We now provide a lower bound on the rate R. We have the following inequalities:

$$\begin{aligned} nR &= \log\Theta - n\epsilon \geq H(M) - n\epsilon \geq I(M; B_R^n)_\omega - n\epsilon\\ &\overset{a}{\geq} I(X^n; B_R^n)_\omega - n\epsilon \overset{b}{\geq} nS(B_R)_\omega - \sum_{i=1}^n S((B_R)_i|X_i)_\omega - n\epsilon\\ &\overset{c}{\geq} nI(X; B_R)_{\omega^Q} - n\epsilon \overset{d}{\geq} nI(X; B_R)_{\tau^Q} - n\tilde{\epsilon}(\epsilon) - n\epsilon, \end{aligned}$$

where inequalities are argued as follows: (a) follows from the quantum data processing inequality [21, Section 11.9.2], (b) follows from the fact that conditioning does not increase quantum entropy, (c) follows from the concavity of conditional quantum entropy [21, Ex. 11.7.5] and by defining $\omega^{X^Q(B_R)^Q} \triangleq \frac{1}{n}\sum_{i=1}^n \mathrm{Tr}_{X^n\backslash i}(B_R)^{n\backslash i}(\omega^{X^n B_R^n})$, and noting that $\omega^{(B_R)^Q} = \rho_B$, and (d) follows from the continuity of quantum mutual information (AFW inequality) [21, Ex. 11.10.2], by defining $\tau^{X^Q(B_R)^Q} \triangleq \frac{1}{n}\sum_{i=1}^n \mathrm{Tr}_{X^n\backslash i}(B_R)^{n\backslash i}(\tau^{X^n B_R^n}) = \sum_x P_{X^Q}(x)|x\rangle\langle x| \otimes W_x$, $\tilde{\epsilon} \triangleq \frac{3}{2}\epsilon\log(\dim H_B) + (2+\epsilon)h_b\left(\frac{\epsilon}{2+\epsilon}\right)$, and observing

$$\begin{aligned}\left\| \rho_B - \sum_x P_{X^Q}(x) W_x \right\|_1 &\leq \left\| \omega^{X^Q(B_R)^Q} - \tau^{X^Q(B_R)^Q} \right\|_1\\ &\leq \left\| \omega^{X^n B_R^n} - \tau^{X^n B_R^n} \right\|_1 \leq \epsilon,\end{aligned} \tag{8}$$

where $P_{X^Q}(x) \triangleq \left(\frac{1}{n}\sum_{i=1}^n \sum_{x^{n\backslash i}} \mathrm{Tr}\{A_{f^{-1}(x^n)}\Omega\}\right)$. We note that $\sum_x P_{X^Q}(x) = 1$. So far, we have shown that $R \in \bigcap_{\epsilon>0} I_\epsilon$, where we have defined for all $\epsilon \geq 0$, $I_\epsilon(\rho_B, W) \triangleq \{R : \exists\, P_X \in A_\epsilon$ such that $R \geq I(X, B_R)_\sigma - g(\epsilon)\}$, $A_\epsilon(\rho_B, W) \triangleq \{P_X \in \mathcal{P}(X) : \left\| \sum_x P_X(x) W_x - \rho_B \right\|_1 \leq \epsilon\}$, and $\sigma^{XB_R} \triangleq \sum_x P_X(x)|x\rangle\langle x| \otimes W_x$, $g(\epsilon) = \tilde{\epsilon} + n\epsilon$. Equation (8) ensures that the set A_ϵ is non-empty for $\epsilon > 0$. Using the continuity of rate-regions I_ϵ at $\epsilon = 0$ (similar to [23, Lemma 5]), we obtain $R \in I_0$. This concludes the converse proof.

VI. Proof of Theorem 2

A. Proof of Achievability

For a given $(P_X, \hat{X}, W_{X|\hat{X}})$ source coding setup, we choose a reconstruction distribution $P_{X'} \in A(P_X, W_{X|X'})$. From now on, we let $\Theta = 2^{nR} + 1$.

Codebook Construction We construct a codebook $\mathcal{C} \triangleq \{\hat{X}^n(1), \hat{X}^n(2), \cdots, \hat{X}^n(2^{nR})\}$, by choosing each codewords randomly and independently according to the following "pruned" distribution:

$$P(\hat{X}^n(m) = \hat{x}^n) = \begin{cases} \frac{P_{\hat{X}}^n(\hat{x}^n)}{1-\epsilon} & \text{if } \hat{x}^n \in T_\delta^{(n)}(\hat{X}), \\ 0 & \text{otherwise.} \end{cases}$$

where $P_{\hat{X}}^n(\hat{x}^n) = \prod_{i=1}^n P_{\hat{X}}(\hat{x}_i)$, $T_\delta^{(n)}(\hat{X})$ is the δ-typical set corresponding to $P_{\hat{X}}$, and $\epsilon(\delta, n) \triangleq \sum_{\hat{x}^n \notin T_\delta^{(n)}(\hat{X})} P_{\hat{X}}^n(\hat{x}^n)$.

Encoder Description For an observed source sequence x^n, construct a randomized encoder that chooses an index $m \in [2^{nR}]$ according to a sub-PMF $E_{M|X^n}(m|x^n)$, which is analogous to the likelihood encoders used in [18], [19]. We now specify $E_{M|X^n}(m|x^n)$ for $x^n \in T_{\hat{\delta}}^{(n)}(X)$ and $m \in [2^{nR}]$, where $\hat{\delta} = \delta(|\mathcal{X}| + |\hat{\mathcal{X}}|)$. For a $\eta \in (0,1)$ (to be specified later), and $\delta > 0$, define

$$E_{M|X^n}(m|x^n) \triangleq \sum_{\hat{x}^n} \frac{1}{2^{nR}} \frac{(1-\epsilon)}{(1+\eta)} \frac{W_{X|\hat{X}}^n(x^n|\hat{x}^n)}{P_X^n(x^n)} \times 1_{\{\hat{x}^n \in T_\delta^{(n)}(\hat{X})\}} 1_{\{x^n \in T_\delta^{(n)}(X|\hat{x}^n)\}} 1_{\{\hat{X}^n(m) = \hat{x}^n\}},$$

Let $1_{\{\text{SPMF}\}}$ denotes the indicator random variable corresponding to the event that $\{E_{M|X^n}(m|x^n)\}_{m \in [\Theta]}$ forms a sub-PMF for all $x^n \in T_\delta^{(n)}(X)$. If $1_{\{\text{SPMF}\}} = 1$, then construct the sub-PMF $P_{M|X^n}(m|x^n) = E_{M|X^n}(m|x^n)$, for all $x^n \in T_\delta^{(n)}(X)$ and $m \in [\Theta]$. We then add an additional PMF element $P_{M|X^n}(0|x^n) = E_{M|X^n}(0|x^n) \triangleq 1 - \sum_{m \in [2^{nR}]} E_{M|X^n}(m|x^n)$ for all $x^n \in T_\delta^{(n)}(X)$, associated with $m = 0$, to form a valid PMF $P_{M|X^n}(m|x^n)$ for all $x^n \in T_\delta^{(n)}(X)$ and $m \in \{0\} \cup [2^{nR}]$. If $x^n \notin T_\delta^{(n)}(X)$, then we define $P_{M|X^n}(m|x^n) = 1_{\{m=0\}}$. If $1_{\{\text{SPMF}\}} = 0$, then $P_{M|X^n}(m|x^n) = 1_{\{m=0\}}$, for all $x^n \in \mathcal{X}^n$. We provide a proposition from [19], which will be helpful later in the analysis.

Proposition 2: For all $\epsilon, \eta \in (0,1)$, for all sufficiently small $\delta > 0$, and sufficiently large n, we have $E\left[1_{\{\text{SPMF}\}}\right] \geq 1 - \epsilon$, if $R > I(X; \hat{X})$.

Decoder Description: For an observed index $m \in \{0\} \cup [2^{nR}]$ communicated by the encoder, the decoder outputs $\hat{X}^n(m)$ if $m = 0$. Otherwise, decoder outputs a fixed $\hat{x}_0^n \in \hat{\mathcal{X}}^n \backslash T_\delta^{(n)}(\hat{X})$, i.e., $D(m) = \hat{X}^n(m)$ if $m = 0$, otherwise $D(m) = \hat{x}_0^n$.

Error Analysis We begin by splitting the error $\Xi(E, D)$ into two terms using the indicator function $1_{\{\text{SPMF}\}}$ as

$$\begin{aligned} \Xi(E,D) &= 1_{\{\text{SPMF}\}} \Xi(E,D) + (1 - 1_{\{\text{SPMF}\}}) \Xi(E,D), \\ &\leq 1_{\{\text{SPMF}\}} \Xi(E,D) + (1 - 1_{\{\text{SPMF}\}}), \end{aligned} \tag{9}$$

where (9) follows from upper bounding the total variation between two PMFs by one. Using the triangle inequality, we now expand $\Xi(E, D)$. Under the condition $1_{\{\text{SPMF}\}} = 1$, $2\,\Xi(E,D) \leq \zeta + 2\tilde{\zeta} + 3\epsilon$, for all sufficiently large n and all $\delta > 0$, where $\zeta \triangleq$

$$\sum_{\substack{m \in [2^{nR}] \\ \hat{x}^n, x^n \in T_\delta^{(n)}(X)}} P_X^n(x^n) E_{M|X^n}(m|x^n) - P_M(m) W_{\hat{X}|X}^n(x^n|\hat{x}^n) \times 1_{\{\hat{X}^n(m) = \hat{x}^n\}}, \quad \tilde{\zeta} \triangleq \sum_{x^n \notin T_\delta^{(n)}(X)} P_X^n(x^n) E_{M|X^n}(0|x^n).$$

The error terms ζ and $\tilde{\zeta}$ capture the error induced by covering and not covering the n-product source sequence, respectively. Below, we summarize the bounds on these error terms. For detailed analysis, we refer to [23]. For all $\epsilon \in (0,1)$, for all sufficiently small $\delta > 0$, and all sufficiently large n, we have $E[1_{\{\text{SPMF}\}}\tilde{\zeta}] \leq \epsilon$, for all sufficiently small $\eta > 0$, and $E[1_{\{\text{SPMF}\}}\zeta] \leq 2\epsilon$, for all $\eta \in (0,1)$. Therefore, we get, for all $\epsilon \in (0,1)$, $E[\Xi(E,D)] \leq E[1_{\{\text{SPMF}\}}\Xi(E,D)] + \epsilon \leq 9\epsilon/2$. Since $E[\Xi(E,D)] \leq 9\epsilon/2$, there exists a code $\mathcal{C}$ such that the associated $\Xi(E,D) \leq 9\epsilon/2$. This completes the achievability proof.

B. Proof of Converse

Let R be an achievable rate. Then from Definition 7, given a triple $(P_X, \hat{X}, W_{X|\hat{X}})$, for all $\epsilon > 0$, and for all sufficiently large n, there exists (n, Θ) lossy compression protocol with an encoding map E and a decoding map D that satisfy the following constraints: $\Xi(E,D) = \|P_{X^nX^n} - P_{\hat{X}^n} W_{X|\hat{X}}^n\|_{TV} \leq \epsilon$, and $\frac{1}{n}\log\Theta \leq R + \epsilon$. Let M denote the transmitted message. We now provide a lower bound on the rate R. We have the following inequalities:

$$\begin{aligned} nR = \log\Theta - n\epsilon &\geq H(M) - n\epsilon \geq I(X^n, M) - n\epsilon \\ &\overset{a}{\geq} I(X^n, \hat{X}^n) - n\epsilon \overset{b}{\geq} \sum_i I(X_i; \hat{X}_i) - n\epsilon \overset{c}{\geq} nI(X_Q; \hat{X}_Q) - n\epsilon \\ &\overset{d}{=} nI(P_{\hat{X}_Q}, P_{X_Q|\hat{X}_Q}) - n\epsilon \overset{e}{\geq} nI(P_{\hat{X}_Q}, W_{X|\hat{X}}) - n\tilde{\epsilon}(\epsilon) - n\epsilon, \end{aligned}$$

where (a) follows from the data processing inequality, (b) follows from the property that conditioning reduces entropy, (c) follows from the convexity of mutual information as the function of varying channel for a fixed source, and by defining $P_{\hat{X}_Q} = \frac{1}{n}\sum_{i=1}^n P_{\hat{X}_i}$ and noting that $P_{X_Q} = P_X$, (d) follows from the change of notation of mutual information [3], and (e) follows from the continuity of mutual information [20, Theorem 17.3.3], by defining $\tilde{\epsilon} \triangleq -2\epsilon\log 4\epsilon^2/(|\mathcal{X}|^2|\mathcal{X}|)$, and observing [23, Lemma 6]: $\|P_X - \sum_{\hat{x}} P_{\hat{X}_Q}(\hat{x}) W_{X|\hat{X}}(\cdot|\hat{x})\|_{TV} \leq \|P_{X_Q X_Q} - P_{\hat{X}_Q} W_{X|\hat{X}}\|_{TV} \leq \|P_{X^nX^n} - P_{\hat{X}^n} W_{X|\hat{X}}^n\|_{TV}$.

So far, we have shown that $R \geq \bigcap_{\epsilon > 0} I_\epsilon$, where we have defined for all $\epsilon \geq 0$, $I_\epsilon(P_X, W_{X|\hat{X}}) \triangleq \{R : \exists P_{\hat{X}} \in A_\epsilon(P_X, W_{X|\hat{X}})$ such that $R \geq I(P_{\hat{X}}, W_{X|\hat{X}}) - g(\epsilon)\}$, $A_\epsilon(P_X, W_{X|\hat{X}}) \triangleq \{P_{\hat{X}} \in \mathcal{P}(\hat{\mathcal{X}}) : \|\sum_{\hat{x}} P_{\hat{X}}(\hat{x}) W_{X|\hat{X}}(\cdot|\hat{x}) - P_X\|_{TV} \leq \epsilon\}$, $g(\epsilon) \triangleq \tilde{\epsilon} + \epsilon$. [23, Lemma 6] ensures that the set A_ϵ is non-empty for $\epsilon > 0$. Using the continuity of rate regions similar to [23, Lemma 5], we obtain $\bigcap_{\epsilon > 0} I_\epsilon = I_0$ and A_0 is non-empty, and hence $R \geq I_0$. This concludes the converse proof.

REFERENCES

[1] N. Datta, M.-H. Hsieh, M. M. Wilde, and A. Winter, "Quantum-to-classical rate distortion coding," Journal of Mathematical Physics, vol. 54, no. 4, p. 042201, 2013.

[2] C. E. Shannon et al., "Coding theorems for a discrete source with a fidelity criterion," IRE Nat. Conv. Rec, vol. 4, no. 142-163, p. 1, 1959.

[3] I. Csiszár and J. Körner, Information theory: coding theorems for discrete memoryless systems. Cambridge University Press, 2011.

[4] T. Berger, "Rate distortion theory and data compression," in Advances in Source Coding. Springer, 1975, pp. 1–39.

[5] R. G. Gallager, Information theory and reliable communication. Springer, 1968, vol. 588.

[6] T. Berger, Rate Distortion Theory: A Mathematical Basis for Data Compression, ser. Prentice-Hall electrical engineering series. Prentice-Hall, 1971.

[7] A. M. Gerrish, "Estimation of information rates," Ph.D. dissertation, Yale University, 1963.

[8] S. S. Pradhan, "Approximation of test channels in source coding," in Proc. Conf. Inform. Syst. Sci.(CISS), 2004.

[9] A. Kanlis, S. Khudanpur, and P. Narayan, "Typicality of a good rate-distortion code," Problems of Information Transmission, vol. 32, no. 1, pp. 96–103, 1996.

[10] T. Weissman and E. Ordentlich, "The empirical distribution of rate-constrained source codes," IEEE transactions on information theory, vol. 51, no. 11, pp. 3718–3733, 2005.

[11] P. W. Cuff, H. H. Permuter, and T. M. Cover, "Coordination capacity," IEEE Transactions on Information Theory, vol. 56, no. 9, pp. 4181–4206, 2010.

[12] C. Schieler and P. Cuff, "A connection between good rate-distortion codes and backward dmcs," in 2013 IEEE Information Theory Workshop (ITW). IEEE, 2013, pp. 1–5.

[13] V. Kostina and S. Verdú, "The output distribution of good lossy source codes," in 2015 Information Theory and Applications Workshop (ITA). IEEE, 2015, pp. 308–312.

[14] R. O. Duda, P. E. Hart et al., Pattern classification. John Wiley & Sons, 2006.

[15] M. K. Steven, "Fundamentals of statistical signal processing," PTR Prentice-Hall, Englewood Cliffs, NJ, vol. 10, p. 151045, 1993.

[16] L. Chennuru Vankadara and U. von Luxburg, "Measures of distortion for machine learning," Advances in Neural Information Processing Systems, vol. 31, 2018.

[17] A. Winter, ""Extrinsic" and "intrinsic" data in quantum measurements: asymptotic convex decomposition of positive operator valued measures," Communication in Mathematical Physics, vol. 244, no. 1, pp. 157–185, 2004.

[18] P. Cuff, "Distributed channel synthesis," IEEE Transactions on Information Theory, vol. 59, no. 11, pp. 7071–7096, 2013.

[19] T. A. Atif, A. Padakandla, and S. S. Pradhan, "Source coding for synthesizing correlated randomness," in 2020 IEEE International Symposium on Information Theory (ISIT), 2020, pp. 1570–1575.

[20] T. M. Cover and J. A. Thomas, "Elements of information theory 2nd edition (wiley series in telecommunications and signal processing)," Acessado em, 2006.

[21] M. M. Wilde, "From classical to quantum shannon theory," arXiv preprint arXiv:1106.1445, 2011.

[22] M. M. Wilde, P. Hayden, F. Buscemi, and M.-H. Hsieh, "The information-theoretic costs of simulating quantum measurements," Journal of Physics A: Mathematical and Theoretical, vol. 45, no. 45, p. 453001, 2012.

[23] T. A. Atif, M. A. Sohail, and S. S. Pradhan, "Lossy quantum source coding with a global error criterion based on a posterior reference map," arXiv preprint arXiv:2302.00625, 2023.