PHYS 7895: Quantum Information Theory Fall 2015

Lecture 21 — November 4, 2015

Prof. Mark M. Wilde *Scribe: Mark M. Wilde*

1 Overview

In this lecture, we now move on to protocols in quantum Shannon theory, beginning with the simplest one: quantum data compression.

2 Introduction

One of the fundamental tasks in classical information theory is the compression of information. Given access to many uses of a noiseless classical channel, what is the best that a sender and receiver can make of this resource for compressed data transmission? Shannon's compression theorem demonstrates that the Shannon entropy is the fundamental limit for the compression rate in the i.i.d. setting. That is, if one compresses at a rate above the Shannon entropy, then it is possible to recover the compressed data perfectly in the asymptotic limit, and otherwise, it is not possible to do so.[1] This theorem establishes the prominent role of the entropy in Shannon's theory of information.

In the quantum world, it very well could be that one day a sender and a receiver would have many uses of a noiseless quantum channel available,[2] and the sender could use this resource to transmit compressed quantum information. But what exactly does this mean in the quantum setting? A simple model of a quantum information source is an ensemble of quantum states $\{p_X(x), |\psi_x\rangle\}$, i.e., the source outputs the state $|\psi_x\rangle$ with probability $p_X(x)$, and the states $\{|\psi_x\rangle\}$ do not necessarily have to form an orthonormal basis. Let us suppose for the moment that the classical data x is available as well, even though this might not necessarily be the case in practice. A naive strategy for compressing this quantum information source would be to ignore the quantum states coming out, handle the classical data instead, and exploit Shannon's compression protocol. That is, the sender compresses the sequence x^n emitted from the quantum information source at a rate equal to the Shannon entropy $H(X)$, sends the compressed classical bits over the noiseless quantum channels, the receiver reproduces the classical sequence x^n at his end, and finally reconstructs the sequence $|\psi_{x^n}\rangle$ of quantum states corresponding to the classical sequence x^n.

The above strategy will certainly work, but it makes no use of the fact that the noiseless quantum channels are quantum! It is clear that noiseless quantum channels will be expensive in practice, and the above strategy is wasteful in this sense because it could have merely exploited classical

[1] Technically, we did not prove the converse part of Shannon's data-compression theorem, but the converse of this chapter suffices for Shannon's classical theorem as well.

[2] How we hope so! If working, coherent fault-tolerant quantum computers come along one day, they stand to benefit from quantum compression protocols.

channels (channels that cannot preserve superpositions) to achieve the same goals. Schumacher compression is a strategy that makes effective use of noiseless quantum channels to compress a quantum information source down to a rate equal to the von Neumann entropy. This has a great benefit from a practical standpoint—recall from the exercises that the von Neumann entropy of a quantum information source is strictly lower than the source's Shannon entropy if the states in the ensemble are non-orthogonal. In order to execute the protocol, the sender and receiver simply need to know the density operator $\rho \equiv \sum_x p_X(x)|\psi_x\rangle\langle\psi_x|$ of the source. Furthermore, Schumacher compression is provably optimal in the sense that any protocol that compresses a quantum information source of the above form at a rate below the von Neumann entropy cannot have a vanishing error in the asymptotic limit.

Schumacher compression thus gives an operational interpretation of the von Neumann entropy as the fundamental limit on the rate of quantum data compression. Also, it sets the term "qubit" on a firm foundation in an information-theoretic sense as a measure of the amount of quantum information "contained" in a quantum information source.

We begin this chapter by giving the details of the general information-processing task corresponding to quantum data compression. We then prove that the von Neumann entropy is an achievable rate of compression and follow by showing that it is optimal (these two respective parts are the direct coding theorem and the converse theorem for quantum data compression). We illustrate how much savings one can gain in quantum data compression by detailing a specific example. The final section of the chapter closes with a presentation of more general forms of Schumacher compression.

3 The Typical Subspace

Our first task is to establish the notion of a quantum information source. It is analogous to the notion of a classical information source, in the sense that the source randomly outputs a quantum state according to some probability distribution, but the states that it outputs do not necessarily have to be distinguishable as in the classical case.

Definition 1 (Quantum Information Source)**.** *A quantum information source is some device that randomly emits pure qudit states in a Hilbert space $\mathcal{H}_A$ of finite dimension.*

We use the symbol A to denote the quantum system for the quantum information source. Suppose that the source outputs states $|\psi_y\rangle$ randomly according to some probability distribution $p_Y(y)$. Note that the states $|\psi_y\rangle$ do not necessarily have to form an orthonormal set. Then the density operator ρ_A of the source is the expected state emitted:

$$\rho_A \equiv \mathbb{E}_Y \left\{|\psi_Y\rangle\langle\psi_Y|_A\right\} = \sum_y p_Y(y)|\psi_y\rangle\langle\psi_y|_A. \tag{1}$$

There are many decompositions of a density operator as a convex sum of rank-one projectors (and the above decomposition is one such example), but perhaps the most important decomposition is a spectral decomposition of the density operator ρ:

$$\rho_A = \sum_{x\in\mathcal{X}} p_X(x)|x\rangle\langle x|_A. \tag{2}$$

The above states $|x\rangle_A$ are eigenvectors of ρ_A and form a complete orthonormal basis for Hilbert space $\mathcal{H}_A$, and the non-negative, convex real numbers $p_X(x)$ are the eigenvalues of ρ_A.

We have written the states $|x\rangle_A$ and the eigenvalues $p_X(x)$ in a suggestive notation because it is actually possible to think of our quantum source as a classical information source—the emitted states $\{|x\rangle_A\}_{x\in\mathcal{X}}$ are orthonormal and each corresponding eigenvalue $p_X(x)$ acts as a probability for choosing $|x\rangle_A$. We can say that our source is classical because it is emitting the orthogonal, and thus distinguishable, states $|x\rangle_A$ with probability $p_X(x)$. This description is equivalent to the ensemble $\{p_Y(y), |\psi_y\rangle\}_y$ because the two ensembles lead to the same density operator (recall that two ensembles that have the same density operator are essentially equivalent because they lead to the same probabilities for outcomes of any measurement performed on the system). Our quantum information source then corresponds to the pure-state ensemble:

$$\{p_X(x), |x\rangle_A\}_{x\in\mathcal{X}}\,. \tag{3}$$

Recall that the von Neumann entropy $H(A)$ of the density operator ρ_A is as follows (Definition **??**):

$$H(A)_\rho \equiv -\operatorname{Tr}\{\rho_A \log \rho_A\}\,. \tag{4}$$

It is straightforward to show that the von Neumann entropy $H(A)_\rho$ is equal to the Shannon entropy $H(X)$ of a random variable X with distribution $p_X(x)$ because the basis states $|x\rangle_A$ are orthonormal.

Suppose now that the quantum information source emits a large number n of random quantum states so that the density operator describing the emitted state is as follows:

$$\rho_{A^n} \equiv \underbrace{\rho_{A_1}\otimes\cdots\otimes\rho_{A_n}}_{n} = (\rho_A)^{\otimes n}. \tag{5}$$

The labels $A_1, \ldots, A_n$ denote the Hilbert spaces corresponding to the different quantum systems, but the density operator is the same for each quantum system $A_1, \ldots, A_n$ and is equal to ρ_A. The above description of a quantum source is within the i.i.d. setting for the quantum domain. A spectral decomposition of the state in (5) is as follows:

$$\begin{aligned}
\rho_{A^n} &= \sum_{x_1\in\mathcal{X}} p_X(x_1)|x_1\rangle\langle x_1|_{A_1}\otimes\cdots\otimes\sum_{x_n\in\mathcal{X}} p_X(x_n)|x_n\rangle\langle x_n|_{A_n} &(6)\\
&= \sum_{x_1,\cdots,x_n\in\mathcal{X}} p_X(x_1)\cdots p_X(x_n)\left(|x_1\rangle\cdots|x_n\rangle\right)\left(\langle x_1|\cdots\langle x_n|\right)_{A_1,\ldots,A_n} &(7)\\
&= \sum_{x^n\in\mathcal{X}^n} p_{X^n}(x^n)|x^n\rangle\langle x^n|_{A^n}, &(8)
\end{aligned}$$

where we employ the shorthand:

$$p_{X^n}(x^n) \equiv p_X(x_1)\cdots p_X(x_n), \qquad |x^n\rangle_{A^n} \equiv |x_1\rangle_{A_1}\cdots|x_n\rangle_{A_n}. \tag{9}$$

The above quantum description of the density operator is essentially equivalent to the classical picture of n realizations of random variable X with each eigenvalue $p_{X_1}(x_1)\cdots p_{X_n}(x_n)$ acting as a probability because the set of states $\{|x_1\rangle\cdots|x_n\rangle_{A_1,\ldots,A_n}\}_{x_1,\cdots,x_n\in\mathcal{X}}$ is an orthonormal set.

We can now "quantize" or extend the notion of typicality to the quantum information source. The definitions follow directly from the classical definitions. The quantum definition of typicality can employ either the weak notion or the strong notion. We do not distinguish the notation for a typical subspace and a typical set because it should be clear from the context which kind of typicality we are employing.

Definition 2 (Typical Subspace)**.** *The δ-typical subspace $T^{\delta}_{A^n}$ is a subspace of the full Hilbert space $\mathcal{H}_{A^n} = \mathcal{H}_{A_1} \otimes \cdots \otimes \mathcal{H}_{A_n}$, associated with many copies of a density operator, such as the one in* (2)*. It is spanned by states $|x^n\rangle_{A^n}$ whose corresponding classical sequences x^n are δ-typical:*

$$T^{\delta}_{A^n} \equiv \text{span}\left\{|x^n\rangle_{A^n} : x^n \in T^{X^n}_{\delta}\right\}, \tag{10}$$

where it is implicit that the typical subspace $T^{\delta}_{A^n}$ on the left-hand side is with respect to a density operator ρ and the typical set $T^{X^n}_{\delta}$ on the right-hand side is with respect to the distribution $p_X(x)$ from the spectral decomposition of ρ in (2)*. We could also denote the typical subspace as $T^{\rho,\delta}_{A^n}$ if we would like to make the dependence of the space on ρ more explicit.*

3.1 The Typical Subspace Measurement

The definition of the typical subspace (Definition 2) gives a way to divide up the Hilbert space of n qudits into two subspaces: the typical subspace and the atypical subspace. The properties of the typical subspace are similar to what we found for the properties of typical sequences. That is, the typical subspace is exponentially smaller than the full Hilbert space of n qudits, yet it contains nearly all of the probability (in a sense that we show below). The intuition for these properties of the typical subspace is the same as it is classically, once we have a spectral decomposition of a density operator.

The *typical projector* is a projector onto the typical subspace, and the complementary projector projects onto the atypical subspace. These projectors play an important operational role in quantum Shannon theory because we can construct a quantum measurement from them. That is, this measurement is the best way of asking the question, "Is the state typical or not?" because it minimally disturbs the state while still retrieving this one bit of information.

Definition 3 (Typical Projector)**.** *Let $\Pi^{\delta}_{A^n}$ denote the typical projector for the typical subspace of a density operator ρ_A with spectral decomposition in* (2)*. It is a projector onto the typical subspace:*

$$\Pi^{\delta}_{A^n} \equiv \sum_{x^n \in T^{X^n}_{\delta}} |x^n\rangle\langle x^n|_{A^n}, \tag{11}$$

where it is implicit that the x^n below the summation is a classical sequence in the typical set $T^{X^n}_{\delta}$, and the state $|x^n\rangle$ is a quantum state given in (9) *and associated with the classical sequence x^n via the spectral decomposition of ρ in* (2)*. We can also denote the typical projector as $\Pi^{\rho,\delta}_{A^n}$ if we would like to make its dependence on ρ explicit.*

The action of multiplying the density operator ρ_{A^n} by the typical projector $\Pi^{\delta}_{A^n}$ is to select out all the basis states of ρ_{A^n} that are in the typical subspace and form a "sliced" operator $\tilde{\rho}_{A^n}$ that is close to the original density operator ρ_{A^n}:

$$\tilde{\rho}_{A^n} \equiv \Pi^{\delta}_{A^n} \rho_{A^n} \Pi^{\delta}_{A^n} = \sum_{x^n \in T^{X^n}_{\delta}} p_{X^n}(x^n) |x^n\rangle\langle x^n|_{A^n}. \tag{12}$$

That is, the effect of projecting a state onto the typical subspace $T^{\delta}_{A^n}$ is to "slice" out any component of the state ρ_{A^n} that does not lie in the typical subspace $T^{\delta}_{A^n}$.

Exercise 4. *Show that the typical projector $\Pi^{\delta}_{A^n}$ commutes with the density operator ρ_{A^n}:*

$$\rho_{A^n}\Pi^{\delta}_{A^n} = \Pi^{\delta}_{A^n}\rho_{A^n}. \tag{13}$$

The typical projector allows us to formulate an operational method for delicately asking the question: "Is the state typical or not?" We can construct a quantum measurement that consists of two outcomes: the outcome "1" reveals that the state is in the typical subspace, and "0" reveals that it is not. This typical subspace measurement is often one of the first important steps in most protocols in quantum Shannon theory.

Definition 5 (Typical Subspace Measurement)**.** *The following map is a quantum instrument that realizes the typical subspace measurement:*

$$\sigma \rightarrow \left(I - \Pi^{\delta}_{A^n}\right)\sigma\left(I - \Pi^{\delta}_{A^n}\right) \otimes |0\rangle\langle 0| + \Pi^{\delta}_{A^n}\sigma\Pi^{\delta}_{A^n} \otimes |1\rangle\langle 1|, \tag{14}$$

where σ is some density operator acting on the Hilbert space $\mathcal{H}_{A^n}$. It associates a classical register with the outcome of the measurement—the value of the classical register is $|0\rangle$ for the support of the state σ that is not in the typical subspace, and it is equal to $|1\rangle$ for the support of the state σ that is in the typical subspace.

The implementation of a typical subspace measurement is currently far from the reality of what is experimentally accessible if we would like to have the measure concentration effects necessary for proving many of the results in quantum Shannon theory. Recall that we required a sequence of about a million bits in order to have the needed measure concentration effects. We would need a similar number of qubits emitted from a quantum information source, and furthermore, we would require the ability to perform noiseless coherent operations over about a million or more qubits in order to implement the typical subspace measurement. Such a daunting requirement firmly places quantum Shannon theory as a "highly theoretical theory," rather than being a theory that can make close connection to current experimental practice.[3]

3.2 Properties of the Typical Subspace

The typical subspace $T^{\delta}_{A^n}$ enjoys several useful properties that are "quantized" versions of the typical sequence properties:

Property 6 (Unit Probability)**.** *Suppose that we perform a typical subspace measurement of a state ρ_{A^n}. Then the probability that the quantum state ρ_{A^n} is in the typical subspace $T^{\delta}_{A^n}$ approaches one as n becomes large:*

$$\text{Tr}\left\{\Pi^{\delta}_{A^n}\rho_{A^n}\right\} \geq 1 - \varepsilon, \tag{15}$$

for all $\varepsilon \in (0,1)$, $\delta > 0$, and sufficiently large n, where $\Pi^{\delta}_{A^n}$ is the typical subspace projector from Definition 3.

[3] We should note that this was certainly the case as well for information theory when Claude Shannon developed it in 1948, but in the many years since then, there has been much progress in the development of practical classical codes for achieving the classical capacity of a classical channel.

Property 7 (Exponentially Smaller Dimension)**.** *The dimension* $\dim(T^{\delta}_{A^n})$ *of the* δ*-typical subspace is exponentially smaller than the dimension* $|\mathcal{X}|^n$ *of the entire space of quantum states when the output of the quantum information source is not maximally mixed. We formally state this property as follows:*

$$\text{Tr}\left\{\Pi^{\delta}_{A^n}\right\} \leq 2^{n(H(A)+c\delta)}, \tag{16}$$

where c *is some positive constant that depends on whether we employ the weak or strong notion of typicality. We can also bound the dimension* $\dim(T^{\delta}_{A^n})$ *of the* δ*-typical subspace from below:*

$$\text{Tr}\left\{\Pi^{\delta}_{A^n}\right\} \geq (1-\varepsilon)\, 2^{n(H(A)-c\delta)}, \tag{17}$$

for all $\varepsilon \in (0,1)$*,* $\delta > 0$*, and sufficiently large* n*.*

Property 8 (Equipartition)**.** *The operator* $\Pi^{\delta}_{A^n}\rho_{A^n}\Pi^{\delta}_{A^n}$ *corresponds to a "slicing" of the density operator* ρ_{A^n} *where we slice out and keep only the part with support in the typical subspace. We can then bound all of the eigenvalues of the sliced operator* $\Pi^{\delta}_{A^n}\rho_{A^n}\Pi^{\delta}_{A^n}$ *as follows:*

$$2^{-n(H(A)+c\delta)}\Pi^{\delta}_{A^n} \leq \Pi^{\delta}_{A^n}\rho_{A^n}\Pi^{\delta}_{A^n} \leq 2^{-n(H(A)-c\delta)}\Pi^{\delta}_{A^n}. \tag{18}$$

The above inequality is an operator inequality. It is a statement about the eigenvalues of the operators $\Pi^{\delta}_{A^n}\rho_{A^n}\Pi^{\delta}_{A^n}$ *and* $\Pi^{\delta}_{A^n}$*, and these operators have the same eigenvectors because they commute. Therefore, the above inequality is equivalent to the following inequality that applies in the classical case:*

$$\forall x^n \in T^{X^n}_{\delta} : 2^{-n(H(A)+c\delta)} \leq p_{X^n}(x^n) \leq 2^{-n(H(A)-c\delta)}. \tag{19}$$

This equivalence holds because each probability $p_{X^n}(x^n)$ *is an eigenvalue of* $\Pi^{\delta}_{A^n}\rho_{A^n}\Pi^{\delta}_{A^n}$*.*

The dimension $\dim(T^{\delta}_{A^n})$ of the δ-typical subspace is approximately equal to the dimension $|\mathcal{X}|^n$ of the entire space only when the density operator of the quantum information source is maximally mixed because

$$\text{Tr}\left\{\Pi^{\delta}_{A^n}\right\} \leq |\mathcal{X}|^n \cdot 2^{n\delta} \simeq |\mathcal{X}|^n. \tag{20}$$

The proofs of the above properties are essentially identical to those from the classical case in Sections **??** and **??**, regardless of whether we employ a weak or strong notion of quantum typicality. We leave the proofs as the three exercises below.

Exercise 9. *Prove the unit probability property of the* δ*-typical subspace (Property 6). First show that the probability that many copies of a density operator is in the* δ*-typical subspace is equal to the probability that a random sequence is* δ*-typical:*

$$\text{Tr}\left\{\Pi^{\delta}_{A^n}\rho_{A^n}\right\} = \Pr\left\{X^n \in T^{X^n}_{\delta}\right\}. \tag{21}$$

Exercise 10. *Prove the exponentially smaller dimension property of the* δ*-typical subspace (Property 7). First show that the trace of the typical projector* $\Pi^{\delta}_{A^n}$ *is equal to the dimension of the typical subspace* $T^{\delta}_{A^n}$*:*

$$\dim(T^{\delta}_{A^n}) = \text{Tr}\left\{\Pi^{\delta}_{A^n}\right\}. \tag{22}$$

Then prove the property.

Exercise 11. *Prove the equipartition property of the* δ*-typical subspace (Property 8). First show that*

$$\Pi^{\delta}_{A^n}\rho_{A^n}\Pi^{\delta}_{A^n} = \sum_{x^n \in T^{X^n}_{\delta}} p_{X^n}(x^n)|x^n\rangle\langle x^n|_{A^n}, \tag{23}$$

and then argue the proof.

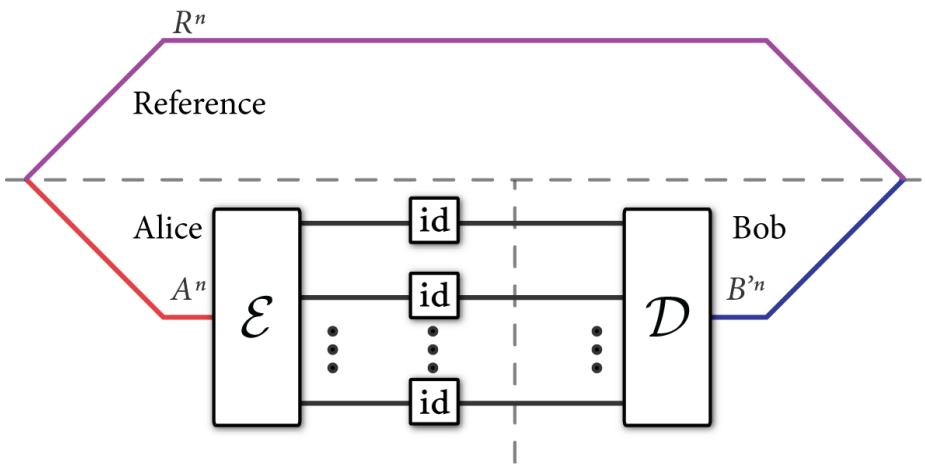

Figure 1: The most general protocol for quantum compression. Alice begins with the output of some quantum information source whose density operator is $\rho^{\otimes n}$ on some system A^n. The inaccessible reference system holds the purification of this density operator. She performs some CPTP encoding map $\mathcal{E}$, sends the compressed qubits through 2^{nR} uses of a noiseless quantum channel, and Bob performs some CPTP decoding map $\mathcal{D}$ to decompress the qubits. The scheme is successful if the difference between the initial state and the final state is negligible in the asymptotic limit $n \to \infty$.

4 The Information-Processing Task

We first overview the general task that any quantum compression protocol attempts to accomplish. Three parameters n, R, and ε corresponding to the length of the original quantum data sequence, the rate, and the error, respectively, characterize any such protocol. An $(n, R+\delta, \varepsilon)$ quantum compression code consists of four steps: state preparation, encoding, transmission, and decoding. Figure 1 depicts a general protocol for quantum compression.

State Preparation. The quantum information source outputs a sequence $|\psi_{x^n}\rangle_{A^n}$ of quantum states according to the ensemble $\{p_X(x), |\psi_x\rangle\}$ where

$$|\psi_{x^n}\rangle_{A^n} \equiv |\psi_{x_1}\rangle_{A_1} \otimes \cdots \otimes |\psi_{x_n}\rangle_{A_n}. \tag{24}$$

The density operator, from the perspective of someone ignorant of the classical sequence x^n, is equal to the tensor power state $\rho^{\otimes n}$ where

$$\rho \equiv \sum_x p_X(x)|\psi_x\rangle\langle\psi_x|. \tag{25}$$

Also, we can think about the purification of the above density operator. That is, an equivalent mathematical picture is to imagine that the quantum information source produces states of the form

$$|\varphi_\rho\rangle_{RA} \equiv \sum_x \sqrt{p_X(x)}|x\rangle_R|\psi_x\rangle_A, \tag{26}$$

where R is the label for an inaccessible reference system (not to be confused with the rate R!). The resulting i.i.d. state produced is $(|\varphi_\rho\rangle_{RA})^{\otimes n}$.

Encoding. Alice encodes the systems A^n according to some CPTP compression map $\mathcal{E}_{A^n \to W}$ where W is a quantum system of size 2^{nR}. Recall that R is the rate of compression:

$$R = \frac{1}{n} \log d_W - \delta, \tag{27}$$

where d_W is the dimension of system W and δ is a small positive number.

Transmission. Alice transmits the system W to Bob using $n\left(R+\delta\right)$ noiseless qubit channels.

Decoding. Bob sends the system W through a decompression map $\mathcal{D}_{W \to \hat{A}^n}$.

The protocol has ε error if the compressed and decompressed state is ε-close in trace distance to the original state $(|\varphi_\rho\rangle_{RA})^{\otimes n}$:

$$\left\| \left(\varphi^\rho_{RA}\right)^{\otimes n} - (\mathcal{D}_{W \to \hat{A}^n} \circ \mathcal{E}_{A^n \to W})(\left(\varphi^\rho_{RA}\right)^{\otimes n}) \right\|_1 \leq \varepsilon. \tag{28}$$

We say that a quantum compression rate R is *achievable* if there exists an $(n, R+\delta, \varepsilon)$ quantum compression code for all $\delta, \varepsilon > 0$ and sufficiently large n.

5 The Quantum Data-Compression Theorem

Schumacher's compression theorem establishes the von Neumann entropy as the fundamental limit on quantum data compression.

Theorem 12 (Quantum Data Compression)**.** *Suppose that ρ_A is the density operator of the quantum information source. Then the von Neumann entropy $H\left(A\right)_\rho$ is the smallest achievable rate R for quantum data compression:*

$$\inf\left\{R : R \text{ is achievable}\right\} = H(A)_\rho. \tag{29}$$

5.1 The Direct Coding Theorem

Schumacher's compression protocol demonstrates that the von Neumann entropy $H(A)_\rho$ is an achievable rate for quantum data compression. It is remarkably similar to Shannon's compression protocol, but it has some subtle differences that are necessary for the quantum setting. The basic steps of the encoding are to perform a typical subspace measurement and an isometry that compresses the typical subspace. The decoder then performs the inverse of the isometry to decompress the state. The protocol is successful if the typical subspace measurement successfully projects onto the typical subspace, and it fails otherwise. Just like in the classical case, the law of large numbers guarantees that the protocol is successful in the asymptotic limit as $n \to \infty$. Figure 2 provides an illustration of the protocol, and we now provide a rigorous argument.

Alice begins with n copies of the state $\left(\varphi^\rho_{RA}\right)^{\otimes n}$. Suppose that a spectral decomposition of ρ is as follows:

$$\rho = \sum_z p_Z(z) |z\rangle\langle z|, \tag{30}$$

where $p_Z(z)$ is some probability distribution, and $\{|z\rangle\}$ is some orthonormal basis. Her first step $\mathcal{E}^1_{A^n \to Y A^n}$ is to perform a typical subspace measurement of the form in (5) onto the typical subspace

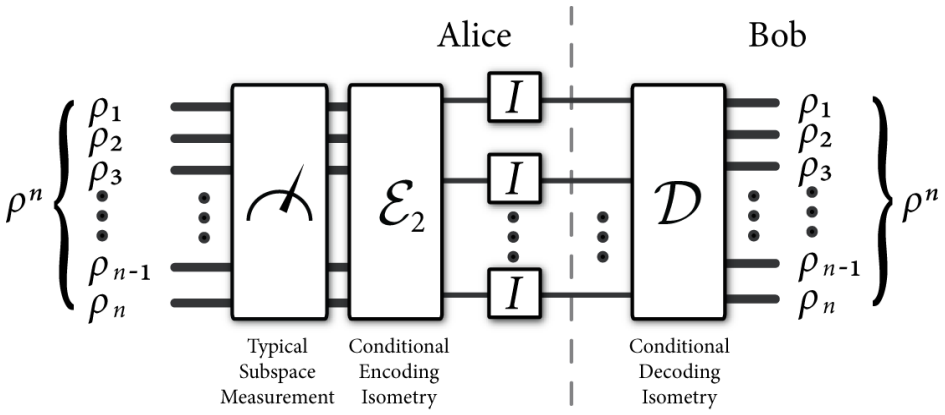

Figure 2: Schumacher's compression protocol. Alice begins with many copies of the output of the quantum information source. She performs a measurement onto the typical subspace corresponding to the state ρ and then performs a compression isometry of the typical subspace to a space of size $2^{n[H(\rho)+\delta]}$ qubits. She transmits these compressed qubits over $n\left[H(\rho)+\delta\right]$ uses of a noiseless quantum channel. Bob performs the inverse of the isometry to uncompress the qubits. The protocol is successful in the asymptotic limit due to the properties of typical subspaces.

of A^n, where the typical projector is with respect to the density operator ρ. The action of $\mathcal{E}^1_{A^n\to YA^n}$ on a general state σ_{A^n} is

$$\mathcal{E}^1_{A^n\to YA^n}(\sigma_{A^n}) \equiv |0\rangle\langle 0|_Y \otimes \left(I - \Pi^\delta_{A^n}\right)\sigma_{A^n}\left(I - \Pi^\delta_{A^n}\right) + |1\rangle\langle 1|_Y \otimes \Pi^\delta_{A^n}\sigma_{A^n}\Pi^\delta_{A^n}, \quad (31)$$

and the classically correlated flag bit Y indicates whether the typical subspace projection $\Pi^\delta_{A^n}$ is successful or unsuccessful. Recall from the Shannon compression protocol in Section **??** that we exploited an invertible function f that mapped from the set of typical sequences to a set of binary sequences $\{0,1\}^{n[H(\rho)+\delta]}$. Now, we can construct a linear map U_f that is a coherent version of this classical function f. It simply maps the orthonormal basis $\{|z^n\rangle_{A^n}\}$ to the basis $\{|f(z^n)\rangle_W\}$:

$$U_f \equiv \sum_{z^n\in T^{Z^n}_\delta} |f(z^n)\rangle_W \langle z^n|_{A^n}, \quad (32)$$

where Z is a random variable corresponding to the distribution $p_Z(z)$ so that $T^{Z^n}_\delta$ is its typical set. The inverse of the above operator is an isometry because the input space $\text{span}\left\{|z^n\rangle_{A^n} : z^n \in T^{Z^n}_\delta\right\}$ is a subspace of size at most $2^{n[H(\rho)+\delta]}$ (recall Property 7) embedded in a larger space of size 2^n (at least for qubits) and the output space is of size at most $2^{n[H(\rho)+\delta]}$. So her next step $\mathcal{E}^2_{YA^n\to YW}$ is to perform the compression conditional on the flag bit Y being equal to one and otherwise declaring an error. The action of $\mathcal{E}^2_{YA^n\to YW}$ on a general classical–quantum state

$$\sigma_{YA^n} \equiv |0\rangle\langle 0|_Y \otimes \sigma^0_{A^n} + |1\rangle\langle 1|_Y \otimes \sigma^1_{A^n} \quad (33)$$

such that $\sigma^0_{A^n} = \left(I - \Pi^\delta_{A^n}\right)\sigma^0_{A^n}\left(I - \Pi^\delta_{A^n}\right)$ and $\sigma^1_{A^n} = \Pi^\delta_{A^n}\sigma^1_{A^n}\Pi^\delta_{A^n}$ is as follows:

$$\mathcal{E}^2_{YA^n\to YW}(\sigma_{YA^n}) \equiv |0\rangle\langle 0|_Y \otimes \text{Tr}\left\{\sigma^0_{A^n}\right\}|e\rangle\langle e|_W + |1\rangle\langle 1|_Y \otimes U_f\sigma^1_{A^n}U_f^\dagger, \quad (34)$$

where $|e\rangle_W$ is some error flag orthogonal to all of the states $\{|f(\phi_{x^n})\rangle_W\}_{\phi_{x^n}\in T_\delta^{Z^n}}$. This last step completes the details of her encoder $\mathcal{E}_{A^n\to YW}$, and the action of it on the initial state is

$$\mathcal{E}_{A^n\to YW}(\left(\varphi_{RA}^{\rho}\right)^{\otimes n}) \equiv (\mathcal{E}^2_{YA^n\to YW}\circ\mathcal{E}^1_{A^n\to YA^n})(\left(\varphi_{RA}^{\rho}\right)^{\otimes n}). \tag{35}$$

Alice then transmits all of the compressed qubits over $n\left[H(\rho)+\delta\right]+1$ uses of the noiseless qubit channel.

Bob's decoding $\mathcal{D}_{YW\to A^n}$ performs the inverse of the linear map U_f conditional on the flag bit being equal to one and otherwise maps to some other state $|e\rangle_{A^n}$ outside of the typical subspace. The action of the decoder on some general classical–quantum state

$$\sigma_{YW} \equiv |0\rangle\langle 0|_Y \otimes \sigma_W^0 + |1\rangle\langle 1|_Y \otimes \sigma_W^1 \tag{36}$$

is

$$\mathcal{D}^1_{YW\to YA^n}(\sigma_{YW}) \equiv |0\rangle\langle 0|_Y \otimes \mathrm{Tr}\left\{\sigma_W^0\right\}|e\rangle\langle e|_{A^n} + |1\rangle\langle 1|_Y \otimes U_f^\dagger \sigma_W^1 U_f. \tag{37}$$

The final part of the decoder is to discard the classical flag bit: $\mathcal{D}^2_{YA^n\to A^n} \equiv \mathrm{Tr}_Y\{\cdot\}$. Then $\mathcal{D}_{YW\to A^n} \equiv \mathcal{D}^2_{YA^n\to A^n}\circ\mathcal{D}^1_{YW\to YA^n}$.

We now can analyze how this protocol performs with respect to our performance criterion in (28). Consider the following chain of inequalities:

$$\begin{aligned}
&\left\|\left(\varphi_{RA}^{\rho}\right)^{\otimes n} - \left(\mathcal{D}_{YW\to A^n}\circ\mathcal{E}_{A^n\to YW}\right)\left(\left(\varphi_{RA}^{\rho}\right)^{\otimes n}\right)\right\|_1 \\
&= \left\|\mathrm{Tr}_Y\left\{|1\rangle\langle 1|_Y \otimes \left(\varphi_{RA}^{\rho}\right)^{\otimes n}\right\} - \left(\mathcal{D}_{YW\to A^n}\circ\mathcal{E}_{A^n\to YW}\right)\left(\left(\varphi_{RA}^{\rho}\right)^{\otimes n}\right)\right\|_1 && (38)\\
&\leq \left\||1\rangle\langle 1|_Y \otimes \left(\varphi_{RA}^{\rho}\right)^{\otimes n} - \left(\mathcal{D}^1_{YW\to YA^n}\circ\mathcal{E}_{A^n\to YW}\right)\left(\left(\varphi_{RA}^{\rho}\right)^{\otimes n}\right)\right\|_1 && (39)\\
&= \left\| \begin{array}{c} |1\rangle\langle 1|_Y \otimes \left(\varphi_{RA}^{\rho}\right)^{\otimes n} - \\ \left(\begin{array}{c} |0\rangle\langle 0|_Y \otimes \mathrm{Tr}\left\{\left(I - \Pi_{A^n}^{\delta}\right)\left(\varphi_{RA}^{\rho}\right)^{\otimes n}\right\}|e\rangle\langle e|_{A^n} \\ +|1\rangle\langle 1|_Y \otimes \Pi_{A^n}^{\delta}\left(\varphi_{RA}^{\rho}\right)^{\otimes n}\Pi_{A^n}^{\delta} \end{array} \right) \end{array} \right\|_1 . && (40)
\end{aligned}$$

The first equality follows by adding a flag bit $|1\rangle_Y$ to $\left(\varphi_{RA}^{\rho}\right)^{\otimes n}$ and tracing it out. The first inequality follows from monotonicity of trace distance under the discarding of subsystems (Corollary **??**). The second equality follows by evaluating the map $\mathcal{D}^1_{YW\to YA^n}\circ\mathcal{E}_{A^n\to YW}$ on the state $\left(\varphi_{RA}^{\rho}\right)^{\otimes n}$. Continuing, we have

$$\begin{aligned}
&\leq \left\||1\rangle\langle 1|_Y \otimes \left(\varphi_{RA}^{\rho}\right)^{\otimes n} - |1\rangle\langle 1|_Y \otimes \Pi_{A^n}^{\delta}\left(\varphi_{RA}^{\rho}\right)^{\otimes n}\Pi_{A^n}^{\delta}\right\|_1 \\
&\qquad + \left\||0\rangle\langle 0|_Y \otimes \mathrm{Tr}\left\{\left(I - \Pi_{A^n}^{\delta}\right)\left(\varphi_{RA}^{\rho}\right)^{\otimes n}\right\}|e\rangle\langle e|_{A^n}\right\|_1 && (41)\\
&= \left\|\left(\varphi_{RA}^{\rho}\right)^{\otimes n} - \Pi_{A^n}^{\delta}\left(\varphi_{RA}^{\rho}\right)^{\otimes n}\Pi_{A^n}^{\delta}\right\|_1 + \mathrm{Tr}\left\{\left(I - \Pi_{A^n}^{\delta}\right)\left(\varphi_{RA}^{\rho}\right)^{\otimes n}\right\} && (42)\\
&\leq 2\sqrt{\varepsilon} + \varepsilon. && (43)
\end{aligned}$$

The first inequality follows from the triangle inequality for trace distance (Lemma **??**). The equality uses the facts $\|\rho\otimes\sigma - \omega\otimes\sigma\|_1 = \|\rho-\omega\|_1\|\sigma\|_1 = \|\rho-\omega\|_1$ and $\|b\rho\|_1 = |b|\,\|\rho\|_1$ for some density operators ρ, σ, and ω and a constant b. The final inequality follows from the first property of typical subspaces:

$$\mathrm{Tr}\left\{\Pi_{A^n}^{\delta}\left(\varphi_{RA}^{\rho}\right)^{\otimes n}\right\} = \mathrm{Tr}\left\{\Pi_{A^n}^{\delta}\rho^{\otimes n}\right\} \geq 1-\varepsilon, \tag{44}$$

and the gentle operator lemma.

We remark that it is important for the typical subspace measurement in (31) to be implemented as a non-destructive quantum measurement. That is, the only information that this measurement should learn is whether the state is typical or not. Otherwise, there would be too much disturbance to the quantum information, and the protocol would fail at the desired task of compression. Such precise control on so many qubits is possible in principle, but it is rather daunting to implement in practice!

5.2 The Converse Theorem

We now prove the converse theorem for quantum data compression by considering the most general compression protocol that meets the success criterion in (28) and demonstrating that such an asymptotically error-free protocol should have its rate of compression above the von Neumann entropy of the source. Alice would like to compress a state $\rho^{\otimes n}$ that acts on a Hilbert space A^n. A purification $\phi_{R^n A^n} \equiv \left(\varphi^{\rho}_{RA}\right)^{\otimes n}$ of this state represents the state of the joint systems A^n and R^n where R^n is the purifying system (again, we should not confuse reference system R^n with rate R). If she can compress any system on A^n and recover it faithfully, then she should be able to do so for the purification of the state. An $(n, R+\delta, \varepsilon)$ compression code has the property that it can compress at a rate $R+\delta$ with only error ε. The quantum data processing is

$$A^n \quad \xrightarrow{\mathcal{E}_{A^n \to W}} \quad W \quad \xrightarrow{\mathcal{D}_{W \to \hat{A}^n}} \quad \hat{A}^n \tag{45}$$

and the following inequality holds for a quantum compression protocol with error ε:

$$\left\| \omega_{R^n \hat{A}^n} - \left(\varphi^{\rho}_{RA}\right)^{\otimes n} \right\|_1 \leq \varepsilon, \tag{46}$$

where

$$\omega_{R^n \hat{A}^n} \equiv \mathcal{D}_{W \to \hat{A}^n}(\mathcal{E}_{A^n \to W}(\left(\varphi^{\rho}_{RA}\right)^{\otimes n})). \tag{47}$$

Consider the following chain of inequalities:

$$\begin{aligned}
2 \log |W| &\geq I\left(W; R^n\right)_{\omega} && (48)\\
&\geq I(\hat{A}^n; R^n)_{\omega} && (49)\\
&\geq I(A^n; R^n)_{\phi} - f(\varepsilon, n) && (50)\\
&= H(A^n)_{\phi} + H\left(R^n\right)_{\phi} - H\left(A^n R^n\right)_{\phi} - f(\varepsilon, n) && (51)\\
&= 2H(A^n)_{\phi} - f(\varepsilon, n). && (52)
\end{aligned}$$

The first inequality is a consequence of a dimension bound for the quantum mutual information $I\left(E; F\right) \leq 2 \log \left(\min\left\{|E|, |F|\right\}\right)$ and the fact that $|W| = 2^{nR}$. The second inequality follows from the quantum data-processing inequality (Bob processes W with the decoder to get $\hat{A}^n$). The third inequality follows from applying the Alicki–Fannes inequality to the success criterion in (46) and setting $f(\varepsilon, n) \equiv n 6 \varepsilon R + 4 h_2(\varepsilon)$. This function has the property that $\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} f(\varepsilon, n) = 0$. The second equality is from the definition of quantum mutual information, and the last equality follows because the entropies of each half of a pure, bipartite state are equal and their joint entropy vanishes. Given that the state ϕ is an i.i.d. state of the form $\left(\varphi^{\rho}_{RA}\right)^{\otimes n}$, the von Neumann entropy is additive so that

$$H(A^n)_{(\varphi^{\rho})^{\otimes n}} = n H(A)_{\varphi^{\rho}}. \tag{53}$$

Putting everything together, we find that

$$R + \delta = \frac{1}{n} \log |W| \geq H(A)_{\varphi^\rho} - \frac{1}{2n} f(\varepsilon, n). \tag{54}$$

Taking the limit as $n \to \infty$ and $\varepsilon, \delta \to 0$ allows us to conclude that an achievable rate R of quantum data compression necessarily satisfies $R \geq H(A)_{\varphi^\rho}$.