PHYS 7895: Quantum Information Theory Fall 2015

Lecture 22 — November 9, 2015

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

1 Overview

In the last lecture, we developed Schumacher quantum data compression. The essential idea for the achievability part was that performing a typical subspace measurement on many copies of the same state does not disturb it too much, while at the same time projecting the state into a subspace which is exponentially smaller than the whole Hilbert space. The converse part relied on quantum data processing.

In this lecture, we develop a protocol known as entanglement concentration, which allows for concentrating many copies of the same pure bipartite state to some number of copies of maximally entangled ebits. We will see that the optimal rate of conversion is given by a quantity known as the entropy of entanglement.

2 Introduction

Entanglement is one of the most useful resources in quantum information processing. If a sender and receiver share noiseless entanglement in the form of maximally entangled states, then we previously showed how they can teleport quantum bits between each other with the help of classical communication, or they can double the capacity of a noiseless qubit channel for transmitting classical information. We will see further applications later on in which they can exploit noiseless entanglement to assist in the transmission of classical or quantum data over a noisy quantum channel.

Given the utility of maximal entanglement, a reasonable question is to ask what a sender and receiver can accomplish if they share pure entangled states that are not maximally entangled. In the quantum Shannon-theoretic setting, we make the further assumption that the sender and receiver can share many copies of these pure entangled states. We find out in this chapter that they can "concentrate" these non-maximally entangled states to maximally entangled ebits, and the optimal rate at which they can do so in the asymptotic limit is equal to the "entropy of entanglement" (the von Neumann entropy of half of one copy of the original state). Entanglement concentration is thus another fundamental task in noiseless quantum Shannon theory, and it gives a different operational interpretation to the von Neumann entropy.

Entanglement concentration is perhaps complementary to Schumacher compression in the sense that it gives a firm quantum information-theoretic interpretation of the term "ebit" (just as Schumacher compression did for the term "qubit"), and it plays a part in demonstrating how the entropy of entanglement is the unique measure of entanglement for pure bipartite states. Despite the similarity to Schumacher compression in this respect, entanglement concentration is a fundamentally different protocol, and we will see that these two protocols are not interchangeable. That is, exploiting the Schumacher compression protocol for the task of entanglement concentration fails at accomplishing the goal of entanglement concentration, and vice versa.

The technique for proving that the von Neumann entropy is an achievable rate for entanglement concentration exploits the method of types for classical and quantum typicality, respectively (the most important property is one which states that the exponentiated entropy is a lower bound on the size of a typical type class). In hindsight, it is perhaps surprising that a typical type class is exponentially large in the large n limit (on the same order as the typical set itself), and we soon discover the quantum Shannon-theoretic consequences of this result.

We begin this chapter by discussing a simple example of entanglement concentration for a finite number of copies of a state. Section 4 then details the information-processing task that entanglement concentration attempts to accomplish, and Section 5 proves both the direct coding theorem and the converse theorem for entanglement concentration. We then discuss how shared randomness concentration is the closest classical analog of the entanglement concentration protocol. Finally, we discuss the differences between Schumacher compression and entanglement concentration, especially how exploiting one protocol to accomplish the other's information-processing task results in a failure of the intended goal.

3 An Example of Entanglement Concentration

A simple example illustrates the main idea underlying the concentration of entanglement. Consider the following partially entangled state:

$$|\Phi_\theta\rangle_{AB} \equiv \cos(\theta)\,|00\rangle_{AB} + \sin(\theta)|11\rangle_{AB}, \tag{1}$$

where θ is some parameter such that $0 < \theta < \pi/2$. The Schmidt decomposition guarantees that the above state is the most general form for a pure bipartite entangled state on qubits. Now suppose that Alice and Bob share three copies of the above state. We can rewrite the three copies of the above state with some straightforward algebra:

$$\begin{aligned}
&|\Phi_\theta\rangle_{A_1B_1}|\Phi_\theta\rangle_{A_2B_2}|\Phi_\theta\rangle_{A_3B_3} \\
&= \cos^3(\theta)|000\rangle_A|000\rangle_B + \sin^3(\theta)|111\rangle_A|111\rangle_B \qquad (2)\\
&\quad + \cos(\theta)\sin^2(\theta)\left(|110\rangle_A|110\rangle_B + |101\rangle_A|101\rangle_B + |011\rangle_A|011\rangle_B\right) \\
&\quad + \cos^2(\theta)\sin(\theta)\left(|100\rangle_A|100\rangle_B + |010\rangle_A|010\rangle_B + |001\rangle_A|001\rangle_B\right) \\
&= \cos^3(\theta)|000\rangle_A|000\rangle_B + \sin^3(\theta)|111\rangle_A|111\rangle_B \qquad (3)\\
&\quad + \sqrt{3}\cos(\theta)\sin^2(\theta)\frac{1}{\sqrt{3}}\left(|110\rangle_A|110\rangle_B + |101\rangle_A|101\rangle_B + |011\rangle_A|011\rangle_B\right) \\
&\quad + \sqrt{3}\cos^2(\theta)\sin(\theta)\frac{1}{\sqrt{3}}\left(|100\rangle_A|100\rangle_B + |010\rangle_A|010\rangle_B + |001\rangle_A|001\rangle_B\right),
\end{aligned}$$

where we relabel all of the systems on Alice and Bob's respective sides as $A \equiv A_1A_2A_3$ and $B \equiv B_1B_2B_3$. Observe that the subspace with coefficient $\cos^3(\theta)$ whose states have zero "ones" is one-dimensional. The subspace whose states have three "ones" is also one-dimensional. But the subspace with coefficient $\cos(\theta)\sin^2(\theta)$ whose states have two "ones" is three-dimensional, and the same holds for the subspace whose states each have one "one."

A protocol for entanglement concentration in this scenario is then straightforward. Alice performs a projective measurement consisting of the operators Π_0, Π_1, Π_2, Π_3 where

$$\Pi_0 \equiv |000\rangle\langle 000|_A, \tag{4}$$
$$\Pi_1 \equiv |001\rangle\langle 001|_A + |010\rangle\langle 010|_A + |100\rangle\langle 100|_A, \tag{5}$$
$$\Pi_2 \equiv |110\rangle\langle 110|_A + |101\rangle\langle 101|_A + |011\rangle\langle 011|_A, \tag{6}$$
$$\Pi_3 \equiv |111\rangle\langle 111|_A. \tag{7}$$

The subscript i of the projection operator Π_i corresponds to the Hamming weight of the basis states in the corresponding subspace. Bob can perform the same "Hamming weight" measurement on his side. With probability $\cos^6(\theta) + \sin^6(\theta)$, the procedure fails because it results in $|000\rangle_A|000\rangle_B$ or $|111\rangle_A|111\rangle_B$ which is not a maximally entangled state. But with probability $3\cos^2(\theta)\sin^4(\theta)$, the state is in the subspace with Hamming weight two, and it has the following form:

$$\frac{1}{\sqrt{3}}\left(|110\rangle_A|110\rangle_B + |101\rangle_A|101\rangle_B + |011\rangle_A|011\rangle_B\right), \tag{8}$$

and with probability $3\cos^4(\theta)\sin^2(\theta)$, the state is in the subspace with Hamming weight one, and it has the following form:

$$\frac{1}{\sqrt{3}}\left(|100\rangle_A|100\rangle_B + |010\rangle_A|010\rangle_B + |001\rangle_A|001\rangle_B\right). \tag{9}$$

Alice and Bob can then perform local operations on their respective systems to rotate either of these states to a maximally entangled state with Schmidt rank three:

$$\frac{1}{\sqrt{3}}\left(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B + |2\rangle_A|2\rangle_B\right). \tag{10}$$

The simple protocol outlined above is the basis for the entanglement concentration protocol, but it unfortunately fails with a non-negligible probability in this case. On the other hand, if we allow Alice and Bob to have a large number of copies of a pure bipartite entangled state, the probability of failing becomes negligible in the asymptotic limit due to the properties of typicality, and each type class subspace contains an exponentially large maximally entangled state. The proof of the direct coding theorem in the book makes this intuition precise.

4 The Information-Processing Task

We first detail the information-processing task that entanglement concentration sets out to accomplish. An (n, E, ε) entanglement concentration protocol consists of just one step of processing. Alice and Bob begin with many copies $(|\varphi\rangle_{AB})^{\otimes n}$ of a pure bipartite, entangled state $|\varphi\rangle_{AB}$. Alice and Bob each then perform local quantum channels $\mathcal{E}_{A^n\to\hat{A}}$ and $\mathcal{F}_{B^n\to\hat{B}}$ in an attempt to concentrate the original state $(|\varphi\rangle_{AB})^{\otimes n}$ to a maximally entangled state:

$$\omega_{\hat{A}\hat{B}} \equiv \left(\mathcal{E}_{A^n\to\hat{A}} \otimes \mathcal{F}_{B^n\to\hat{B}}\right)\left(\varphi_{A^nB^n}\right). \tag{11}$$

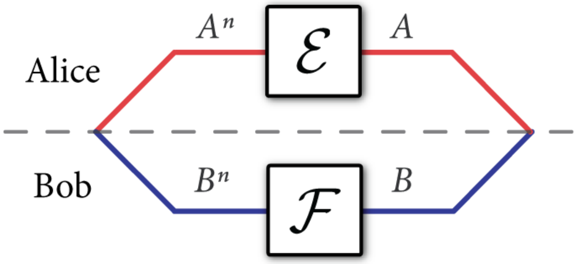

Figure 1: The most general protocol for entanglement concentration. Alice and Bob begin with many copies of some pure bipartite state $|\varphi\rangle_{AB}$. They then perform local operations to concentrate this state to a maximally entangled state.

The protocol has ε error if the final state $\omega_{\hat{A}\hat{B}}$ is ε-close to a maximally entangled state $|\Phi\rangle_{\hat{A}\hat{B}}$:

$$\frac{1}{2}\left\|\omega_{\hat{A}\hat{B}} - \Phi_{\hat{A}\hat{B}}\right\|_1 \leq \varepsilon, \tag{12}$$

where

$$|\Phi\rangle_{\hat{A}\hat{B}} \equiv \frac{1}{\sqrt{D}}\sum_{i=0}^{D-1} |i\rangle_{\hat{A}}|i\rangle_{\hat{B}}, \tag{13}$$

and the rate E of ebit extraction is

$$E = \frac{1}{n}\log_2(D), \tag{14}$$

where δ is some small positive number.

We say that a particular rate E of entanglement concentration is *achievable* if there exists an $(n, E-\delta, \varepsilon)$ entanglement concentration protocol for all $\varepsilon \in (0,1)$, $\delta > 0$, and sufficiently large n. The entanglement concentration limit for $|\varphi\rangle_{AB}$ is equal to the supremum of all achievable rates. Figure 1 displays the operation of a general entanglement concentration protocol.

5 The Entanglement Concentration Theorem

We first state the entanglement concentration theorem and then prove it below in two parts (the direct coding theorem and the converse theorem).

Theorem 1 (Entanglement Concentration)**.** *Let $|\varphi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ be a pure bipartite state. The entanglement concentration limit for $|\varphi\rangle_{AB}$ is equal to the entropy of entanglement, i.e., the quantum entropy $H(A)_\varphi$.*

5.1 The Direct Coding Theorem

We only prove this for a particular case to illustrate the main idea and point to the book for a full proof.

Generalizing the procedure outlined above to an arbitrary number of copies is straightforward. Suppose Alice and Bob share n copies of the partially entangled state $|\Phi_\theta\rangle$. We can then write the state as follows:

$$|\Phi_\theta\rangle_{A^nB^n} = \sum_{k=0}^{n} \sqrt{\binom{n}{k}} \cos^{n-k}(\theta)\sin^k(\theta) \left(\frac{1}{\sqrt{\binom{n}{k}}} \sum_{x\ :\ w(x)=k} |x\rangle_{A^n}|x\rangle_{B^n} \right), \tag{15}$$

where $w(x)$ is the Hamming weight of the binary vector x. Alice performs a "Hamming weight" measurement whose projective operators are as follows:

$$\Pi_k = \sum_{x\ :\ w(x)=k} |x\rangle\langle x|_{A^n}, \tag{16}$$

and the Schmidt rank of the maximally entangled state that they then share is $\binom{n}{k}$.

We can give a rough analysis of the performance of the above protocol when n becomes large by exploiting Stirling's approximation (we just need a handle on the term $\binom{n}{k}$ for large n). Recall that Stirling's approximation is $n! \approx \sqrt{2\pi n}\,(n/e)^n$, and this gives

$$\begin{aligned}
\binom{n}{k} &= \frac{n!}{k!n-k!} &(17)\\
&\approx \frac{\sqrt{2\pi n}\,(n/e)^n}{\sqrt{2\pi k}\,(k/e)^k\,\sqrt{2\pi\,(n-k)}\,((n-k)\,/e)^{n-k}} &(18)\\
&= \sqrt{\frac{n}{2\pi k\,(n-k)}}\frac{n^n}{(n-k)^{n-k}\,k^k} &(19)\\
&= \text{poly}\,(n)\ \left(\frac{n-k}{n}\right)^{-(n-k)}\left(\frac{k}{n}\right)^{-k} &(20)\\
&= \text{poly}\,(n)\ 2^{n[-((n-k)/n)\log((n-k)/n)-(k/n)\log(k/n)]} &(21)\\
&= \text{poly}\,(n)\ 2^{nh_2(k/n)}, &(22)
\end{aligned}$$

where h_2 is the binary entropy function and poly(n) indicates a term at most polynomial in n. When n is large, the exponential term $2^{nh_2(k/n)}$ dominates the polynomial $\sqrt{n/2\pi k\,(n-k)}$, so that the polynomial term begins to behave merely as a constant. So, the protocol is for Alice to perform a typical subspace measurement with respect to the distribution $\left(\cos^2(\theta), \sin^2(\theta)\right)$, and the state then collapses to the following one with high probability:

$$\frac{1}{\mathcal{N}} \sum_{\substack{k=0\ :\\ \left|k/n-\sin^2(\theta)\right|\ \leq\ \delta,\\ \left|(n-k)/n-\cos^2(\theta)\right|\ \leq\ \delta}}^{n} \sqrt{\binom{n}{k}} \cos^{n-k}(\theta)\sin^k(\theta) \left(\frac{1}{\sqrt{\binom{n}{k}}} \sum_{x\ :\ w(x)=k} |x\rangle_{A^n}|x\rangle_{B^n} \right), \tag{23}$$

where $\mathcal{N}$ is an appropriate normalization constant. Alice and Bob then both perform a Hamming weight measurement and the state collapses to a state of the form:

$$\frac{1}{\sqrt{\text{poly}\,(n)\ 2^{nh_2(k/n)}}} \sum_{x\ :\ w(x)=k} |x\rangle_{A^n}|x\rangle_{B^n}, \tag{24}$$

depending on the outcome k of the measurement. The above state is a maximally entangled state with Schmidt rank $\text{poly}(n)\ 2^{nh_2(k/n)}$, and it follows that

$$h_2\left(k/n\right) \geq h_2\left(\cos^2(\theta)\right) - \delta, \tag{25}$$

from the assumption that the state first projects into the typical subspace. Alice and Bob can then perform local operations to rotate this state to approximately $nh_2\left(\cos^2(\theta)\right)$ ebits. Thus, this procedure concentrates the original non-maximally entangled state to ebits at a rate equal to the entropy of entanglement of the state $|\Phi_\theta\rangle_{AB}$ in (1). The above proof is a bit rough, and it applies only to entangled qubit systems in a pure state. The direct coding theorem in the book generalizes this proof to pure entangled states on d-dimensional systems.

5.2 The Converse Theorem

We now prove the converse theorem for entanglement concentration, i.e., that the entanglement concentration limit for $|\varphi\rangle_{AB}$ does not exceed $H(A)_\varphi$. Alice and Bob begin with many copies of the pure state $|\varphi\rangle_{AB}$. In the most general protocol given in Figure 1, they both perform local quantum channels $\mathcal{E}_{A^n\to\hat{A}}$ and $\mathcal{F}_{B^n\to\hat{B}}$ to produce the following state:

$$\omega_{\hat{A}\hat{B}} \equiv \left(\mathcal{E}_{A^n\to\hat{A}} \otimes \mathcal{F}_{B^n\to\hat{B}}\right)\left(\varphi_{A^nB^n}\right). \tag{26}$$

If the protocol is successful, then the actual state $\omega_{\hat{A}\hat{B}}$ is ε-close to the ideal maximally entangled state $\Phi_{\hat{A}\hat{B}}$:

$$\frac{1}{2}\left\|\omega_{\hat{A}\hat{B}} - \Phi_{\hat{A}\hat{B}}\right\|_1 \leq \varepsilon. \tag{27}$$

Consider the following chain of inequalities:

$$\begin{aligned}
2n\left(E-\delta\right) &= 2H(\hat{A})_\Phi && (28)\\
&= H(\hat{A})_\Phi + H(\hat{B})_\Phi - H(\hat{A}\hat{B})_\Phi && (29)\\
&= I(\hat{A};\hat{B})_\Phi && (30)\\
&\leq I(\hat{A};\hat{B})_\omega + f(n,\varepsilon) && (31)\\
&\leq I(A^n;B^n)_{\varphi^{\otimes n}} + f\left(n,\varepsilon\right) && (32)\\
&= H(A^n)_{\varphi^{\otimes n}} + H(B^n)_{\varphi^{\otimes n}} - H(A^nB^n)_{\varphi^{\otimes n}} + f(n,\varepsilon) && (33)\\
&= 2nH(A)_\varphi + f(n,\varepsilon). && (34)
\end{aligned}$$

The first equality follows because the entropy of entanglement $H(\hat{A})_\Phi$ of a maximally entangled state $\Phi_{\hat{A}\hat{B}}$ is equal to the logarithm of its Schmidt rank. The next equality follows because $H(\hat{B})_\Phi = H(\hat{A})_\Phi$ and $H(\hat{A}\hat{B})_\Phi = 0$ for a pure bipartite entangled state. The third equality follows from the definition of quantum mutual information. The first inequality follows from applying the AFW inequality for quantum mutual information to (27) with $f\left(n,\varepsilon\right) \equiv 3\varepsilon nE + 2(1+\varepsilon)h_2(\varepsilon/[1+\varepsilon])$. This function has the property that $\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{1}{n}f(n,\varepsilon) = 0$. The second inequality follows from quantum data processing of both A^n and B^n. The final equalities follow from the same arguments as the first two equalities and because the entropy of a tensor product state is additive. Putting everything together, we find that the entanglement concentration rate E for any $(n, E-\delta, \varepsilon)$ entanglement concentration protocol satisfies

$$E - \delta \leq H(A)_\varphi + \frac{1}{2n}f(n,\varepsilon). \tag{35}$$

Taking the limit as $n \to \infty$ and $\varepsilon, \delta \to 0$ allows us to conclude that an achievable rate E of entanglement concentration necessarily satisfies $E \leq H(A)_\varphi$.