Prepared for submission to JHEP

A brief overview of quantum information theory and related topics

Xiao-Liang Qi

Abstract: The note is updated Wednesday 27$^{\text{th}}$ April, 2022.

Contents

1 Classical information theory

1.1 Shannon entropy

The amount of information in a message depends on its prior probability. If I expect "the sun rises from the east" with probability 1, then hearing this statement gives me no additional information. On the contrary, if I heard some rare event (say an earthquake is happening in 10 seconds) it contains a huge amount of information.

How to quantify this intuition? If a message has prior probability p, we can assume that the amount of information is a function of p, denoted as $I(p)$. If we consider two independent events (say the weather tomorrow and the stock market tomorrow), which has probability distribution p_i and q_a for possible outputs i, a, then the joint probability is $P_{ia} = p_i q_a$. We expect the amount of information in the message ia is just the sum of the two, so that the function $f(p)$ is required to satisfy

$$I(p_i q_a) = I(p_i) + I(q_a) \tag{1.1}$$

This equation suggests that I is a linear function of $\log p$. Without losing generality, we can define

$$I(p) = -\log p \tag{1.2}$$

The minus sign is chosen to be consistent with the intuition that rarer events contain higher information. Probability 1 events contains zero information. Since $p \leq 1$, $I(p)$ is non-negative.

For a probability distribution p_i, the average information amount is the Shannon entropy:

$$S(\{p_i\}) = \sum_i p_i I(p_i) = -\sum_i p_i \log p_i \tag{1.3}$$

Since the amount of information is related to the uncertainty (before we heard the message), the entropy is also a measure of uncertainty (how much we don't already know before getting the message). Historically, entropy as a measure of uncertainty was proposed by Boltzman in statistical physics.

1.2 Typical states

Consider a coin that has half probability up and half probability down when thrown. If we throw it 100 times we expect that roughly half of the time it is up. As we throw more and more, the percentage of times it is up should approach half more and more accurately. This is how we measure the probability by doing the experiment many times. To make this more precise, consider a coin with two states with probability p for up and $1-p$ for down. If we throw the coin N times, the probability that it has k up is

$$P_k = \binom{N}{k} p^k (1-p)^{N-k} \tag{1.4}$$

This probability distribution peaks sharply at $\frac{k}{N} = p$. To see this we assume $k \gg 1, N \gg 1$ and use the Stirling formula

$$-\log P_k = -k\log p - (N-k)\log(1-p) - N\log N + k\log k + (N-k)\log(N-k) \quad (1.5)$$

$$= N\left(q\log\frac{q}{p} + (1-q)\log\frac{1-q}{1-p}\right) \quad (1.6)$$

with $q = \frac{k}{N}$. This quantity is actually N times the relative entropy between two probability distributions $\{q, 1-q\}$ and $\{p, 1-p\}$, but we will postpone that discussion to later. The key thing to know is that P_k strongly peaks around $q = p$. If we expand $q = p + x$, for small x we get

$$-\log P_k \simeq \frac{N}{2p(1-p)}x^2 \quad (1.7)$$

Thus the width of the peak is $\propto N^{-1/2}$. If we take a range $|q-p| \leq \frac{c}{\sqrt{N}}$, or $|k - pN| \leq c\sqrt{N}$, the probability of k falling in this range can be arbitrarily close to 1 by choosing a bigger c. In other words, with a very high probability, if we throw a coin N times, we don't see all possible bit strings in the results, but see one of the typical strings satisfying $|k - pN| \leq c\sqrt{N}$. The number of typical string is roughly speaking

$$\binom{N}{k} \simeq e^{NS(\{p,1-p\})} \quad (1.8)$$

The discussion above was made for a coin with two states, but it can be generalized straightforwardly to multistate case. In general, entropy of a probability distribution tells us that the number of typical states in N copies is $e^{NS(\{p_i\})}$.

Since not all bit strings appear, we could store this data more efficiently if we know this probability distribution in advance. For example, if a book contains words w_i with probability p_i, $i = 1, 2, ..., M$, and the book has N words, then a typical state contains $p_i N$ word w_i, and we can store the book with $NS(\{p_i\})/\log M$ code words (if the number of code word is also M).

1.3 Conditional entropy and mutual information

We will introduce some concepts that are related to entropy. For a joint probability distribution $P(x_i, y_j)$, we can define the conditional probability

$$P(y_j|x_i) = \frac{p(x_i, y_j)}{\sum_{y_j} p(x_i, y_j)} \equiv \frac{p(x_i, y_j)}{p(x_i)} \quad (1.9)$$

which is the probability distribution of y conditioned on a given value of x. The entropy of this probability distribution is

$$S(Y|x_i) = S\left[P(y_j|x_i)\right] = -\sum_j P(y_j|x_i)\log P(y_j|x_i)$$

$$= -\frac{1}{p(x_i)}\sum_j p(x_i, y_j)\log p(x_i, y_j) + \log p(x_i) \quad (1.10)$$

If we average this quantity over x_i, we obtain

$$\begin{aligned} S(Y|X) \equiv \sum_i p(x_i) S(Y|x_i) &= -\sum_j p(x_i, y_j) \log p(x_i, y_j) + \sum_i p(x_i) \log p(x_i) \\ &= S(XY) - S(X) \end{aligned} \tag{1.11}$$

We can interpret this as information that are unknown in XY subtracting unknown information in X, which is equal to the remaining uncertainty if we already know X but does not know Y.

If X and Y are independent from each other, $S(XY) = S(X) + S(Y)$, and $S(Y|X) = S(Y)$, which means the uncertainty in Y is the same whether or not we know X. If we take the difference

$$I(X:Y) = S(Y) - S(Y|X) = S(Y) + S(X) - S(XY) \tag{1.12}$$

this measures how much knowing X can reduce the uncertainty in Y. Alternatively we can also say, it measures how much the amount of information in Y is reduced if we already know X. In other words, it measures how much we learn about Y by knowing X. Interestingly, $I(X:Y) = I(Y:X)$ is symmetric. This quantity is called the mutual information.

Mutual information is related to the classical channel capacity. A classical channel takes an input x_i and generates an output y_j with the probability distribution $p(y_j|x_i)$. If Alice encode a message with probability distribution $p(x_i)$, the joint probability distribution is $p(x_i, y_j) = p(x_i)p(y_j|x_i)$. The channel capacity C is defined by the amount of information that can be sent through the channel, per each use of the channel. More precisely, if we take an input message m_I and encode it into a sequence $x_{i_1}x_{i_2}...x_{i_n}$ and fed into n copies of the channel. The output is a sequence $y_{i_1}y_{i_2}...y_{i_n}$. Then a decoding algorithm is needed to map it back to message m_I. If $I = 1, 2, ..., K$ and m_I can be transmitted faithfully, we say the channel has a capacity $C \geq \frac{\log K}{n}$. Take $n \to \infty$ and take an upperbound over encoding and decoding, we define $C = \lim_{n\to\infty} \sup \frac{\log K}{n}$.

To see how to compute C, we again take a two-state example, when $x_i = 0, 1$, $y_i = 0, 1$. If $x_{i_1}x_{i_2}...x_{i_n}$ contains k 0's, then y_i corresponding to these x are generated by the probability distribution $p(y_j|0)$. The 1's correspond to distribution $p(y_j|1)$. Consequently, the number of typical states for the bitstring $y_{i_1}y_{i_2}...y_{i_n}$ is

$$e^{kS(p(y_j|0))} \times e^{(N-k)S(p(y_j|1))} = e^{NH(Y|X)} \tag{1.13}$$

A message about $x_{i_1}x_{i_2}...x_{i_n}$ can be transmitted if typical states in the y bitstring are not totally random. Each given x string with k 0's corresponds to the ensemble of $e^{NS(Y|X)}$ states. The total number of y states with the same y counting is $e^{NS(Y)}$. Thus the number of different x string that can be possibly recovered from y is

$$e^{N(S(Y)-S(Y|X))} = e^{NI(X:Y)} \tag{1.14}$$

Note that $I(X:Y) \leq S(X)$, so that if we have an initial distribution $p(x_i)$, it corresponds to $e^{NS(X)}$ typical states and in general not all of them can be recovered from Y.

This discussion then suggest that the channel capacity is the supreme of mutual information over $p(x_i)$, which measures the maximal number of bits that can be transferred using this noisy channel.

$$C = \sup_{p(x_i)} I(X:Y) = \sup_{p(x_i)} \left(S(Y) - S(Y|X)\right) \tag{1.15}$$

C measures the maximal amount of information that can be transmitted *per use of the channel* with arbitrarily low error. One interesting feature of the classical channel capacity is that it is subadditive. If we take two identical channels and input a joint probability distribution $p(x_1, x_2)$, we can define the output distribution is

$$p(x_1, x_2, y_1, y_2) = \sum_{x_1, x_2} M(y_1|x_1) M(y_2|x_2) p(x_1, x_2) \tag{1.16}$$

The conditional entropy

$$S(Y_1, Y_2|X_1, X_2) = \sum_{x_1, x_2} p(x_1, x_2) S\left(M(y_1|x_1) M(y_2|x_2)\right) = S(Y_1|X_1) + S(Y_2|X_2) \tag{1.17}$$

Here $S(Y_1|X_1) = \sum_{x_1,x_2} p(x_1, x_2) S\left(M(y_1|x_1)\right)$ is the conditional entropy for the reduced input distribution $p(x_1) = \sum_{x_2} p(x_1, x_2)$. On the other hand, the output distribution satisfies

$$S(Y_1, Y_2) \leq S(Y_1) + S(Y_2) \tag{1.18}$$

Thus

$$I(X_1, X_2 : Y_1, Y_2) \leq I(X_1 : Y_1) + I(X_2 : Y_2) \tag{1.19}$$

This discussion can be generalized to more copies. Therefore if we take N copies of the channel and try to send message through a correlated code word choice, it won't be better than using an uncorrelated choice.

$$C = \frac{1}{N} \sup_{p(x_1, x_2, \ldots, x_N)} I(X_1, X_2, ..., X_N|Y_1, Y_2, ..., Y_N) = \sup_{p(x_1)} I(X:Y) \tag{1.20}$$

As we will see later, the quantum channel case is very different.

1.4 Relative entropy

As we discussed in Eq. (1.6), if we have a probability distribution $\{p, 1-p\}$, the chance that k out of N measurements return 0 is given by $P_k = e^{-NS(\{q,1-q\}|\{p,1-p\})}$ with

$$S(\{q, 1-q\} \,|\, \{p, 1-p\}) = q \log \frac{q}{p} + (1-q) \log \frac{1-q}{1-p} \tag{1.21}$$

More generally, for two probability distribution $Q = \{q_i\}, P = \{p_i\}$ we have

$$S(Q|P) = \sum_i q_i \log \frac{q_i}{p_i} \tag{1.22}$$

$S(Q|P)$ measures how different Q is from P. A typical bit string from N copies of Q has the probability $P_N = e^{-NS(Q|P)}$ if the probability distribution were P. In other words, if we made the hypothesis that the probability distribution is P, and actually the distribution is Q, $S(Q|P)$ measures how fast we can find out that we are wrong. The relative entropy is also called Kullback–Leibler (KL) divergence.

To see that the relative entropy is non-negative, we can check that

$$\partial_{q_i} S(Q|P) = 1 + \log \frac{q_i}{p_i} \tag{1.23}$$

Note that q_i satisfies the constraint $\sum_i q_i = 1$, so that the minimum of $S(Q|P)$ is given by the requirement

$$\begin{aligned} \partial_{q_i} \left(S(Q|P) - \mu \left(\sum_i q_i - 1 \right) \right) = 0 \\ 1 + \log \frac{q_i}{p_i} = \mu \\ \Rightarrow \frac{q_i}{p_i} = e^{\mu - 1} \end{aligned} \tag{1.24}$$

Since q_i and p_i are both normalized, this requires $p_i = q_i,\ \mu = 1$. In addition,

$$\partial_{q_i} \partial_{q_j} S(Q|P) = \delta_{ij} q_i^{-1} \tag{1.25}$$

Thus $S(Q|P)$ is convex and has $q_i = p_i$ as the only minimum (with value 0).

Mutual information is actually a relative entropy. If we consider a joint probability distribution $P(x_i, y_j)$, with the reduced distribution $P(x_i) = \sum_{y_j} P(x_i, y_j)$ and $P(y_j) = \sum_{x_i} P(x_i, y_j)$, the relative entropy

$$S(P(x_i, y_j)|P(x_i)P(y_j)) = I(X:Y) \tag{1.26}$$

Physically, mutual information measures how different a probability distribution is from product distribution.

The relative entropy has a lot of important properties. One of the most important one is its monotonicity under a classical channel. If we have Q and P, and we have a classical channel defined by the conditional entropy $M(y_j|x_i) = M_{ji}$. Then the output distribution

$$\tilde{q}_j = \sum_i M_{ji} q_i,\ \tilde{p}_j = \sum_i M_{ji} p_i \tag{1.27}$$

has a relative entropy that is smaller or the same.

$$S(Mq|Mp) \leq S(q|p) \tag{1.28}$$

Intuitively, this means that applying M can only make it more difficult to distinguish the two distributions. (For example if $M_{ji} = M(y_j|x_i)$ is independent from the input x_i, it will erase all differences between P and Q.) To prove this inequality, we first consider a special case of it. Consider p as a joint probability distribution $p_{i\alpha} = p(x_i, y_\alpha)$ and the same for

$q_{i\alpha} = q(x_i, y_\alpha)$. We want to compare the relative entropy $S(q|p)$ with that of the reduced distribution $p^x(i) = \sum_\alpha p_{i\alpha}$ and similar for q^x.

$$\begin{aligned} S(q|p) - S(q_x|p_x) &= \sum_{i\alpha} q_{i\alpha} \log \frac{q_{i\alpha}}{p_{i\alpha}} - \sum_{i\alpha} q_{i\alpha} \log \frac{\sum_\beta q_{i\beta}}{\sum_\gamma p_{i\gamma}} \\ &= \sum_{i\alpha} q_{i\alpha} \log \frac{q_{i\alpha}/q_i^x}{p_{i\alpha}/p_i^x} = \sum_i q_i^x \sum_\alpha q(\alpha|i) \log \frac{q(\alpha|i)}{p(\alpha|i)} \\ &\equiv \sum_i q_i^x S\left(q(y|x_i)|p(y|x_i)\right) \geq 0 \end{aligned} \tag{1.29}$$

where $q(\alpha|i) = q_{i\alpha}/q_i^x$ is the conditional entropy of y_α condition on $x = x_i$. This is an average of the relative entropy between condition probability distribution $S(q(y|x = x_i)|p(y|x = x_i))$, which is thus non-negative. Eq. (1.29) tells us that neglecting part of the variable (y) will only make it more difficult to distinguish the two distributions, which is reasonable. To see how this imply the more general monotonicity under classical channel, we can introduce a uniform random variable $r \in [0, 1]$ with uniform probability. Defining the following function

$$f(i, r) = j, \text{ if } r \in \left[\sum_{k=1}^{j-1} M_{ji}, \sum_{k=1}^{j} M_{ji}\right) \tag{1.30}$$

This makes sure that the probability $P(y_j|x_i) = \int_0^1 dr \delta_{f(i,r),j} = M_{ji}$. Since the mapping from (i, r) to (j, i, r) is one-to-one, no information is lost. If we define $\tilde{q}_{i,j,r} = q_i P(r) = q_i$ (since $P(r) = 1$) for $j = f(i, r)$, and zero otherwise, then

$$S(\tilde{q}|\tilde{p}) = S(q|p) \tag{1.31}$$

since this is just a (redundant) relabeling of the same data. Now if we forget i and r, the reduced distribution of $\tilde{q}$ and $\tilde{p}$ are

$$\sum_i \int_0^1 dr \tilde{q}\,(j, i, r) = \sum_i \int_0^1 dr q_i \delta_{f(i,r),j} = \sum_i M_{ji} q_i \tag{1.32}$$

Thus according to the result we have proven, we have

$$S\left(\sum_i M_{ji} q_i \middle| \sum_i M_{ji} p_i\right) \leq S\left(\tilde{q}|\tilde{p}\right) = S(q|p) \tag{1.33}$$

The monotonicity of relative entropy plays an important role in information theory, which has many useful consequences.

Monotonicity of mutual information. Since mutual information is a relative entropy, it is also monotonous under classical channel. In particular, if we take $p(x_i, y_j)$ and apply separate channel M_{ji} and N_{kl} to it to obtain $\tilde{p}(x_j, y_l) = \sum_{i,k} M_{ji} N_{lk} p(x_i, y_k)$, the mutual information can only decrease. $I(X : Y)[MNp] \leq I(X : Y)[p]$. As a special case of this, for any three variables $I(X : YZ) \geq I(X : Y)$, since forgetting Z is a special case of a channel.

Strong subadditivity. If we write $I(X:YZ) \geq I(X:Y)$ in term of Shannon entropy, we obtain

$$\begin{aligned} S(X) + S(YZ) - S(XYZ) &\geq S(X) + S(Y) - S(XY) \\ \Rightarrow S(YZ) + S(XY) &\geq S(Y) + S(XYZ) \end{aligned} \tag{1.34}$$

This is called strong subadditivity, which is a key property of entropy.

Joint convexity of relative entropy. The relative entropy is jointly convex:

$$S\left(\lambda q_1 + (1-\lambda)q_2 | \lambda p_1 + (1-\lambda)p_2\right) \leq \lambda S\left(q_1|p_1\right) + (1-\lambda) S\left(q_2|p_2\right) \tag{1.35}$$

for $\lambda \in [0,1]$. To prove this we can define a joint probability of $x_i, s = 1, 2$:

$$Q(x_i, 1) = \lambda q_1(x_i),\ Q(x_i, 2) = (1-\lambda) q_2(x_i) \tag{1.36}$$

and the same for $P(x_i, s)$. The relative entropy is

$$S(Q|P) = \sum_i q_1(x_i) \lambda \log \frac{q_1(x_i)}{p_1(x_i)} + \sum_i q_2(x_i)(1-\lambda) \log \frac{q_2(x_i)}{p_2(x_i)} = \lambda S\left(q_1|p_1\right) + (1-\lambda) S\left(q_2|p_2\right) \tag{1.37}$$

If we forget s variable, the induced probability distribution is

$$q(x_i) = \sum_{s=1,2} Q(x_i, s) = \lambda q_1(x_i) + (1-\lambda) q_2(x_i) \tag{1.38}$$

and similar for $p(x_i)$. Thus according to the monotonicity we obtain Eq. (1.35).

Entropy growth. For a probability $q(x_i)$, we can consider it's relative entropy with the uniform distribution $p(x_i) = \frac{1}{M}$ when there are M states.

$$S\left(q|p\right) = -S\left(\{q(x_i)\}\right) - \sum_i q(x_i) \log \frac{1}{M} = \log M - S\left(\{q(x_i)\}\right) \tag{1.39}$$

Thus the entropy "deficit" comparing with the maximal entropy is a relative entropy. As a consequence, if we consider any classical channel N_{ji} which satisfies

$$\sum_i N_{ji} = 1 \tag{1.40}$$

(which is not always true), this channel preserves the maximal entropy state. Such channel applying to a generic q_i will only increase its entropy, since the relative entropy $S(Nq|Np) = S(Nq|p) = \log M - S\left(Nq\right) \leq \log M - S\left(q\right)$.

2 Quantum information theory

2.1 Relation between classical and quantum physics

There is an intrinsic relation between quantum mechanics and classical probability. The foundation of quantum mechanics is quantum state, unitary evolution and projective measurement. A quantum state $|\Psi\rangle$ is a vector in the Hilbert space, a complex linear space, modular a $U(1)$ phase. In other words, $|\Psi\rangle$ and $e^{i\theta}|\Psi\rangle$ is considered as the same state. Alternatively, we can also denote states by the projection operator $\rho = |\Psi\rangle\langle\Psi|$.

We assume the dynamics of the state is described by Schroedinger equation

$$i\frac{\partial}{\partial t}|\Psi(t)\rangle = H(t)|\Psi(t)\rangle \tag{2.1}$$

The key thing is that H is Hermitian, so that the time evolution is unitary. $|\Psi(t)\rangle = U(t)|\Psi(0)\rangle$ with $U^\dagger(t)U(t) = \mathbb{I}$ the identity matrix.

Projective measurements are defined by projectors P_n satisfying

$$\Pi_n\Pi_m = \delta_{nm}\Pi_n,\ \Pi_n = \Pi_n^\dagger,\ \sum_n \Pi_n = \mathbb{I} \tag{2.2}$$

In general each projector may have a rank higher than 1. If we have an orthonormal basis $|n\rangle$, then $P_n = |n\rangle\langle n|$ is a special set of projectors, which have rank 1.

A set of projectors define a quantum measurement. The wavefunction $|\Psi\rangle$ is related to measurement result by

$$P_n = \langle\Psi|\Pi_n|\Psi\rangle \tag{2.3}$$

which satisfies $P_n \geq 0,\ \sum_n P_n = 1$.

In probability theory, when we have multiple variables x_i, y_j we in general need to think about joint probability $P(x_i, y_j)$. In the same way, in quantum mechanics if we have multiple degrees of freedom such as different spins, or different (distinguishable) particles, the Hilbert space has a direct product structure $\mathbb{H} = \mathbb{H}_1 \otimes \mathbb{H}_2$. If we are measuring the first degrees of freedom with projectors $\Pi_n : \mathbb{H}_1 \to \mathbb{H}_2$ in the first Hilbert space, in the bigger Hilbert space we can express the operator as $\Pi_n \otimes \mathbb{I}_2$. For such measurements, the probability

$$P_n = \langle\Psi|\Pi_n \otimes \mathbb{I}_2|\Psi\rangle \tag{2.4}$$

Independent from Π_n, we see that P_n does not depend on the full $|\Psi\rangle$ but only depends on its reduction to the first Hilbert space $\langle\Psi|...\otimes\mathbb{I}_2|\Psi\rangle$. This defines the density operator

$$\rho_1 = \mathrm{tr}_2|\Psi\rangle\langle\Psi| \tag{2.5}$$

which always satisfy $\mathrm{tr}\rho_1 = 1$ and $\langle\Phi|\rho_1|\Phi\rangle \geq 0$ for all $|\Phi\rangle \in \mathbb{H}_1$. Thus ρ_1 is positive semi-definite. For the conceptual simplicity one could also forget about the pure state $|\Psi\rangle$ and start from general mixed state ρ. Unitary evolution $\rho \to U\rho U^\dagger$ preserves the normalization and positive semi-definiteness of ρ.

$$P_n = \mathrm{tr}\,(\Pi_n\rho) \tag{2.6}$$

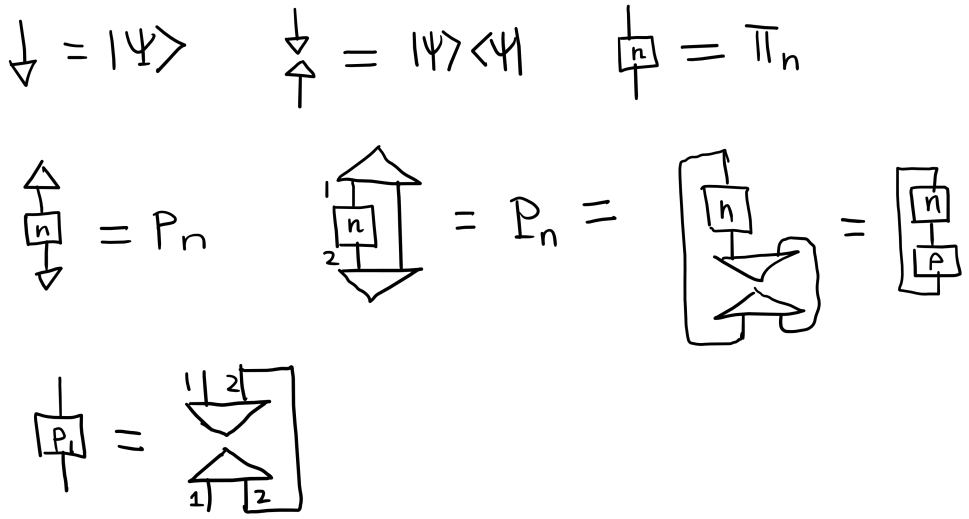

Figure 1. Diagrammatic representation of basic quantum information concepts. An open line pointing up represents a vector in $\mathbb{H}$. A line pointing down represents a vector index in $\mathbb{H}^*$. All internal lines are summed over, just like in Feynman diagrams. A triangle with a single line represents a ket or a bra. A box with two lines represents an operator. In the second and the third row we illustrated the definition of reduced density operator.

is the general way how quantum mechanics is related to classical probability theory. It is helpful to introduce diagrammatic representation of quantum information concepts, as is shown in Fig. 1.

An alternative way of connecting quantum mechanics with classical probability is the simple statement that *a diagonal density operator ρ corresponds to a classical probabilistic mixture of pure states.* A diagonal density operator can be written as $\rho = \sum_n p_n |n\rangle \langle n|$ with $|n\rangle$ an orthogonal matrix. ρ corresponds to a classical probability p_n. However, it should be noted that this statement is meaningless if we only talk about a single state, since every state can be diagonalized. It becomes a meaningful statement if we consider a family of states which are all diagonal in the same basis. In that case if we are only interested in such states, we can say the physics reduces to classical physics.

It is interesting to note that these two point of views are related. For a state ρ with Hilbert space dimension d, and a set of projection operators $\Pi_n,\ n = 1, 2, .., M\ (M \leq d)$, we can define an ancilla with Hilbert space dimension M^k. Denote an orthonormal basis of the ancilla as $|a_1 a_2 ... a_k\rangle$ with $a_i = 1, 2, ..., M$. Denote a cyclic permutation operator R by

$$R\,|a_1 a_2 ... a_k\rangle = |a_1 + 1, a_2 + 1, ..., a_k + 1\rangle \tag{2.7}$$

where the addition is in the cyclic group, such that $M + 1 \sim 1$. R is a unitary operator. Now we can define an unitary operator acting on the system and ancilla

$$U = \sum_{n=1}^{M} \Pi_n \otimes R^n \tag{2.8}$$

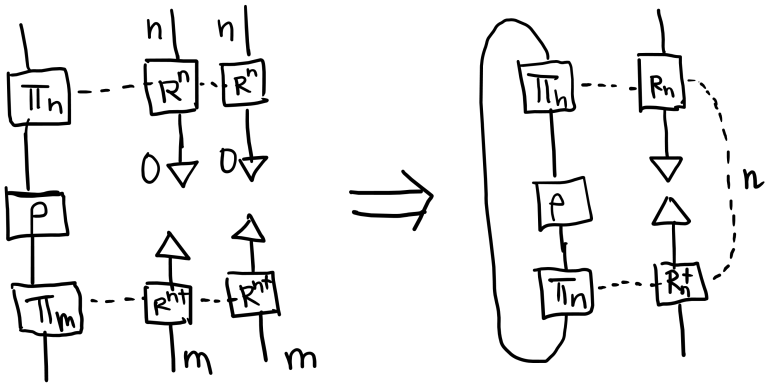

Figure 2. Diagrammatic representation of how a unitary followed by a partial trace can lead to a diagonal density operator of the ancilla.

This is unitary because

$$U^\dagger U = \sum_{n,m} \Pi_m \Pi_n \otimes R^{\dagger m} R^n = \sum_n \Pi_n \otimes \mathbb{I}_A = \mathbb{I}_S \otimes \mathbb{I}_A \tag{2.9}$$

If we prepare the ancilla in the state $|00...0\rangle$ ($|0\rangle$ is the same as $|M\rangle$), then we obtain

$$\sigma_{SA} = U\rho \otimes |00...0\rangle \langle 00...0| U^\dagger = \sum_{n,m} \Pi_n \rho \Pi_m \otimes |nn...n\rangle \langle mm...m| \tag{2.10}$$

The reduced density operator of any one of the ancilla is

$$\rho_{A1} = \mathrm{tr}_{S,A_2A_3...A_k} (\sigma_{SA}) = \sum_n \mathrm{tr} (\Pi_n \rho) |n\rangle \langle n| \tag{2.11}$$

This unitary U followed by partial trace is the physical way to realize a quantum measurement (See. Fig. 2). As long as we "write down" the measurement result in more than one ancilla, if we only look at one of the ancilla system it will always have a diagonal density operator in this basis. (This is often referred to as "decoherence".) We can use this ancilla to measure different physical states ρ_1, ρ_2.... Since arbitrary states always lead to ρ_{A1} that are diagonal in the same basis, we can claim that the measurement maps quantum mechanics to classical probability theory.

2.2 Quantum channel

If we consider two systems A, B and a product initial state $\rho_A \otimes \rho_B$, a unitary evolution generically entangle these two systems and lead to

$$\sigma_{AB} = U\rho_A \otimes \rho_B U^\dagger \tag{2.12}$$

If we only focus on the evolution of A subsystem, we can carry a partial trace over B and obtain

$$\sigma_A = \mathrm{tr}_B \left(U\rho_A \otimes \rho_B U^\dagger \right) \tag{2.13}$$

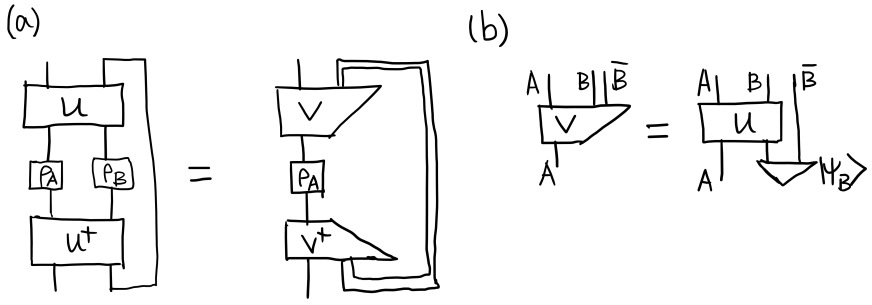

Figure 3. (a) Illustration of a general quantum channel induced from a unitary operator. A unitary acting on $\rho_A \otimes \rho_B$ followed by tracing over B is equivalent to an isometry V from A to $AB\overline{B}$. (b) Explicit definition of V, where $|\Psi_B\rangle$ is a purification of ρ_B.

This is a linear map from ρ_A to σ_A, which maps a density operator to another density operator. Such a map is called completely positive trace-preserving (CPTP) map. Most general CPTP map can always be viewed as a unitary in a bigger system followed by a partial trace. We could also always set ρ_B to be a pure state. If it's not pure, we can introduce a purification of ρ_B by introducing a system $\bar{B}$ with the same Hilbert space dimension as B. Denote $\rho_B = \sum_n p_n |n_B\rangle \langle n_B|$, we define a pure state

$$\left|\psi_{B\overline{B}}\right\rangle = \sum_n \sqrt{p_n} |n_B\rangle \left|n_{\overline{B}}\right\rangle \tag{2.14}$$

The reduced density operator of this state on B is ρ_B. We can take $\left|\psi_{B\overline{B}}\right\rangle \left\langle\psi_{B\overline{B}}\right|$ and apply $U \otimes \mathbb{I}_{\overline{B}}$. Tracing over $B\overline{B}$ gives the same σ_A. The action of U

$$U \otimes \mathbb{I}_{\overline{B}} |\psi_A\rangle \otimes \left|\psi_{B\overline{B}}\right\rangle = \left|\Phi_{AB\overline{B}}\right\rangle \tag{2.15}$$

can be viewed as a linear map from $\mathbb{H}_A$ to $\mathbb{H}_{AB\overline{B}}$, which preserves norm of each state. Such a linear map is called an isometry (Fig. 3). In general, each CPTP map is equivalent to an isometry followed by a partial trace. Mathematically, this is known as Stinespring's dilation theorem.

$$\mathcal{C}(\rho_A) = \mathrm{tr}_E \left(V \rho_A V^\dagger\right) \tag{2.16}$$

$$\text{with } V : \mathbb{H}_A \to \mathbb{H}_A \otimes \mathbb{H}_E \tag{2.17}$$

A CPTP map is also called a quantum channel, which is the quantum analog of a classical channel.

A CPTP map can be defined between two different Hilbert spaces. For example, we can view Eq. (2.13) as a linear map from ρ_B to σ_A. The measurement procedure we described earlier can be viewed as a quantum channel from $\mathbb{H}_S$ to $\mathbb{H}_{A_1}$. This is an example of a quantum-to-classical channel:

$$\mathcal{C}(\rho) = \sum_n |n\rangle \langle n| \,\mathrm{tr}_n (\Pi_n \rho) \tag{2.18}$$

A quantum state can be measured, and if the measurement result is n (with probability $p_n \mathrm{tr}_n (\Pi_n \rho)$), a state $\sigma_n = |n\rangle \langle n|$ is prepared.

2.3 Kraus operator and Lindbladian evolution

The relation of quantum channel with isometry in Eq. (2.16) allows us to take a complete basis $|\alpha\rangle$ in E, and define

$$K_\alpha = \langle\alpha| V \tag{2.19}$$

V maps A to AE and α projects the E part into a fixed state, so that K_α is a linear operator in $\mathbb{H}_A$. Eq. (2.16) can be rewritten as

$$\mathcal{C}(\rho_A) = \sum_\alpha K_\alpha \rho_A K_\alpha^\dagger \tag{2.20}$$

This decomposition can be taken for any quantum channel. The operators K_α are called Kraus operators. They satisfy

$$\sum_\alpha K_\alpha^\dagger K_\alpha = V^\dagger V = \mathbb{I}_A \tag{2.21}$$

which is required for preserving the norm of arbitrary input state ρ_A.

For a unitary evolution U, we can consider an infinitesimal evolution $U = 1 - iH\delta t$ with $\delta t \to 0$ and H a Hermitian matrix. Similarly, we can ask what is the infinitesimal form of quantum channel. Since an identity channel corresponds to $K_0 = \mathbb{I}$ as the only nonzero Kraus operator, it is natural to consider a nearly identity channel as corresponding to a K_0 close to identity and other $K_\alpha,\ \alpha \neq 0$ that are close to zero. If we want $\mathcal{C}(\rho_A) = \rho_A + O(\delta t)$, then $K_\alpha,\ \alpha \neq 0$ must be proportional to $\sqrt{\delta t}$. In contrast, $K_0 - \mathbb{I}$ is of order δt since it can have a nontrivial commutator with ρ_A:

$$K_0 = \mathbb{I} + (-iH + R)\delta t \tag{2.22}$$

$$K_\alpha = \sqrt{\delta t} L_\alpha,\ \alpha \neq 0 \tag{2.23}$$

$$\mathcal{C}(\rho_A) = \rho_A + \delta t \left(-i\left[H, \rho_A\right] + \{R, \rho_A\} + \sum_\alpha L_\alpha \rho_A L_\alpha^\dagger \right) + O(\delta t^2) \tag{2.24}$$

Requiring the channel to preserve trace leads to the condition

$$\begin{aligned} \mathrm{tr}\left(\rho_A R\right) &= \frac{1}{2} \sum_{\alpha \neq 0} \mathrm{tr}\left(L_\alpha \rho_A L_\alpha^\dagger\right) \\ \Rightarrow R &= -\frac{1}{2} \sum_{\alpha \neq 0} L_\alpha^\dagger L_\alpha \end{aligned} \tag{2.25}$$

Thus we have

$$\mathcal{C}(\rho_A) - \rho_A = \delta t \left(-i\left[H, \rho_A\right] + \sum_{\alpha \neq 0} \left[\left[L_\alpha, \rho_A\right], L_\alpha^\dagger\right] \right) \tag{2.26}$$

or

$$\frac{d\rho_A}{dt} = -i\left[H, \rho_A\right] + \sum_{\alpha \neq 0} \left[\left[L_\alpha, \rho_A\right], L_\alpha^\dagger\right] \tag{2.27}$$

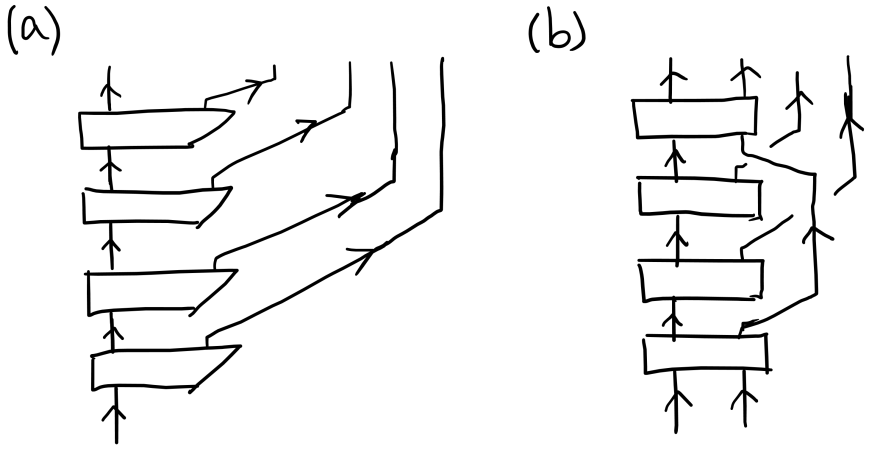

Figure 4. Illustration of a Lindbladian evolution (a) and a more general evolution that couples the system with bath (b).

This is the Lindbladian equation, which can be used to describe the evolution of certain open systems. As is illustrated in Fig. 4, Lindbladian equation physically assumes that the system couples with the environment, and after every step, the environment qubit leaves the system and never returns, so that we can trace over it. This is called a Markovian approximation. For example, if the system is an atom and the environment is a photon system, and photons never return the system after being emitted, the evolution of the system can be described well by a Lindbladian evolution. In more general cases, a qubit can enter the environment and later return to the system, leading to a time evolution that is beyond Lindbladian.

As a simple example of Lindbladian evolution, consider $L_\alpha = \sqrt{J}\,|\alpha\rangle\langle\alpha|$ with $|\alpha\rangle$ an orthonormal basis. $\sqrt{J}$ is an energy scale. For simplicity take the Hamiltonian $H = 0$. Then the Lindbladian equation is

$$\frac{d\rho}{dt} = J\sum_\alpha \left(|\alpha\rangle\langle\alpha|\langle\alpha|\rho|\alpha\rangle - \rho\right) \tag{2.28}$$

This equation describes an exponential decay of off-diagonal components of ρ:

$$\rho_{\alpha\beta}(t) = \rho_{\alpha\beta}(0)e^{-Jt} \text{ for } \alpha \neq \beta \tag{2.29}$$

$$\rho_{\alpha\alpha}(t) = \rho_{\alpha\alpha}(0) \tag{2.30}$$

This is a simplest model of decoherence, which describes how a quantum state (with all quantum phase information in the off-diagonals of density operator) evolves into a diagonal classical state by coupling to the bath.

2.4 More general measurements

In the same way how we formulate the projective measurement as a coupling with ancilla, we can consider more general unitary coupling with an ancilla. Without losing generality we can consider the ancilla with a pure state initial state (since otherwise we can always

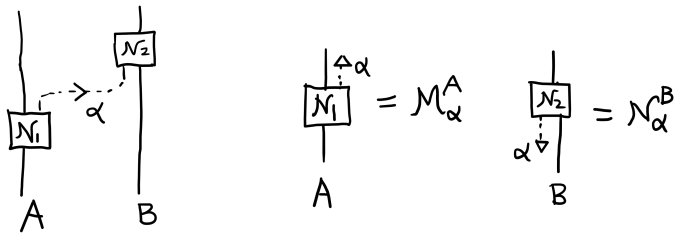

Figure 5. Illustration of LOCC from A to B.

purify it), in which case we can define the coupling as an isometry V from the system $\mathbb{H}_S$ to the system and ancilla $\mathbb{H}_S \otimes \mathbb{H}_A$. Then we can carry a projective measurement on the ancilla (which physically requires to introduce another ancilla and do the special kinds of coupling shown in Fig. 2)). The outcome of this measurement is

$$p_n = \mathrm{tr}_{AS}\left[\Pi_{nA}\left(V\rho_S V^\dagger\right)\right] \tag{2.31}$$

In general, this probability is not the result of any direct projective measurement on ρ_S. We can express it in a form

$$p_n = \mathrm{tr}_S\left(\rho_S M_n\right) \tag{2.32}$$

$$M_n = V^\dagger \Pi_{nA} V \tag{2.33}$$

M_n are not projectors, but they are positive, and they satisfy

$$\sum_n M_n = \mathbb{I}_S \tag{2.34}$$

These are called positive operator valued measurement (POVM). Sometimes they are also called weak measurements. As a related concept, the mapping

$$\rho_S \to \Pi_{nA} V \rho_S V^\dagger \Pi_{nA} \equiv \mathcal{M}_n\left(\rho_A\right) \tag{2.35}$$

is a linear map that maps ρ_S to a positive operator but with a generically smaller trace. This is called a completely positive (CP) map. In this example, $\sum_n \mathcal{M}_n$ is a CPTP map.

2.5 Local operation and classical communication

A concept related to measurements is local operation and classical communication (LOCC). This is a set of operations that do not create quantum entanglement. In general, LOCC is defined between multiple parties, but here we will only discuss two-party case with two subsystems A and B. A local operation refers to quantum channels of the form $N_A \otimes \mathcal{N}_B$. A classical communication requires to first measure one subsystem, such as A, and send the classical result to B. After B receive this message, a local operation, i.e. quantum channel can be applied to B. This operation is called a one-way LOCC $LOCC_{A\to B}$. We can decompose this process into two parts. Sending the information from A is a channel

$$\mathcal{N}_1 :\ \mathbb{H}_A \to \mathbb{H}_A \otimes \mathbb{H}_C \tag{2.36}$$

$$\mathcal{N}_1(\rho_A) = \sum_\alpha \mathcal{M}_\alpha\left(\rho_A\right) \otimes |\alpha\rangle\langle\alpha| \tag{2.37}$$

Here $\mathcal{M}_\alpha$ is a CP map, and $\sum_\alpha \mathcal{M}_\alpha$ is a CPTP map. This channel carries the POVM to A and record the result in C.[1]

The next step is to take the classical information in C and bring it to B. The influence of this message to B corresponds to a quantum channel

$$\mathcal{N}_2 : \mathbb{H}_B \otimes \mathbb{H}_C \to \mathbb{H}_B \tag{2.38}$$

$$\mathcal{N}_2 \left(|\alpha\rangle \langle\alpha| \otimes \rho_B \right) \equiv \mathcal{N}_\alpha \left(\rho_B \right) \tag{2.39}$$

Here $\mathcal{N}_\alpha$ is a channel from B to B defined by restricting the initial state of C to the basis state. Combining the two steps we get the LOCC channel

$$\mathcal{N} = \mathcal{N}_2 \circ \mathcal{N}_1 = \sum_\alpha \mathcal{M}_\alpha \otimes \mathcal{N}_\alpha \tag{2.40}$$

with $\mathcal{M}_\alpha$ acting on A and $\mathcal{N}_\alpha$ acting on B. Any channel with this decomposition is a one-way LOCC.

If there can be two-way classical communication, we can define the class LOCC_r which denotes the channel that combines r rounds of back-and-forth classical communication. LOCC plays an important role in distinguishing classical correlation and quantum entanglement. LOCC cannot create quantum entanglement, since it can be realized by purely classical communication. For a review on LOCC, see [1].

2.6 von Neumann entropy and mutual information

Since each density operator ρ is diagonal in some basis, it is natural to define the Shannon entropy of this diagonal distribution as the entropy of ρ:

$$\rho = \sum_i p_i |i\rangle \langle i| \tag{2.41}$$

$$S(\rho) = -\sum_i p_i \log p_i = -\text{tr}\left(\rho \log \rho\right) \tag{2.42}$$

The second expression makes it explicit that the von Neumann entropy is invariant under unitary: $S(\rho) = S\left(U\rho U^\dagger\right)$. In analogy of the classical discussion of typical states, if we take N copies of the same state ρ,

$$\rho^{\otimes N} = \sum_{i_1 i_2 \ldots i_N} p_{i_1} p_{i_2} \ldots p_{i_N} |i_1 i_2 \ldots i_N\rangle \langle i_1 i_2 \ldots i_N| \tag{2.43}$$

We can define the operator N_i as the number of $i_k = i$:

$$N_i = \sum_{i_1 i_2 \ldots i_N} \sum_{a=1}^{N} \delta_{i_a, i} |i_1 i_2 \ldots i_N\rangle \langle i_1 i_2 \ldots i_N| \tag{2.44}$$

Then we have

$$\text{tr}\left(N_i \rho^{\otimes N}\right) = \sum_{n_i=0}^{N} \binom{N}{n_i} p_i^{n_i} \left(1 - p_i\right)^{N - n_i} n_i \tag{2.45}$$

[1] More precisely, the POVM operator M_α is $M_\alpha = \mathcal{M}_\alpha^\dagger \left(\mathbb{I}_A\right)$.

In the same way as the classical case, one can show that $\rho^{\otimes N}$ is almost an eigenstate of $\frac{N_i}{N}$ with eigenvalue $\frac{N_i}{N} = p_i$. The number of typical states is the subspace defined by

$$\frac{N_i}{N} = p_i + O\left(N^{-1/2}\right), \ \forall i \tag{2.46}$$

which has the dimension $e^{NS(\rho)+O(\log N)}$.

The quantum analog of joint probability distribution is a state ρ living in the direct product Hilbert space $\mathbb{H}_A \otimes \mathbb{H}_B$. The quantum mutual information can be defined as

$$I(A:B) = S\left(\rho_A \otimes \rho_B\right) - S\left(\rho_{AB}\right) = S_A + S_B - S_{AB} \tag{2.47}$$

If there is an orthogonal basis $|n_A\rangle$ of A and $|n_B\rangle$ of B, such that

$$\rho_{AB} = \sum_{n_A, m_B} p_{nm} \left|n_A\right\rangle \left\langle n_A\right| \otimes \left|m_B\right\rangle \left\langle m_B\right| \tag{2.48}$$

then the quantum mutual information is the same as the classical one. However, in general this is not possible. This expression requires that the basis in which ρ_{AB} is diagonal is a product basis between A and B. In general, the eigenstates of ρ_{AB} are not product states.

One major difference with the classical case is that $I(A:B)$ can be bigger than S_A or S_B. For example if AB is in a pure state

$$\left|I_{AB}\right\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} \left|i_A\right\rangle \left|i_B\right\rangle \tag{2.49}$$

for A, B having the same Hilbert space dimension d. For this state we have

$$S_{AB} = 0, \ S_A = S_B = \log d \Rightarrow I(A:B) = 2\log d \tag{2.50}$$

If we write $I(A:B)$ in term of conditional entropy, we obtain

$$I(A:B) = S_A - S(A|B), \ S(A|B) \equiv S_{AB} - S_B \tag{2.51}$$

In quantum case, $S(A|B)$ can be negative, while classically the conditional entropy is always non-negative.

2.7 Quantum relative entropy

The classical relative entropy also has a quantum counter part. If we consider two states ρ, σ that are commuting, we can diagonalize them as $\rho = \sum_i q_i \left|i\right\rangle \left\langle i\right|$, $\sigma = \sum_i p_i \left|i\right\rangle \left\langle i\right|$, and write the classical relative entropy

$$S\left(q|p\right) = \sum_i q_i \log \frac{q_i}{p_i} = \mathrm{tr}\left(\rho \log \rho - \rho \log \sigma\right) \tag{2.52}$$

We can generalize this definition to the quantum case when ρ, σ do not commute. In general

$$S\left(\rho|\sigma\right) = \mathrm{tr}\left(\rho \log \rho - \rho \log \sigma\right) \tag{2.53}$$

Just like the classical case, the relative entropy is non-negative, and monotonous under quantum channels. There are direct proofs of the monotonicity[2] which was based on the strong subadditivity of entropy[3]. Alternatively, we discuss the physical interpretation of relative entropy from the point of view of hypothesis testing. We will see that this interpretation implies monotonicity of the relative entropy. This discussion largely follows[4].

We first discuss the non-negativity of relative entropy. Denote $|n\rangle$ as a basis in which σ is diagonal, such that $\sigma = \sum_n p_n |n\rangle \langle n|$. Then

$$S(\rho|\sigma) = \mathrm{tr}\,(\rho \log \rho) - \sum_n \langle n|\, \rho\, |n\rangle \log p_n \tag{2.54}$$

In the second term, only diagonal terms of ρ appears. We can define a diagonal state

$$\rho_D \equiv \sum_n |n\rangle \langle n|\, \tilde{q}_n,\ \ \tilde{q}_n \equiv \langle n|\, \rho\, |n\rangle \tag{2.55}$$

then

$$S\,(\rho|\sigma) = S\,(\rho_D) - S\,(\rho) + S\,(\rho_D|\sigma) \tag{2.56}$$

Since ρ_D and σ commute with each other by definition, $S\,(\rho_D|\sigma)$ is a classical relative entropy which is non-negative. The remaining task is to show that $S\,(\rho_D) - S\,(\rho)$ is also non-negative. To see that, denote ρ in its diagonal basis as $\rho = \sum_m q_m |u_m\rangle \langle u_m|$. Then

$$\tilde{q}_n = \sum_m q_m \left|\langle n|\, u_m\rangle\right|^2 \tag{2.57}$$

$M_{nm} \equiv |\langle n|\, u_m\rangle|^2$ defines a classical channel, which satisfies $\sum_m M_{nm} = 1$. Thus as we discussed earlier, $S_{\max} - S\,(\rho)$ is equal to a classical relative entropy which decreases under this classical channel, which means $S\,(\rho_D) \geq S\,(\rho)$. Thus we have proven $S\,(\rho|\sigma) \geq 0$.

Now we discuss the relation of relative entropy and hypothesis testing. As we discussed in the classical case, if we conjecture that there is a classical state σ (which means the probability distribution given by the eigenvalues of σ), but actually the state is ρ_D (in the same sense), then after N measurements we conclude that the probability the state is actually σ is

$$P_N = e^{-NS(\rho_D|\sigma)} \tag{2.58}$$

Note that for ρ_D and σ, if we measure them in a different basis, in the same way as in Eq. (2.57) we see that the basis change corresponds to mapping the two ensembles corresponding to ρ_D and σ by the same classical channel, which can only decrease the relative entropy and make the two states more difficult to distinguish.

Compared with the classical case, in the quantum case when we have N copies of the system, we can carry a joint measurement by applying an operator in the N-copied Hilbert space. This is a new possibility that does not exist in the classical case. If the states commute with each other, so is $\rho_D^{\otimes N}$ and $\sigma^{\otimes N}$, such that the classical calculation above still apply. However, it should be noticed that $\sigma^{\otimes N}$ has a lot of degeneracies. For example if σ is

2×2, all eigenstates $|n_1 n_2 ... n_N\rangle$ with the same number of 0 and 1 corresponds to the same eigenvalue $p_0^k (1-p_0)^{N-k}$. Therefore we have a lot of freedom in choosing a diagonal basis for $\sigma^{\otimes N}$. Denote a basis choice as Π_I with I labeling the states in the N-copied Hilbert space $\mathbb{H}^{\otimes N}$, then for each choice of Π_I that makes σ diagonal, one can define

$$\rho_D^{(N)} = \sum_I \Pi_I \rho^{\otimes N} \Pi_I \tag{2.59}$$

Note that Π_I does not have to be a direct product of each copy, so that $\rho_D^{(N)}$ is not necessarily the product of N density operators. Naively, $\rho_D^{(N)}$ could have a very different entropy from $\rho^{\otimes N}$, but it turns out that we can reduce this difference by noticing the symmetry of $\rho^{\otimes N}$ and $\sigma^{\otimes N}$. In the N-copied Hilbert space we can consider the symmetry action of permutation group S^N and unitary group $U(d)$ (with d the Hilbert space dimension of each copy). The permutation acts naturally by permuting different replicas, and $U(d)$ refers to the same unitary operator acting on each replica. Thus in this action S^N commutes with $U(d)$ and the group is $S^N \times U(d)$. According to the Schur-Weyl duality, the Hilbert space $\mathbb{H}^{\otimes N}$ factorizes into a direct sum of different representations:

$$\mathbb{H}^{\otimes N} = \oplus_Y \mathbb{H}_Y^S \otimes \mathbb{H}_Y^U \tag{2.60}$$

Here the sum is over Young tableaux, which label the representation of S^N and $U(d)$. Since $\rho^{\otimes N}$ and $\sigma^{\otimes N}$ commute with all permutations, if we expand them in this decomposition of Hilbert space, they looks like

$$\sigma^{\otimes N} = \sum_Y \mathbb{I}_Y^S \otimes p_Y \sigma_Y \tag{2.61}$$

with $\sigma_Y \in \mathbb{H}_Y^U$. We have defined σ_Y to be normalized, so that there is a coefficient $p_Y \in [0,1]$ satisfying $\sum_Y p_Y = 1$. Similarly

$$\rho^{\otimes N} = \sum_Y \mathbb{I}_Y^S \otimes q_Y \rho_Y \tag{2.62}$$

This decomposition tells us a strong restriction on how non-commuting these two operators could be. Since both of them are block-diagonal already in this form, when we choose a diagonal basis of $\sigma^{\otimes N}$ we only need to apply projection to $\rho^{\otimes N}$ within each subspace. One can prove

$$S\left(\rho^{\otimes N} \middle| \sigma^{\otimes N}\right) = \sum_Y q_Y S\left(\rho_Y | \sigma_Y\right) + S\left(\{q_Y\} | \{p_Y\}\right) \tag{2.63}$$

The second term is the classical relative entropy between q_Y and p_Y probability distributions. Now we can choose a diagonal basis of σ_Y and take the diagonal elements of ρ_Y, denoted as ρ_{DY}. This will leave the second term invariant and decrease the first term.

$$S\left(\rho^{\otimes N} \middle| \sigma^{\otimes N}\right) = \sum_Y q_Y \left(S\left(\rho_{DY}\right) - S\left(\rho_Y\right)\right) + S\left(\rho_D^{(N)} \middle| \sigma^{\otimes N}\right) \tag{2.64}$$

$$S\left(\rho_D^{(N)} \middle| \sigma^{\otimes N}\right) = \sum_Y q_Y S\left(\rho_{DY} | \sigma_Y\right) + S\left(\{q_Y\} | \{p_Y\}\right) \tag{2.65}$$

Now the key point is that the entropy difference $S(\rho_{DY}) - S(\rho_Y)$ is bounded by the maximal entropy of this subspace:

$$\begin{aligned} & S(\rho_{DY}) - S(\rho_Y) \leq \log D_Y \\ & \Rightarrow S\left(\rho^{\otimes N} \middle| \sigma^{\otimes N}\right) \leq \max_Y \log D_Y + S\left(\rho_D^{(N)} \middle| \sigma^{\otimes N}\right) \end{aligned} \tag{2.66}$$

The maximal dimension D_Y scales polynomially with N:

$$D_Y \leq (N+1)^{d(d-1)/2} \tag{2.67}$$

For example for $d = 2$, the maximal dimensional representation has $SU(2)$ spin $N/2$, with dimension $N + 1$. Consequently we obtain

$$S\left(\rho_D^{(N)} \middle| \sigma^{\otimes N}\right) \leq S\left(\rho^{\otimes N} \middle| \sigma^{\otimes N}\right) \leq \frac{d(d-1)}{2} \log(N+1) + S\left(\rho_D^{(N)} \middle| \sigma^{\otimes N}\right) \tag{2.68}$$

$$\Rightarrow S(\rho|\sigma) = \lim_{N \to +\infty} \frac{1}{N} S\left(\rho_D^{(N)} \middle| \sigma^{\otimes N}\right) \tag{2.69}$$

The right-hand side controls the optimal distinguishability of the two states when we carry measurements on N copies. Thus we conclude that for large N, the probability of measured state ρ be mistaken as σ is at least

$$P_N = e^{-NS(\rho|\sigma)} \tag{2.70}$$

This result gives an operational definition of the relative entropy, and also implies its monotonicity under quantum channels. According to the dilation theorem, a quantum channel $\mathcal{N}$ can be viewed as an isometry followed by partial trace. Denote the isometry as $V : \mathbb{H}_A \to \mathbb{H}_A \otimes \mathbb{H}_E$, one can prove

$$S\left(V\rho V^\dagger \middle| V\sigma V^\dagger\right) = S(\rho|\sigma) \tag{2.71}$$

Now if we carry measurements on $\left(V\rho V^\dagger\right)^{\otimes N}$ and $\left(V\sigma V^\dagger\right)^{\otimes N}$, but restrict the measurements to A, then the probability P_N will only go up (since it is more difficult to distinguish the two states). Restricting the operators to (N-copies of) A is equivalent of computing the relative entropy between the reduced states $\mathrm{tr}_E\left(V\rho V^\dagger\right) = \mathcal{N}(\rho)$ and $\mathcal{N}(\sigma)$. Thus we obtain

$$S(\mathcal{N}(\rho)|\mathcal{N}(\sigma)) \leq S(\rho|\sigma) \tag{2.72}$$

for any quantum channel $\mathcal{N}$.

In the same way as the classical case, the monotonicity of quantum relative entropy implies the monotonicity of mutual information:

$$I(A:B) = S(\rho_{AB}|\rho_A \otimes \rho_B) \tag{2.73}$$

$$I(A:B)(\mathcal{N}_A \otimes \mathcal{N}_B(\rho_{AB})) \leq I(A:B)(\rho_{AB}) \tag{2.74}$$

In particular, it implies the strong subadditivity of entropy:

$$\begin{aligned} & I(A:BC) \geq I(A:B) \\ \Rightarrow\ & S_{BC} + S_{AB} \geq S_{ABC} + S_B \end{aligned} \tag{2.75}$$

If we consider N copies of ρ_{AB} and carry joint measurements by applying POVM $M_{A\alpha} \in \mathbb{H}_A^{\otimes N}$ and $M_{B\beta} \in \mathbb{H}_B^{\otimes N}$[2], the measurement leads to the classical joint probability distribution $P_{\alpha\beta}$, which has a classical mutual information $I_c^{(N)}(A:B)$. The monotonicity implies $I_c^{(N)}(A:B) \leq NI(A:B)$. Note that even if we optimize over all POVM, $\frac{1}{N}I_c^{(N)}(A:B)$ may not reach $I(A:B)$. The classical mutual information always satisfy $\frac{1}{N}I_c^{(N)}(A:B) \leq S_A$ (and also S_B) but $I(A:B) > S_A$ is possible. For example for an EPR pair state $I(A:B) = 2\log 2$ while the classical MI is upper bounded by $\log 2$. On the other hand, the discussion above suggests that the relative entropy $I(A:B) = \frac{1}{2}S\left(\rho_{AB}^{\otimes N}\middle|\rho_A^{\otimes N} \otimes \rho_B^{\otimes N}\right)$ is close to classical relative entropy. This suggests that the optimal measurement (the one in the diagonal basis of $\rho_A^{\otimes N} \otimes \rho_B^{\otimes N}$ as we discussed above) cannot factorize into measurements in $\mathbb{H}_A^{\otimes N}$ and $\mathbb{H}_B^{\otimes B}$. In other words, the optimal measurements must be in entangled basis if $I(A:B) > S_A$ or $I(A:B) > S_B$. This is a consequence of the intrinsic nonlocality of quantum mechanics.

We can also obtain the joint convexity of relative entropy in the same way as the classical case:

$$S\left(p\rho_1 + (1-p)\rho_2 | p\sigma_1 + (1-p)\sigma_2\right) \leq pS\left(\rho_1|\sigma_1\right) + (1-p)S\left(\rho_2|\sigma_2\right) \tag{2.76}$$

In the quantum language we can introduce an ancilla and construct the state $\pi = p\rho_1 \otimes |1\rangle\langle 1| + (1-p)\rho_2 \otimes |2\rangle\langle 2|$ and $\eta = p\sigma_1 \otimes |1\rangle\langle 1| + (1-p)\sigma_2 \otimes |2\rangle\langle 2|$. The convexity follows from the monotonicity upon tracing over the ancilla.

2.8 Holevo information

When we have a quantum channel, one natural question is how much classical information can be transmitted through this quantum channel. For classical information in the probability distribution $p(x_i) = p_i$, we can define states $|i\rangle\langle i|$ and encode this classical information in a diagonal density operator $\rho = \sum_i p_i |i\rangle\langle i|$. A quantum channel brings this state to

$$\sigma = \mathcal{C}(\rho) = \sum_i p_i \mathcal{C}\left(|i\rangle\langle i|\right) \equiv \sum_i p_i \sigma_i \tag{2.77}$$

How much information is transferred depends on σ_i. For example, σ_i could be all the same, in which case there is no information transferred. To measure the amount of classical information transferred, one need to apply a POVM and obtain

$$P(\alpha|i) = \mathrm{tr}\left(M_\alpha \sigma_i\right) \tag{2.78}$$

[2]Here I mean the operator acts in this Hilbert space. More precisely we should say $O_A \in \mathbb{H}_A^{\otimes N} \otimes \mathbb{H}_A^{*\otimes N}$.

Just like in the classical channel capacity discussion, the information transferred through the channel can be measured by the classical mutual information

$$I_{\mathcal{C}} = \sup_{M_\alpha} \sum_{p_i} I[P(\alpha|i)p_i] \tag{2.79}$$

We can define an auxiliary state

$$\pi = \sum_i p_i \left|i\right\rangle \left\langle i\right| \otimes \sigma_i \tag{2.80}$$

where the $\left|i\right\rangle \left\langle i\right|$ lives in an additional copy of the input state. This is obtained by copying the input classical information before sending it to $\mathcal{C}$. The probability distribution $P(\alpha|i)p_i$ corresponds to the diagonal state

$$\begin{aligned} \eta &= \sum_i p_i \left|i\right\rangle \left\langle i\right| \otimes \mathcal{M}\left(\sigma_i\right) \\ &= \sum_i p_i \mathrm{tr}\left(M_\alpha \sigma_i\right) \left|i\right\rangle \left\langle i\right| \otimes \left|\alpha\right\rangle \left\langle \alpha\right| \end{aligned} \tag{2.81}$$

Thus the mutual information in η is upper bounded by that in π:

$$I\left[P(\alpha|i)p_i\right] = I_\eta \leq I_\pi \tag{2.82}$$

One can show that

$$I_\pi = S\left(\sum_i p_i \sigma_i\right) - \sum_i p_i S\left(\sigma_i\right) \equiv \chi \tag{2.83}$$

This is known as the Holevo χ quantity, which provides an upper bound of the classical information that can be transmitted.

As a special case, if $\sigma_i = \left|\psi_i\right\rangle \left\langle \psi_i\right|$ is a pure state, we have $\chi = S\left(\sum_i p_i \left|\psi_i\right\rangle \left\langle \psi_i\right|\right)$. From a different point view, this means that for any mixed state that is an ensemble of pure states $\sigma = \sum_i p_i \left|\psi_i\right\rangle \left\langle \psi_i\right|$, there is a way to encode classical information of the amount $S(\sigma)$ if we can pick the pure state in the ensemble.

Similar to the discussion about classical channel capacity, it is natural to ask what happens when we use a quantum channel N times in parallel. Denote the input Hilbert space of each channel as $\mathbb{H}_{\text{in}}$, In general we can choose a set of basis states $\left|I\right\rangle$ in $\mathbb{H}_{\text{in}}^{\otimes N}$ and a probability distribugtion p_I. Then the Holevo information is

$$\chi^{(N)} = S\left(\sum_I p_I \mathcal{C}^{\otimes N}\left(\left|I\right\rangle \left\langle I\right|\right)\right) - \sum_I p_I S\left(\mathcal{C}^{\otimes N}\left(\left|I\right\rangle \left\langle I\right|\right)\right) \tag{2.84}$$

In general, $\frac{1}{N}\chi^{(N)}$ could increase as a function of N. However, if we assume that $\left|I\right\rangle$ are all product states of the N copies, *i.e.*

$$\left|I\right\rangle = \prod_{a=1}^{N} \left|i_a\right\rangle \tag{2.85}$$

Then we have

$$\chi^{(N)} = S\left(\sum_I p_I \otimes_{a=1}^N \sigma_{i_a}\right) - \sum_I p_I \left(\sum_{a=1}^N S\left(\sigma_{i_a}\right)\right) \tag{2.86}$$

Here $\sigma_{i_a} = \mathcal{C}\left(|i_a\rangle \langle i_a|\right)$. Denote the reduced probability distribution of $p_I = p_{i_1 i_2 \ldots i_N}$ as p_{i_a}, the reduced state of $\sum_I p_I \otimes_{a=1}^N \sigma_a$ on a-th Hilbert space is $\sum_{i_a} p_{i_a} \sigma_{i_a}$. Thus due to strong subadditivity we have

$$S\left(\sum_I p_I \otimes_{a=1}^N \sigma_{i_a}\right) \leq \sum_{a=1}^N S\left(\sum_{i_a} p_{i_a} S\left(\sigma_{i_a}\right)\right) \tag{2.87}$$

This implies that

$$\chi^{(N)} \leq \sum_{a=1}^N \chi_a \tag{2.88}$$

Thus for product state, the Holevo information is subadditive. In other words, for transmitting information, it is not useful to use classically correlated code words (but it is useful to use quantum entangled code word states $|I\rangle$).

2.9 Quantum channel capacity

In analogy with the classical channel capacity, we are interested in the amount of quantum information that can be transmitted through a quantum channel. In the classical case, the channel capacity is C if we can (faithfully) transmit C bits of classical information per each use of the channel. In classical case, the quantum channel capacity is defined as the maximal number of qubits that can be transmitted errorlessly per each use of the channel. More precisely, consider an arbitrary M qubit state $|\psi\rangle \in \mathbb{H}_I$. For a channel $\mathcal{C} : \mathbb{H}_A \to \mathbb{H}_B$, the quantum information in $|\psi\rangle$ can be transmitted accurately with N copies of the channel, if there exists an encoding map $\mathcal{E} : \mathbb{H}_I \to \mathbb{H}_A^{\otimes N}$ and a decoding map $\mathcal{D} : \mathbb{H}_B^{\otimes N} \to \mathbb{H}_I$, such that

$$\mathcal{D} \circ \mathcal{C}^{\otimes N} \circ \mathcal{E}(|\psi\rangle) = |\psi\rangle \tag{2.89}$$

Since we require this to be true for arbitrary $|\psi\rangle$, this is equivalent to the requirement

$$\mathcal{D} \circ \mathcal{C}^{\otimes N} \circ \mathcal{E} = \mathbb{I} \tag{2.90}$$

The quantum channel capacity is the upper limit of $\frac{M}{N}$:

$$Q\left(\mathcal{C}\right) = \lim_{N\to\infty} \sup_{\mathcal{D},\mathcal{E}} \frac{M}{N} \log 2 \tag{2.91}$$

This definition is illustrated in Fig. 6 (a). (There shouldn't be $\log 2$ if we are measuring the channal capacity in term of qubits, but we add it because it is more convenient for relating this quantity to von Neumann entropy, since we have been using natural log in the definition of the latter.)

We would want to have a more explicit formula, ideally N independent, for quantum channel capacity. This is an open question. In the following we will discuss some more explicit formula, although it is still N-dependent. First, if we apply the channel $\mathcal{D}\circ\mathcal{C}^{\otimes N}\circ\mathcal{E}$ to a maximally entangled state between I and an auxiliary system R, the identity condition translates to the statement that after applying $\mathcal{D}\circ\mathcal{C}^{\otimes N}\circ\mathcal{E}$ the state of IR is still maximally entangled. (It's ok if it's a different maximally entangled state, since that just means the quantum channel is applying a unitary to I, which can be easily absorbed by redefinition of $\mathcal{D}$.) The encoding map is an isometry which preserves information. The key question is whether $\mathcal{C}^{\otimes N}$ also preserves the quantum information.

To obtain a more explicit formula for quantum channel capacity, we consider the setup in Fig. 6 (c). The dilation of channel $\mathcal{C}$, denoted as $V_{\mathcal{C}}$, is applied to a state $|\psi_{RA}\rangle$. Really we should be discussing N copies of A, B and $\mathcal{C}$ but we leave out the N for simplicity. We will discuss the N copy case later. The state $\mathcal{E}|IR\rangle$ in (b) is a special case of $|\psi_{RA}\rangle$. The requirement of information preservation is that one can find a decoding channel $\mathcal{D}$ applying to B which maps the state of RBE to $\rho_{R\tilde{A}E} = |\psi_{RA}\rangle\langle\psi_{RA}|\otimes\rho_E$. Now remember that the monotonicity of mutual information requires

$$I(R:A) \geq I(R:B) \geq I(R:\tilde{A}) \tag{2.92}$$

so if we want $I(R:A) = I(R:\tilde{A})$, the necessary condition is $I(R:A) = I(R:B)$. Conversely, if $I(R:A) = I(R:B)$, it implies that

$$S(A) = S(B) - S(RB) = S(B) - S(E) \tag{2.93}$$

where we have used the fact that RA together is a purestate. Thus

$$S(R) = S(A) = S(RE) - S(E) \Rightarrow I(R:E) = 0 \tag{2.94}$$

This equation says that if the mutual information did not decrease when A is mapped to B, then the environment E does not know anything about R. The quantity on the right-hand side can be written as

$$S(A) = S(B) - S(RB) = -S(R|B) \tag{2.95}$$

with $S(R|B)$ the conditional entropy. In classical case the conditional entropy is always positive, but in quantum case it can be negative. It actually must be negative in order for the equation above to hold. The quantity $-S(R|B)$ is important below, which is named as coherent information, denoted as

$$I_c(A:B)_{\rho_A} = S(\mathcal{C}(\rho_A)) - S\left(\mathcal{C}\left(|\psi_{RA}\rangle\langle\psi_{RA}|\right)\right) \tag{2.96}$$

It should be noted that the coherent information only depends on ρ_A, since different $|\psi_{RA}\rangle$ that corresponds to the same ρ_A can always be related by a unitary acting on R, which does not change $S(B)$ or $S(RB)$.

Because $I(R:E) = 0$ we have $\rho_{RE} = \rho_R\otimes\rho_E$. A state like this can be purified by separately purifying ρ_R and ρ_E. We denote the purified state as

$$|\psi_{REB_1B_2}\rangle = |\psi_{RB_1}\rangle\otimes|\psi_{B_2E}\rangle \tag{2.97}$$

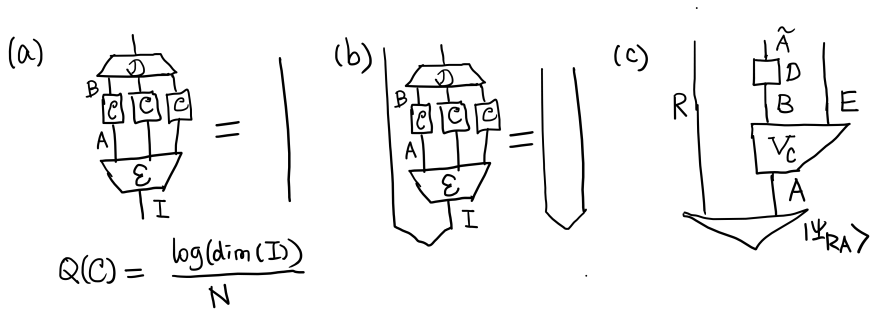

Figure 6. (a) Definition of the quantum channel capacity. (b) For an input state in A that is purified to $|\psi_{RA}\rangle$, we consider the dilation of the channel $\mathcal{C}$ to an isometry from A to BE.

where B_1 and B_2 are two independent systems. Since all different purifications of a state are related by unitaries, we could simply take B_1 to have the same size as A and take $|\psi_{RB_1}\rangle = |\psi_A\rangle$. For the same reason, there exists a unitary U_B acting on B such that

$$|\psi_{RB_1}\rangle \otimes |\psi_{EB_2}\rangle = U_B |\psi_{RB}\rangle \tag{2.98}$$

Thus we have proven that there exists a channel $\mathcal{D}$, obtained by applying U_B and tracing over B_2, which recovers the state $|\psi_{RA}\rangle$.

The discussion above can be easily generalized to the case of N copies. For a general state of RA^N, the condition of recovery is

$$\begin{aligned} &I\left(R:A^N\right) = I\left(R:B^N\right) \\ &\text{or } S(A^N) = I_c\left(A^N:B^N\right)_{\rho_{A^N}} \end{aligned} \tag{2.99}$$

Now if we consider the case in Fig. 6 (b), $S(A) = S(I) = M \log 2$ is the number of qubits. Thus

$$Q\left(\mathcal{C}\right) = \sup_{\mathcal{E}} \frac{S(I)}{N} = \sup_{|\psi_{\mathcal{E}}\rangle} \frac{1}{N} I_c\left(A^N:B^N\right)_{\rho_{A^N}} \tag{2.100}$$

The supreme is taken for all states of the form $\mathcal{E}|IR\rangle$. Thus we have an upper bound if we replace it by a supreme over all possible states:

$$Q\left(\mathcal{C}\right) \leq \lim_{N\to\infty} \sup_{\rho_{A^N}} \frac{1}{N} I_c\left(A^N:B^N\right)_{\rho_{A^N}} \tag{2.101}$$

This upper bound can actually be achieved. Let's assume that the state $|\psi_{RA^N}\rangle$ is one that maximizes I_c in the N copy case. Then we can take another large number M and look at NM copies with the state $|\psi_{RA^N}\rangle^{\otimes M}$. In large M, such a state can be approximated by a different state $\left|\tilde{\psi}_{\tilde{R}A^{NM}}\right\rangle$, in which $\rho_{\tilde{R}}$ is maximally mixed, with a Hilbert space dimension $\dim_{\tilde{R}} \simeq e^{MS_R}$. Thus this state has the form of $|\psi_{\mathcal{E}}\rangle = \mathcal{E}|IR\rangle$. Of course, to be more precise we need to quantify how close are these two states, and prove that the small deviation corresponds to a small change in the coherent information. We won't go into these details here.

As an example to illustrate this formula, we note that measure-and-prepare channels have zero quantum channel capacity. A measure-and-prepare channel is defined as

$$\mathcal{C}(\rho) = \sum_{\alpha} \mathrm{tr}\left(M_\alpha \rho\right) \sigma_\alpha \tag{2.102}$$

which carries a POVM M_α to ρ and prepare a state σ_α. For this channel, with initial state $|\psi_{RA}\rangle$ we have

$$\mathcal{C} \otimes \mathbb{I}\left(|\psi_{RA}\rangle \langle\psi_{RA}|\right) = \sum_{\alpha} \mathrm{tr}_A \left(M_\alpha |\psi_{RA}\rangle \langle\psi_{RA}|\right) \otimes \sigma_\alpha \tag{2.103}$$

This is a separable state. Denote $\mathrm{tr}_A \left(M_\alpha |\psi_{RA}\rangle \langle\psi_{RA}|\right) = p_\alpha \pi_\alpha$, this state is denoted as

$$\eta_{RB} \equiv \mathcal{C} \otimes \mathbb{I}\left(|\psi_{RA}\rangle \langle\psi_{RA}|\right) = \sum_{\alpha} p_\alpha \pi_\alpha \otimes \sigma_\alpha \tag{2.104}$$

For this state, the coherent information $S_B - S_{RB}$ is always non-positive. To see this we define an auxiliary state

$$\eta_{R\tilde{B}} = \sum_{\alpha} p_\alpha \pi_\alpha \otimes |\alpha\rangle \langle\alpha| \tag{2.105}$$

where we defined an ancilla system $\tilde{B}$ with orthonormal basis $|\alpha\rangle$. Obviously one can apply a measure-and-prepare channel $\mathbb{M} : \tilde{B} \to B$ to map $\eta_{R\tilde{B}}$ to η_{RB}. Since $I_c = S_B - S_{RB} = I(R:B) - S_R$ is monotonous under quantum channel applying to B, we have

$$I_c\left[\eta_{R\tilde{B}}\right] \geq I_c\left[\eta_{RB}\right] \tag{2.106}$$

For $\eta_{R\tilde{B}}$ we have

$$I_c\left[\eta_{R\tilde{B}}\right] = S_{\tilde{B}} - S_{R\tilde{B}} = -\sum_{\alpha} p_\alpha S\left(\pi_\alpha\right) \leq 0 \tag{2.107}$$

Therefore we have proven $I_c\left[\eta_{RB}\right] \leq 0$.

2.10 Quantum teleportation

By definition, one cannot send quantum information by a purely classical channel. However, if the sender and the receiver already shared quantum entanglement, it is possible to make use of that and send (teleport) a quantum state by sending only a classical signal. The simplest example of teleportation is shown in Fig. 7. AB together is controlled by the sender, and C is the receiver. B and C are in a maximally entangled EPR pair state. We also prepared the maximally entangled state of A with an auxiliary system $\overline{A}$ for the convenience of keeping track of input quantum information in A. Quantum teleportation means that we can apply a measurement on AB and send the result to C. Then applying a unitary transformation to C controlled by the measurement result, the output state of C is equal to the initial state of A.

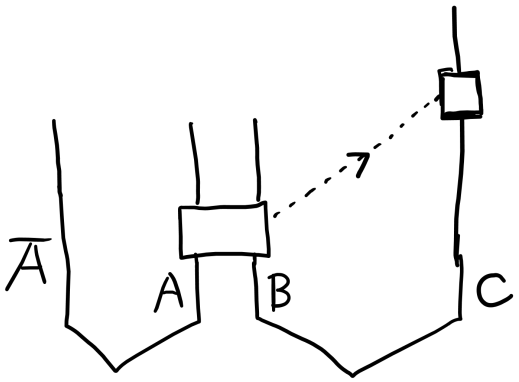

Figure 7. Illustration of quantum teleportation.

To see how this works, we define four matrices σ^μ, with $\sigma^0 =$, $\sigma^a, a = x, y, z$ are Pauli matrices. Then we define the four Bell basis states of AB:

$$|\Psi_\mu\rangle = \sigma_A^\mu |I_{AB}\rangle \tag{2.108}$$

Here $|I_{AB}\rangle$ is a reference state that is maximally entangled. We choose $|I_{AB}\rangle$ to be the same as $|I_{CB}\rangle$, with C replaced by A.

Now we measure AB in the Bell basis and send the result to C. The measurement is defined by the projectors

$$P_\mu = |\psi_\mu\rangle \langle\psi_\mu| = \sigma_A^\mu |I_{AB}\rangle \langle I_{AB}| \sigma_A^\mu \tag{2.109}$$

Since the state of AB is $\rho_{AB} = \frac{1}{4}\mathbb{I}_{AB}$, the probability of result μ is $\frac{1}{4}$ for all μ. For a given measurement output μ, the state of $\overline{A}C$ becomes

$$|\phi_\mu\rangle_{\overline{A}C} = \langle\psi_\mu| I_{CB} I_{\overline{A}A}\rangle = \sigma_C^\mu \langle I_{AB}| I_{CB} I_{\overline{A}A}\rangle \tag{2.110}$$

σ_A^μ can be moved to C because the state of BC and that of AB are maximally entangled. Now after C receive the classical signal μ, she can apply another gate σ_a^μ. After applying this gate, we obtain the final state

$$\left|I_{\overline{A}C}\right\rangle = \left|I_{\overline{A}A}\right\rangle \tag{2.111}$$

which is identical to the original state. If we input a state $|\psi_A\rangle$ in A, after teleportation C will be in the same state $|\psi_C\rangle = |\psi_A\rangle$.

It should be noted that the teleportation process is a one-way LOCC from AB to C, which cannot increase entanglement between AB and C, but the teleportation protocol shows that $LOCC_{AB\to C}$ can increase the entanglement of C with another region A.

If we have an initial state $|\psi_A\rangle$ (Fig. 7 (b)), it should be noted that the probability of each μ is still

$$p_\mu = \mathrm{tr}\left(P_\mu |\psi_A\rangle \langle\psi_A| \otimes \mathbb{I}_B\right) = \frac{1}{4} \tag{2.112}$$

Therefore there is no classical information about the initial state in this distribution. In other words, if someone intercepted the classical communication, she/he still does not learn anything about the initial quantum state. Similarly, if someone measures the quantum state of C before the classical signal reaches C, she/he will only see a maximally mixed state which contains no information about input state A.

The name "teleportation" often gives the impression that we can send quantum state faster than speed of light, but actually because the quantum state can only be recovered after the classical signal arrives, teleportation does not achieve any faster communication than usual signal propagation. What distinguishes teleportation from ordinary signal propagation is its non-locality: the quantum information in the input state A is neither in the classical signal nor in the entangled pair between BC, but nonlocally stored in both of them.

There is a classical analog of teleportation which is a simple encryption procedure. If we have a random number generator which gives B a random number r, and gives C the opposite number $-r$, then if AB wants to send a signal m, she will send $m+r$ to C, and C carries the computation $m+r-r=m$ to extract the information. The main difference from quantum teleportation is that the classical messages can be intercepted without being noticed.

Beyond this simple example, in general, quantum teleportation is defined by three systems A, B, C with A and C having the same size, and a state ω_{BC}. The teleportation is successful if there exists an LOCC channel $\mathcal{T}_{AB\to C}$ such that the reduced channel

$$\mathcal{C}_{AC}\left(\rho_A\right) \equiv \mathrm{tr}_{AB}\left(\mathcal{T}_{AB\to C}\left(\rho_A \otimes \omega_{BC}\right)\right) \tag{2.113}$$

is identity channel. If we purify A by a maximally entangled state $\left|I_{A\overline{A}}\right\rangle$, the identity channel requirement is equivalent to

$$I\left(\overline{A} : C\right) = 2S_A^{\max} \tag{2.114}$$

after applying the LOCC. If we consider a dilation of the LOCC channel by including the environment W (Fig. 8), we have

$$I\left(\overline{A} : WABC\right) = 2S_A^{\max} \tag{2.115}$$

One can prove that for a tripartite pure state

$$I\left(\overline{A} : WABC\right) = I\left(\overline{A} : C\right) + I\left(\overline{A} : ABW\right) \tag{2.116}$$

Thus the sufficient and necessary condition for $I\left(\overline{A} : C\right)$ to be maximal is that $\overline{A}$ has zero mutual information with the complement ABW. This proves that successful teleportation requires the state of AB after the measurement and the classical measurement output μ to know nothing about the input state.

Another question is the amount of entanglement in the initial state ω_{BC} needed for the teleportation. Before applying the channel, $I(C : BA\overline{A}) = I(C : B)\left[\omega_{BC}\right] = 2S_C\left[\omega_{BC}\right]$. After the channel, C is maximally entangled with $\overline{A}$ and thus not entangled with the

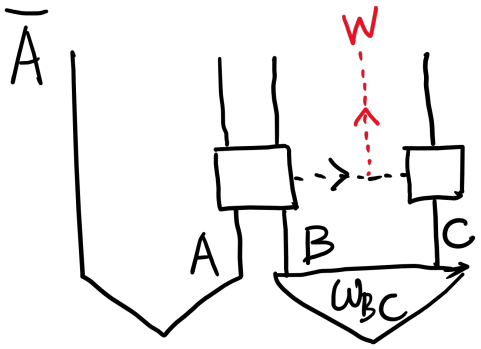

Figure 8. Illustration of a general teleportation process. W is the environment that we introduce in dilation of the LOCC channel.

remainder of the system, such that $I(C : BA\overline{A}) = 2S_A^{\max}$. Thus due to monotonicity we have

$$S_A^{\max} \leq S_C\left[\omega_{BC}\right] \tag{2.117}$$

Thus the initial state must have at least the same amount of entanglement entropy as $S_A^{\max}$. (In general it is possible to consider a C that is bigger than A, and the LOCC maps ABC to $AB\tilde{C}$ with $\tilde{C}$ the same size as A.)

2.11 Quantum error correction

When we send a signal in a noisy channel, it may be damaged. To make sure the signal is sent, we can try to use some code that has some redundancy and is able to correct errors. For example, we want to send one bit 0 or 1 through a classical channel, but this channel has a probability p of flipping the bit. Thus our information may be destroyed with this probability. To reduce this probability, we can duplicate the message by k copies to 00...0 and 11...1. Sending these signals through k copies of the same channel, the probability that m copies of 0 are flipped to 1 is given by

$$p_m = \begin{pmatrix} k \\ m \end{pmatrix} p^m \tag{2.118}$$

This probability peaks at $m = kp$. As long as $p < \frac{1}{2}$, for large enough k the probability that more than half of 0 is flipped $P = \sum_{m \geq k/2} p_m$ is very low (it decays exponentially with k). Thus we can decode the original message $0, 1$ by majority vote, with an exponentially small error.

The idea of quantum error correction is similar. A quantum channel is generally noisy because the coupling with environment. Denote the quantum channel describing such errors as $\mathcal{E}$. Quantum information is generally not preserved by a quantum channel for a generic input. Error correction is possible if we restrict the input. If we define an encoding map $\mathcal{C}$ and a decoding map $\mathcal{D}$, the condition of error correction is

$$\mathcal{D} \circ \mathcal{E} \circ \mathcal{C} = \mathbb{I} \tag{2.119}$$

The image of encoding map $\mathcal{C}$ defines a subspace of the Hilbert space, called the code subspace. The discussion here is very similar to that in the channel capacity. If we consider a dilation of $\mathcal{E}$, the error correction condition requires that the environment introduced in the dilation knows nothing about the input. Denote the Kraus operators of $\mathcal{E}$ as E_i, and the projection operator onto the code subspace as P_C, the complement channel maps the input state $\rho \in \mathbb{H}_C$ to a state of the environment:

$$\sigma_E^{ij} = \mathrm{tr}_S \left(E_i \rho E_j^\dagger \right) \tag{2.120}$$

The requirement that environment knows nothing about the input state requires that σ_E^{ij} is independent from the input. We can write the equation above as $\mathrm{tr}\,(\rho)\, \sigma_E^{ij} = \mathrm{tr}\left(E_j^\dagger E_i \rho \right)$ for any $\rho \in \mathbb{H}_C$. This requires

$$P_C E_j^\dagger E_i P_C = P_C \sigma_E^{ij} \tag{2.121}$$

This is called the *quantum error correction condition*.[5] Each E_i labels a possible kind of error.

The decoding map can be explicitly constructed if the quantum error-correction condition is satisfied. Denote the unitary that diagonalize σ_E as u_{ia}, such that $\sigma_E^{ij} = \sum_a u_{ia} p_a u_{ja}^*$. Denote

$$\tilde{E}_a = \sum_i E_i u_{ia} \tag{2.122}$$

we have

$$P_C \tilde{E}_a^\dagger \tilde{E}_b P_C = P_C p_a \delta_{ab} \tag{2.123}$$

This equation means that

$$M_a = p_a^{-1} \tilde{E}_a P_C \tag{2.124}$$

is an isometry from the code subspace to the output Hilbert space, satisfying $M_a^\dagger M_a = P_C$. In addition $M_a^\dagger M_b = 0$ for $a \neq b$.

The error channel maps an input state $\rho = P_C \rho P_C$ to

$$\sigma_S = \sum_i E_i P_C \rho P_C E_i^\dagger = \sum_a \tilde{E}_a P_C \rho P_C \tilde{E}_a^\dagger = \sum_a M_a \rho M_a^\dagger p_a \tag{2.125}$$

Because $M_a^\dagger M_b = 0$ for $a \neq b$, σ_S is a direct sum over multiple blocks, each of which contains the information about the original state ρ. The original state can be decoded from each of them, as long as $p_a \neq 0$.

The decoding channel can be defined as

$$\mathcal{D}\left(\sigma_S\right) = \sum_a M_a^\dagger \sigma_S M_a \tag{2.126}$$

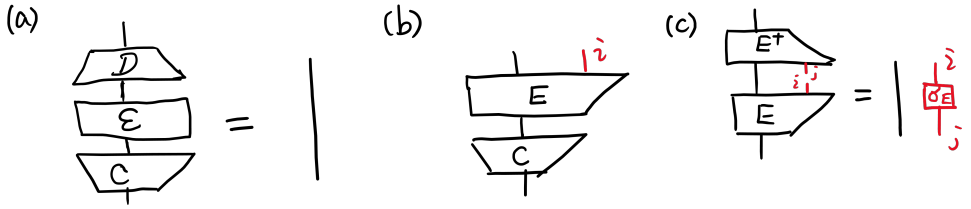

Figure 9. Illustration of the quantum error correction condition. (a) The requirement of quantum error channel $\mathcal{E}$ being corrected for the code subspace states defined by encoding $\mathcal{C}$. (b) The dilation of error channel $\mathcal{E}$ into isometry E, with i labeling the environment state. (c) The QECC condition (Eq. (2.121)).

$\mathcal{D}$ satisfies

$$\mathcal{D}\left(\sum_a M_a \rho M_a^\dagger p_a\right) = \rho \tag{2.127}$$

Note that $M_a^\dagger$ are not all Kraus operators of $\mathcal{D}$. In general $\sum_a M_a M_a^\dagger$ is a projector with some zero eigenvalues. We can introduce an orthogonal projector $P_\perp$ such that $\sum_a M_a M_a^\dagger + P_\perp = \mathbb{I}$. $M_a^\dagger$ and $P_\perp$ forms the Kraus operators of $\mathcal{D}$.

As an example of quantum error correction, we consider the qutrit code (which I learned from Ref. [6]). The encoding map is defined from one qutrits (Hilbert space dimension 3) to three qutrits. The code subspace consists of three states $|\psi_a\rangle$, $a = 0, 1, 2$ which is defined by

$$|\psi_a\rangle = \frac{1}{\sqrt{3}} \sum_{b=0,1,2} |b\rangle_1 |a+b\rangle_2 |a-b\rangle_3 \tag{2.128}$$

Here the states are labeled by $0, 1, 2$ and the addition and subtraction are defined in $\mathbb{Z}_3$. It is easy to see that $|\psi_a\rangle$ are three orthogonal states. The error model we consider is erasure of the third qubit, which means the error channel $\mathcal{E}$ is defined by tracing out the third qubit. To write down the Kraus operators we can define

$$E_a = \mathbb{I}_{12} \otimes \langle a|_3, \ a = 0, 1, 2 \tag{2.129}$$

which maps the 3-qutrit state to a 2-qutrit state. Clearly $\sum_a E_a \rho E_a^\dagger = \mathrm{tr}_3 \rho = \rho_{12}$.

The QECC condition (2.121) requires that for $|\psi_a\rangle$,

$$\langle\psi_a| E_c^\dagger E_d |\psi_b\rangle = \delta_{ab}\sigma_{cd} \tag{2.130}$$

Note that

$$E_d |\psi_b\rangle = \frac{1}{\sqrt{3}} |b-d\rangle_1 \otimes |2b-d\rangle_2 \tag{2.131}$$

we see that $\langle\psi_a| E_c^\dagger E_d |\psi_b\rangle = \frac{1}{3}\delta_{ab}\delta_{cd}$.

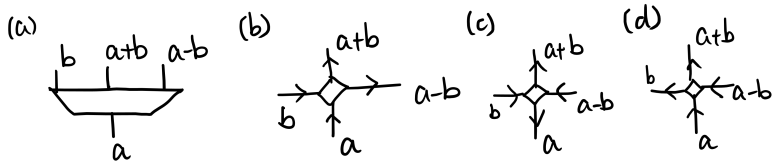

Figure 10. Illustration of the qutrit code. (a) Definition of the state. (b) A more symmetric way of drawing the state $|\Psi\rangle$, which defines a unitary mapping from any two qutrits to the other two.

One can show that the qutrit code is also able to correct the erasure error of qutrits 1 or 2, if we know which site is erased. An equivalent way of staying this QECC property (for erasure of any one of the three qutrits) is to consider the auxiliary state

$$|\Psi\rangle = \frac{1}{\sqrt{3}} |\psi_a\rangle \otimes |a\rangle_4 = \frac{1}{3} \sum_{a,b} |b\rangle_1 |a+b\rangle_2 |a-b\rangle_3 |a\rangle_4 \tag{2.132}$$

This state has the property that any two qutrits among the four are maximally entangled with the other two. In other words, for any choice $i, j = 1, 2, 3, 4$, the subsystem entropy $S_{ij} = \log 9$ is maximal. The wavefunction of this state is often referred to as a "perfect tensor" in the holographic duality literature[6], which we will return to in later part of the note.

3 Entanglement properties of many-body systems

In this section, I will discuss entanglement properties of various quantum many-body systems. In general it is difficult to compute entanglement properties such as the von Neumann entropy, because it involves the density operator which has an exponentially high dimension. Therefore it is useful to study example systems in which the calculation is easier.

3.1 Toric code

We start with a quantum state called "toric code", which is also known as the ground state of $(2+1)$-dimensional Z_2 gauge theory. This state provides a simple example of quantum many-body state for which quantum entanglement entropy can be computed exactly. It is also an example of topologically ordered state and an example of stabilizer states. We will discuss these related concepts after discussing the toric code state.

The toric code model proposed by Alexei Kitaev[7] is a spin model with a single qubit defined for each link of a given lattice. For example if we take a two-dimensional square lattice, there is a qubit for each link. Denote the Pauli matrices for link ij as Z_{ij} and X_{ij}, the toric code Hamiltonian is defined as

$$H = -U \sum_i \prod_{j \text{ nn } i} X_{ij} - V \sum_I \prod_{\langle ij \rangle \in \square_I} Z_{ij} \tag{3.1}$$

with $U, V > 0$. Here i labels the vertices of the lattice and $\square_I$ labels the plaquettes. One can check that these two terms commute with each other for all i, I. Thus the ground state is given by the lowest eigenstate of all terms. If we take eigenstates of X_{ij} with eigenvalues ± 1, we can view $X_{ij} = -1$ as a string and $X_{ij} = +1$ as "no string". The ground state satisfies

$$\prod_{j \text{ nn } i} X_{ij} |G\rangle = |G\rangle \tag{3.2}$$

for all i. This requires that the string does not end, so that we obtain close loops of $X_{ij} = -1$ strings. The second term is a kinetic energy of the string, which creates a new small loop, deforms an existing loop, or annhilates a loop. The requirement

$$\prod_{\langle ij \rangle \in \square_I} Z_{ij} |G\rangle = |G\rangle \tag{3.3}$$

requires the ground state to be an equal weight superposition of all different string configurations that can be connected by the fundamental move. If the spatial geometry is a sphere with trivial first homotopy, the ground state is uniquely determined, which is a sum over all close loop configurations, *i.e.* a "loop gas". If the geometry has nontrivial homotopy, such as a torus, there are degenerate ground states depending on the number of noncontractable loops (0 or 1) around each cycle.

This model can be viewed as a special limit of the Z_2 gauge field. Take Z_{ij} to be the gauge vector potential on each link, the term $\prod_{j \text{ nn } i} X_{ij}$ generates the Z_2 gauge transformation (flipping the sign of Z_{ij} on all neighboring links). Thus the condition (3.2) is the

gauge invariance requirement. The V term in the Hamiltonian is the energy of magnetic flux. In general, the gauge theory Hamiltonian contains another term $H = -\lambda \sum_{\langle ij \rangle} X_{ij}$ which is the Z_2 version of the electric field energy (analog of $\mathbf{E}^2$ term in the Maxwell theory). This term does not commute with the toric code Hamiltonian (3.1). Toric code can be viewed as a weakly coupled (*i.e.* deconfined) limit of the Z_2 gauge theory.

Now we discuss entanglement entropy of a subsystem in the toric code. Let's consider a region with the topology of disk, illustrated in Fig. 11 (a). For simplicity let's assume the entire system is defined on a disk with open boundary condition, in which case there is a unique ground state. Since the ground state is a superposition of all string configurations, the string configuration in a region A and its complement is only correlated by the boundary values. For example, $X_{13} = -1$ in Fig. 11 requires that $X_{12}X_{14}X_{15} = -1$ in the ground state. We can write the ground state in the following decomposition:

$$|G\rangle = N_\partial^{-1/2} \sum_{X_{\partial A} \equiv \{X_{ij}, \langle ij \rangle \in \partial A\}} |G_A \{X_{\partial A}\}\rangle \left|G_{\overline{A}} \{X_{\partial A}\}\right\rangle \tag{3.4}$$

Here N_∂ is the number of boundary configurations $X_{\partial A}$. Obviously states with different boundary configurations are orthogonal to each other. It is slightly more nontrivial to confirm that the number of configurations with different $X_{\partial A}$ are all the same, which allows us to write the expression above with $|G_A \{X_{\partial A}\}\rangle$ and $\left|G_{\overline{A}} \{X_{\partial A}\}\right\rangle$ normalized. This is a Schmit decomposition of the state $|G\rangle$, which tells us that all nonzero eigenvalues of the reduced density matrices are equal to N_∂^{-1}. Thus the entanglement entropy and all Renyi entropies are equal to

$$S_A^{(n)} = S_A = \log N_\partial \tag{3.5}$$

Now the question is what is N_∂, the number of boundary string configurations. Each boundary link ij can have $X_{ij} = \pm 1$, so naively one would thought $N_\partial = 2^{|\partial A|}$, with $|\partial A|$ the number of boundary links, *i.e.* the area of A. However, the strings are closed, which means for a disk region A the number of -1 must be even. This is a consequence of the Gauss law.

$$\prod_{\langle ij \rangle \in \partial A} X_{ij} |G\rangle = \prod_{i \in A} \prod_{j \text{ nn } i} X_{ij} |G\rangle = |G\rangle \tag{3.6}$$

Therefore the number of allowed configuration of $X_{\partial A}$ is reduced by half, leading to

$$\begin{aligned} N_\partial &= 2^{|\partial A|-1}, \\ S_A &= S_A^{(n)} = \log 2 \, |\partial A| - \log 2 \end{aligned} \tag{3.7}$$

The negative constant $-\log 2$ is called topological entropy[8, 9], which is a signature of the non-local correlation between X_{ij} at different locations. This is a signature of topological order, since it is something that one cannot observe locally. The existence of topological entropy, as a constant correction to the area law of entanglement entropy, is a generic signature of $(2+1)$-dimensional topologically ordered states. In order to avoid possible ambiguities in defining the area law (which is not a problem for the special toric code

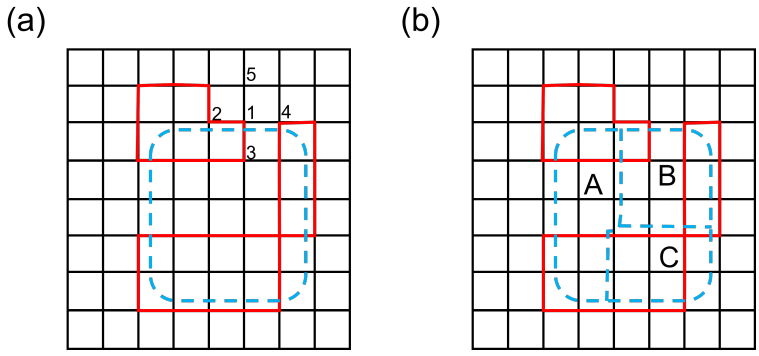

Figure 11. Illustration of the toric code model. The red and black links represent $X_{ij} = -1$ and $X_{ij} = +1$ states, respectively. The blue dashed line defines a subsystem A.

state but may be more suble for a more generic state, where the area law term is only proportional to the area in the limit when A is large), Ref. [9] proposed to consider three regions A, B, C, illustrated in Fig. 11 (b). They considered the tripartite information

$$\begin{aligned} I_3(A:B:C) &= I(A:B) + I(A:C) - I(A:BC) \\ &= S_A + S_B + S_C - S_{AB} - S_{AC} - S_{BC} + S_{ABC} \end{aligned} \tag{3.8}$$

If there is an EPR pair between A and B, it will contribute to both $I(A:B)$ and $I(A:BC)$, so $I_3(A:B:C)$ measures the entanglement between A, B and C which are not pair-wise between any two of them. For the toric code state with entanglement entropy (3.7), the area law terms all cancel, and we are left with only the topological entropy:

$$I_3(A:B:C) = -\log 2 \tag{3.9}$$

As we mentioned above, the toric code model have multiple degenerate ground states when the lattice has nontrivial topology. For example on a torus it has 4 ground states, corresponding to 0 or 1 closed loops in each non-contractable cycle. On an annulus, if there is a fixed boundary condition on the open boundaries (for example, if no string can end on the boundary), there are two ground states. These ground states are locally indistinguishable: if we take a topologically trivial disk region A, the reduced density operator ρ_A is identical for the different ground states. If we define an isometry that maps two qubit states $|ab\rangle$, $a, b = 0, 1$ to the four ground states of toric code on the torus $|\Psi_{ab}\rangle$, this is an encoding map that has quantum error correction properties. Quantum erasure errors on any topologically trivial region can be corrected, since in the auxiliary state

$$|\Phi\rangle = \frac{1}{2} \sum_{a,b} |ab\rangle_W |\Psi_{ab}\rangle \tag{3.10}$$

The mutual information between ancilla W and a disk region A is zero. Such error correction property suggests that quantum information stored in topological ground states is robust against local perturbation. This is a generic property of topologically ordered states, which is the key idea behind topological quantum computation.

3.2 Stabilizer states

The toric code is an example of a large family of quantum states which are determined by constraints like Eq. (3.2) and (3.3). As another example, we can consider a 1-dimensional Ising model

$$H = -J \sum_i Z_i Z_{i+1} \tag{3.11}$$

There are two degenerate ground states $|G_\pm\rangle$ with all Z_i taking the same value $+1$ or -1. To lift this degeneracy we can introduce another operator $F = X_1 X_2 ... X_N$. This term commutes with all $Z_i Z_{i+1}$. The N conditions

$$Z_i Z_{i+1} |G\rangle = |G\rangle \,, \;\; i = 1, 2, ..., N-1, \;\; F|G\rangle = |G\rangle \tag{3.12}$$

fix the state $|G\rangle = \frac{1}{\sqrt{2}} (|++...+\rangle + |--...-\rangle)$. (Alternatively, we could also use $Z_i Z_{i+1}$, $i = 1, 2, 3, ..., N-1$ and Z_1 to fix the state $|G_+\rangle = |++...+\rangle$.) Compare this example with the toric code example, the common things is that there are N commuting operators, each with eigenvalue ± 1, which have $|G\rangle$ as their only common eigenstate with eigenvalue 1. These operators are called *stabilizers*. Stabilizer states have various nice properties. They could be highly entangled and their entanglement properties can be studied relatively easily.

Consider a system with N qubits labeled by $i = 1, 2, ..., N$. A complete basis of operators is given by the product of Pauli operators. Denote X_i and Z_i as the Pauli X and Z operators, with $X_i Z_i = -Z_i X_i$. The Pauli $Y_i = i Z_i X_i$. For each site i there are 4 operators $\mathbb{I}_i, X_i, Y_i, Z_i$ which we could instead denote as

$$T_i^{ab} \equiv (-i)^{ab} X_i^a Z_i^b, \;\; a = 0, 1, \;\; b = 0, 1 \tag{3.13}$$

Thus we can label a generic Pauli string by a $2N$-dimensional binary vector $v = (a_1, b_1, a_2, b_2, ..., a_N, b_N)$:

$$S_v = \otimes_{i=1}^N (-i)^{a_i b_i} X_i^{a_i} Z_i^{b_i} \tag{3.14}$$

In general two such Pauli strings do not commute. In order to find a set of stabilizers we need to find S_v that commutes with each other. The commutator of two Pauli strings S_u and S_v can be expressed as

$$T_i^{ab} T_i^{cd} = T_i^{cd} T_i^{ab} (-1)^{ad-bc} \tag{3.15}$$

For two stabiliziers corresponding to $v = (a_1, b_1, a_2, b_2, ..., a_N, b_N)$ and $u = (c_1, d_1, c_2, d_2, ..., c_N, d_N)$, we have

$$\begin{aligned} S_v S_u &= S_{v+u} i^{\sum_i (a_i b_i + c_i d_i)} (-i)^{\sum_i (a_i + c_i)(b_i + d_i)} (-1)^{\sum_i b_i c_i} \\ &= S_{v+u} (-i)^{\sum_i (a_i d_i - b_i c_i)} \end{aligned} \tag{3.16}$$

$$S_v S_u = S_u S_v \, (-1)^{\sum_i a_i d_i - b_i c_i} \tag{3.17}$$

Therefore the commutation relation is completely determined by the inner product

$$v^T J u = \sum_i a_i d_i - b_i c_i \tag{3.18}$$

modular 2. Here J is an antisymmetric matrix $J = \mathbb{I}_{N\times N} \otimes \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Since $S_u^2 = \mathbb{I}$, and $\mathrm{tr}(S_u) = 0$ if $u \neq 0$, then there are half of the 2^N states with eigenvalue $S_u = +1$. If $[S_u, S_v] = 0$ and $u \neq v,\ v \neq 0$, we have

$$\mathrm{tr}\left[\frac{\mathbb{I} + S_u}{2} S_v\right] = 0 \tag{3.19}$$

which means that in the subspace of $S_u = +1$, there are half which has $S_v = +1$ and the other half has $S_v = -1$. Thus the number of states satisfying $S_u |G\rangle = |G\rangle\,,\ S_v |G\rangle = |G\rangle$ is exactly 2^{N-2}. Continue this reasoning, we see that N mutually commuting and independent stabilizers will determine a unique state. The vectors $v_n,\ n = 1, 2, ..., N$ needs to be linearly independent in Z_2 field, and satisfy

$$v_n^T J v_m = 0 \bmod 2, \tag{3.20}$$

In this way, we have translate the problem of defining a stabilizer state to a linear algebra problem on Z_2 field.

Now for such a state defined by conditions $S_{v_n} |G\rangle = |G\rangle$, we want to determine the entanglement entropy of a subsystem. Naturally we will try to find local stabilizers that are supported on A. For example if we consider an Ising model with only two sites, the stabilizers

$$S_1 = Z_1 Z_2,\ S_2 = Z_1 \tag{3.21}$$

determine a product state $|G_+\rangle = |+\rangle\,|+\rangle$, because $S_2 = 1$ already determines the state of the site 1. In contrast, if we take

$$S_1 = Z_1 Z_2,\ S_2' = X_1 X_2 \tag{3.22}$$

there is no local stabilizer on site 1. The state stabilized by S_1 and S_2' is maximally entangled with entropy $\log 2$. In general, if we can M linear superpositions of v_n that are only supported in A (which means these vectors are zero outside A), the entanglement entropy will be $(|A| - M)\log 2$. Each local stabilizer reduces one qubit of entanglement. To see that, we denote the vectors $v_n, n = 1, 2, ..., N$ correspond to stabilizers as

$$v_n = \begin{pmatrix} q_n \\ p_n \end{pmatrix} \tag{3.23}$$

with q_n the components in A region. If the size of A is $|A|$, q_n has dimension $2|A|$. This corresponds to a factorization of stabilizer operator

$$S_{v_n} = S_{q_n}^A S_{p_n}^{\overline{A}} \tag{3.24}$$

with $S^A_{q_n}$ a stabilizer operator acting on the Hilbert space of A, and similar for $S^{\overline{A}}_{p_n}$. Since

$$S_{v_n} |G\rangle \langle G| S_{v_n} = |G\rangle \langle G| \tag{3.25}$$

tracing over $\overline{A}$ we obtain

$$S^A_{q_n} \rho_A S^A_{q_n} = \rho_A, \text{ or } \left[S^A_{q_n}, \rho_A\right] = 0 \tag{3.26}$$

Note that $S^A_{q_n}$ is not a set of stabilizers, since they don't necessarily commute each other. For example in the Ising model example $S_1 = Z_1 Z_2, S_2' = X_1 X_2$ the truncated operators Z_1, X_1 do not commute with each other. To determine ρ_A we can take a Z_2 linear superposition of v_n to define a new set of stabilizers

$$\tilde{v}_n = \begin{pmatrix} \tilde{q}_1 & \tilde{q}_2 & ... & \tilde{q}_M & \tilde{q}_{M+1} & ... & \tilde{q}_N \\ 0 & 0 & ... & 0 & \tilde{p}_{M+1} & ... & \tilde{p}_N \end{pmatrix} \tag{3.27}$$

Then $S^A_{\tilde{q}_n},\ n = 1, 2, ..., M$ commute with each other. These are local stabilizers on A, and we have

$$S^A_{\tilde{q}_n} \rho_A = \rho_A,\ n = 1, 2, ..., M \tag{3.28}$$

If $M < |A|$, then the condition above does not determine a unique state. In that case, it is possible to find $|A| - M$ other stabilizers in A, denoted by vectors $r_s, s = 1, 2, ...|A| - M$. The operators $S^A_{r_s}$ by definition commute with $S^A_{\tilde{q}_n},\ n = 1, 2, ..., M$ but they must anticommute with some of $S^A_{\tilde{q}_m},\ m > M$. Otherwise, they will be additional stabilizers which commute with all $S_{\tilde{v}_n}$, but there is no such an operator, since $S_{\tilde{v}_n}$ is already complete. Therefore the condition

$$\left[S^A_{\tilde{q}_m}, \rho_A\right] = 0,\ m > M \tag{3.29}$$

requires

$$\rho_A = 2^{M-|A|} \prod_{n=1}^{M} \frac{\mathbb{I}_A + S_{\phi^A_n}}{2} \tag{3.30}$$

which is a maximally mixed state in the subspace of $S^A_{\tilde{q}_n} = 1,\ n = 1, 2, ..., M$. Thus we have proven that the entropy

$$S_A = (|A| - M) \log 2 \tag{3.31}$$

For example, the Ising model with stabiliziers $Z_i Z_{i+1}$ and F has $L - 1$ stabilizers in region A if its size is L. Thus the entropy is $\log 2$ for all subsystems $L < N$. As another example, in the Z_2 gauge theory, if we consider a disk region with N internal vertices and P plaquettes, L links, the number of local stabilizer is $N + P$, and the number of qubits is L. Thus the entropy is

$$S_A = (L - N - P) \log 2 = (|\partial A| - \chi) \log 2 \tag{3.32}$$

Here $\chi = (N + N_d) + P - L$ is the Euler number, and N_d is the number of vertices at the end of dangling legs. $N_d = |\partial A|$. This reproduces the formula of topological entanglement entropy.

The stabilizer states can be generalized from qubits to qudits. In general when the dimension of each site is d, we can define

$$Z = \text{diag}\left[1,\ \omega,\ \omega^2, ..., \omega^{d-1}\right],\ \omega = e^{i\frac{2\pi}{d}} \tag{3.33}$$

$$X = \begin{pmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ ... & & & & 1 \\ 1 & 0 & 0 & ... & 0 \end{pmatrix} \tag{3.34}$$

They satisfy

$$Z^d = X^d = \mathbb{I},\ XZ = \omega ZX \tag{3.35}$$

A complete basis of operators on each site can be defined as

$$T_i^{ab} = X_i^a Z_i^b,\ a, b = 0, 1, ..., d-1 \tag{3.36}$$

These are unitary operators. The stabilizers can be defined by a $\mathbb{Z}_d$ vector $v = (a_1, b_1, a_2, b_2, ..., a_i, b_i)$:

$$S_v = \otimes X_i^{a_i} Z_i^{b_i} \tag{3.37}$$

They satisfy the same commutation equation as in the $\mathbb{Z}_2$ case:

$$S_u S_v = S_v S_u \omega^{u^T J v} \tag{3.38}$$

Stabiliziers are determined by a set of vectors v_n satisfying

$$v_n^T J v_m = 0 \text{ mod } d \tag{3.39}$$

Following the same argument as in the qubit case, for N qubits we can determine a state completely by N linearly independent stabilizers.

For example, the state (2.132) we discussed in previous section is a stabilizer state of 4 qutrits. It has the following stabilizers:

$$S_1 = Z_1^{-1} Z_2^{-1} Z_3,\ S_2 = Z_1 Z_2^{-1} Z_4,\ S_3 = X_1 X_2 X_3^{-1},\ S_4 = X_2 X_3 X_4 \tag{3.40}$$

One can check that there is no local stabilizers on any 2 of the 4 sites, which leads to $S_{ij} = 2\log 3$ maximal for any two sites i, j. This is the "perfect tensor" condition we mentioned earlier, that the mapping from any two qutrits to any other two qutrits is unitary.

The stabilizer state above is an example of CSS (CalderbankShor-Steane) state, which has all stabilizers only depend on either Z or X, but not both.

In the same way as defining stabilizer states, we can define stabilizer operators. For example the 4-qutrit state in Eq. (2.132) can be mapped to operators corresponding to its Schmit matrices:

$$U = \sum_{a,b=0,1,2} |a+b\rangle |a-b\rangle \langle a| \langle b| \tag{3.41}$$

This is a two-qubit gate which is unitary. The stabilizer conditions $S_n |\Psi\rangle = |\Psi\rangle$, $n = 1, 2, 3, 4$ becomes conditions of the form

$$UA_nU^\dagger = B_n,\ n = 1, 2, 3, 4 \tag{3.42}$$

$$A_1 = Z_1,\ B_1 = Z_1^{-1}Z_2^{-1} \tag{3.43}$$

$$A_2 = Z_2,\ B_2 = Z_1^{-1}Z_2 \tag{3.44}$$

$$A_3 = X_1,\ B_3 = X_1X_2 \tag{3.45}$$

$$A_4 = X_2,\ B_4 = X_1X_2^{-1} \tag{3.46}$$

Such unitaries are called Clifford gates. They are not generic unitaries since they map Pauli operators to Pauli operators, rather than superposition of Pauli's.

3.3 Gaussian states

Gaussian states refer to the ground state or thermal states of free bosons or free fermions. They are special because correlation functions satisfy the Wick theorem. The easiest way to see the Wick theorem is to write down a path integral for the partition with source. For example for the thermal state of a boson system with Hamiltonian

$$H = \sum_{i,j} b_i^\dagger h_{ij} b_j \tag{3.47}$$

we can define the path integral

$$Z[J] = \int Db_i(\tau) e^{-\int_0^\beta d\tau \left(\sum_i b_i^* \partial_\tau b_i + \sum_{i,j} b_i^* h_{ij} b_j - \left(b_i^* J_i + J_i^* b_i\right)\right)} \tag{3.48}$$

Time-ordered multi-point functions can be obtained by taking variation over $J(\tau)$:

$$\left\langle T_\tau b_{i_1}(\tau_1) b_{i_2}^\dagger(\tau_2) ... b_{i_n}(\tau_n) \right\rangle = \frac{1}{Z[0]} \frac{\delta^n Z[J]}{\delta J_{i_1}^*(\tau_1)\delta J_{i_2}(\tau_2)...\delta J_{i_n}^*(\tau_n)}\bigg|_{J=0} \tag{3.49}$$

Here $\langle ... \rangle$ refers to the thermal average $Z^{-1}\mathrm{tr}\left(e^{-\beta H}\right)$. Because (3.48) is Gaussian in b field, after the integration $Z[J]$ is also Gaussian in J:

$$\log Z[J] = \frac{1}{2}\int_0^\beta d\tau \int_0^\beta d\tau' J^*(\tau) C(\tau, \tau') J\left(\tau'\right) \tag{3.50}$$

This leads to the Wick theorem. All higher point correlation functions are determined by the two-point function

$$\left\langle T_\tau b_{i_1}(\tau_1) b_{i_2}^\dagger(\tau_2) \right\rangle = C_{i_1 i_2}(\tau_1, \tau_2) \tag{3.51}$$

When we discuss entanglement properties, we focus on a quantum state, so that we only need to consider equal time correlation functions. The equal time correlators can be taken as a special limit of the time-ordered correlators. For example if we normal order the correlators,

$$\left\langle b_{i_1}^\dagger b_{i_2}^\dagger b_{i_3} b_{i_4} \right\rangle = \lim_{\epsilon \to 0^+} \left\langle T_\tau b_{i_1}^\dagger(\epsilon) b_{i_2}^\dagger(\epsilon) b_{i_3}(0) b_{i_4}(0) \right\rangle \tag{3.52}$$

The Wick theorem applies in general to states obtained from Gaussian path integral, which includes the thermal state, the ground state, and their time-dependent generalization. For example if $|G\rangle$ is a Gaussian state, and we have a time-dependent Hamiltonian $H(t) = \sum_i b_i^\dagger h_{ij}(t) b_j$, the time-evolved state

$$|\Psi(t)\rangle = T_t e^{-i\int dt H(t)} |G\rangle \tag{3.53}$$

is also a Gaussian state.

An interesting observation is that the Wick theorem also applies to a subsystem. If we take a subsystem A, and restrict $i \in A$ in all indices in a correlation function, then it is determined by the two point function in the same region A following the same rules of the Wick theorem. Since the Wick theorem determines all correlation functions, it actually determines the reduced density operator. Denote

$$C^A_{i_1 i_2} = \left\langle b_{i_1} b^\dagger_{i_2} \right\rangle, \ i_1, i_2 \in A \tag{3.54}$$

as the two-point function in A region, C^A is simply a block in the matrix of two-point correlator in the entire system. Since C^A determines all higher-point correlation functions, it also determines the reduced density operator ρ_A. Furthermore, we know that ρ_A is actually Gaussian—because the correlation functions satisfy the Wick theorem. Following this reasoning, we can explicitly compute ρ_A by assuming the Gaussian form

$$\rho_A = Z_A^{-1} \exp\left[-\sum_{i,j\in A} b_i^\dagger K^A_{ij} b_j \right] \tag{3.55}$$

$H^A_E \equiv \sum_{i,j\in A} b_i^\dagger K^A_{ij} b_j = -\log \rho_A$ is usually called the "entanglement Hamiltonian". To write K^A explicitly in term of C^A, we assume K^A is diagonalized into

$$K^A = V \Lambda V^\dagger \tag{3.56}$$

Then defining $\gamma_n = \left(V^\dagger b\right)_n = \sum_j V^*_{jn} b_j$, and denote the eigenvalues of K^A as λ_n, we have

$$H^A_E = \sum_n \lambda_n \gamma_n^\dagger \gamma_n \tag{3.57}$$

such that the two-point function is

$$\left\langle \gamma_n \gamma_m^\dagger \right\rangle = \delta_{nm} \frac{1}{1 - e^{-\lambda_n}} \tag{3.58}$$

Express this back in b we obtain

$$\left\langle b_i b_j^\dagger \right\rangle = V_{in} \left\langle \gamma_n \gamma_n^\dagger \right\rangle V^*_{jn} = \left[\left[\mathbb{I} - e^{-K^A} \right]^{-1} \right]_{ij} \tag{3.59}$$

In other words, in the matrix form

$$C^A = \left[\mathbb{I}_A - e^{-K^A} \right]^{-1} \tag{3.60}$$

$$K^A = -\log\left[\mathbb{I}_A - C_A^{-1} \right] \tag{3.61}$$

Since entanglement quantities such as the von Neumann entropy and Renyi entropy can be completely determined by eigenvalues of ρ_A, we can directly express them in C^A. The Renyi entropy is defined as

$$S_A^{(n)} = -\frac{1}{n-1}\log \mathrm{tr}\rho_A^n \tag{3.62}$$

which approaches the von Neumann entropy in the $n \to 1$ limit. Using the diagonalized entanglement Hamiltonian we can write

$$\begin{aligned}\mathrm{tr}\left(\rho_A^n\right) &= \frac{\mathrm{tr}e^{-n\sum_k \lambda_k \gamma_k^\dagger \gamma_k}}{\left(\mathrm{tr}e^{-\sum_k \lambda_k \gamma_k^\dagger \gamma_k}\right)^n} \\ &= \prod_k \frac{\left(1-e^{-\lambda_k}\right)^n}{1-e^{-n\lambda_k}} = \frac{\det^n\left[\mathbb{I}_A - e^{-K^A}\right]}{\det\left[\mathbb{I}_A - e^{-nK^A}\right]}\end{aligned} \tag{3.63}$$

We can express everything in C^A:

$$\begin{aligned}\mathrm{tr}\left(\rho_A^n\right) &= \frac{\det^n\left[C^{A-1}\right]}{\det\left[\mathbb{I}_A - \left(\mathbb{I}_A - C^{A-1}\right)^n\right]} \\ &= \det{}^{-1}\left[C^{A^n} - \left(C^A - \mathbb{I}\right)^n\right]\end{aligned} \tag{3.64}$$

Thus

$$\begin{aligned}S_A^{(n)} &= \frac{1}{n-1}\log\det\left[C^{A^n} - \left(C^A - \mathbb{I}\right)^n\right] \\ &= \frac{1}{n-1}\mathrm{tr}\log\left[C^{A^n} - \left(C^A - \mathbb{I}\right)^n\right]\end{aligned} \tag{3.65}$$

Take the $n \to 1$ limit we get

$$S_A^{(n)} = \mathrm{tr}\left[C_A \log C_A - \left(C^A - \mathbb{I}_A\right)\log\left(C^A - \mathbb{I}_A\right)\right] \tag{3.66}$$

In summary, we can compute the Renyi entropy and von Neumann entropy of the boson system by computing the two-point correlation function in the entire system, taking a truncation to the subsystem and applying Eq. (3.65) and Eq. (3.66).[10]

The discussion here can be generalized to more general boson systems. The most general Gaussian boson system has one-point function and two-point function. For region A with boson operators $b_i, i \in A$, define

$$\langle b_i \rangle = \phi_i, \quad \left\langle \begin{pmatrix} b_i \\ b_i^\dagger \end{pmatrix} \begin{pmatrix} b_j^\dagger & b_j \end{pmatrix} \right\rangle = G_{nm}^A \tag{3.67}$$

Here G^A is a $2N \times 2N$ matrix if $i = 1, 2, ..., N$. We want to use these expectation values to determine the entanglement Hamiltonian

$$H_E^A \equiv -\log\rho_A = \frac{1}{2}\begin{pmatrix} b_i^\dagger & b_i \end{pmatrix} K^A \begin{pmatrix} b_j \\ b_j^\dagger \end{pmatrix} + \sum_i \left(b_i^\dagger v_i + v_i^* b_i\right) \tag{3.68}$$

Define

$$\begin{pmatrix} h_i \\ h_i^* \end{pmatrix} = K^{A^{-1}} \begin{pmatrix} v_i \\ v_i^* \end{pmatrix} \tag{3.69}$$

we can write

$$H_E^A = \frac{1}{2} \begin{pmatrix} \tilde{b}_i^\dagger & \tilde{b}_i \end{pmatrix} K^A \begin{pmatrix} \tilde{b}_j \\ \tilde{b}_j^\dagger \end{pmatrix} - \text{const.} \tag{3.70}$$

$$\tilde{b}_i = b_i + h_i \tag{3.71}$$

$\tilde{b}_i$ satisfies the same canonical commutation relation as b_i, so that we can compute correlation functions in $\tilde{b}_i$. K^A can be diagonalized by a symplectic transformation:

$$K^A = V \Lambda V^\dagger, \; V^\dagger Z V = Z \tag{3.72}$$

Here $Z = \text{diag}[1, 1, .., 1, -1, -1, ..., -1]$ is the commutator matrix

$$Z = \left[\begin{pmatrix} b_i \\ b_i^\dagger \end{pmatrix}, \begin{pmatrix} b_j^\dagger & b_j \end{pmatrix} \right] \tag{3.73}$$

The symplectic condition $V^\dagger Z V = Z$ guarantees that

$$\begin{pmatrix} \gamma_n \\ \gamma_n^\dagger \end{pmatrix} = V^\dagger \begin{pmatrix} b_i \\ b_i^\dagger \end{pmatrix} \tag{3.74}$$

satisfies the boson commutation relation. K^A needs to satisfy

$$X K^A X = K^{A^T} \tag{3.75}$$

to remove redundancies. (For example if $K^A = Z$, the Hamiltonian is a constant $-N/2$.) Due to this condition, the eigenvalues are $\Lambda = \text{diag}\left[\lambda_1, \lambda_2, ..., \lambda_N, \lambda_1, \lambda_2, ..., \lambda_N\right]$. The Hamiltonian is (up to a constant)

$$H_E^A = \sum_n \lambda_n \gamma_n^\dagger \gamma_n \tag{3.76}$$

The one-point and two-point functions are

$$\phi_i = -h_i + \left\langle \tilde{b}_i \right\rangle = -h_i \tag{3.77}$$

$$\begin{aligned} G^A &= \left\langle \begin{pmatrix} \tilde{b}_i - h_i \\ \tilde{b}_i^* - h_i^* \end{pmatrix} \begin{pmatrix} \tilde{b}_j^\dagger - h_j^* & \tilde{b}_j - h_j \end{pmatrix} \right\rangle \\ &= V \begin{pmatrix} \left\langle \gamma_n \gamma_n^\dagger \right\rangle & 0 \\ 0 & \left\langle \gamma_n^\dagger \gamma_n \right\rangle \end{pmatrix} V^\dagger + \begin{pmatrix} h_i \\ h_i^* \end{pmatrix} \begin{pmatrix} h_j^* & h_j \end{pmatrix} \\ &= \frac{1}{2} V \left(Z + \coth \frac{\Lambda}{2} \right) V^\dagger + \begin{pmatrix} h_i \\ h_i^* \end{pmatrix} \begin{pmatrix} h_j^* & h_j \end{pmatrix} \end{aligned} \tag{3.78}$$

Using $V^\dagger = ZV^{-1}Z$ and the fact that $\coth(x)$ is an odd function, we can write

$$V \coth \frac{\Lambda}{2} V^\dagger = V \coth\left(\frac{\Lambda Z}{2}\right) V^{-1} Z = \coth \frac{K^A Z}{2} Z \tag{3.79}$$

Thus

$$G^A = \frac{1}{2}\left(\mathbb{I} + \coth \frac{K^A Z}{2}\right) Z + \begin{pmatrix} h_i \\ h_i^* \end{pmatrix} \begin{pmatrix} h_j^* & h_j \end{pmatrix} \tag{3.80}$$

The Renyi entropy and von Neumann entropy can be expressed as

$$\begin{aligned} S_A^{(n)} &= \frac{1}{n-1} \sum_k \log \frac{1 - e^{-n\lambda_k}}{(1 - e^{-\lambda_k})^n} \\ &= \frac{1}{n-1} \frac{1}{2} \mathrm{tr} \log \left[\left(\tilde{G}^A Z\right)^n Z^{n+1} - \left(\tilde{G}^A Z - \mathbb{I}_A\right)^n Z^{n+1} \right] \qquad (3.81) \\ S_A &= \frac{1}{2} \mathrm{tr} \left[\tilde{G}^A Z \log\left(\tilde{G}^A\right) - \left(\tilde{G}^A Z - \mathbb{I}_A\right) \log\left(\tilde{G}^A - Z\right) \right] \qquad (3.82) \end{aligned}$$

with

$$\tilde{G}^A = G^A - \begin{pmatrix} \phi_i \\ \phi_i^* \end{pmatrix} \begin{pmatrix} \phi_j^* & \phi_j \end{pmatrix} \tag{3.83}$$

One can check that for charge conserved boson Eq. (3.81) and (3.82) return to Eq. (3.65) and (3.66) respectively.

The discussion here can be carried completely in parallel for fermions. For general fermion models we can express the entanglement Hamiltonian in Majorana fermion basis χ_i satisfying the anticommutation relation

$$\{\chi_i, \chi_j\} = 2\delta_{ij} \tag{3.84}$$

The entanglement Hamiltonian is

$$H_E^A = \frac{1}{4} \sum_{i,j} \chi_i K_{ij}^A \chi_j \tag{3.85}$$

with h a purely imaginary, anti-symmetric matrix. Diagonalization of this matrix is done

by an orthogonal matrix

$$K^A = O \begin{pmatrix} 0 & -i\lambda_1 & ... & & \\ i\lambda_1 & 0 & & & \\ ... & ... & ... & & \\ & & & 0 & -i\lambda_N \\ & & & i\lambda_N & 0 \end{pmatrix} O^T \tag{3.86}$$

$$\begin{pmatrix} \chi_1 \\ \chi_2 \\ ... \\ chi_{2N} \end{pmatrix} = O \begin{pmatrix} a_1 \\ b_1 \\ ... \\ a_N \\ b_N \end{pmatrix} \tag{3.87}$$

$$H_E^A = -\sum_{k=1}^{N} i a_k b_k \frac{\lambda_k}{2} = \sum_k \left(\gamma_k^\dagger \gamma_k - \frac{1}{2}\right) \lambda_k \tag{3.88}$$

$$\text{with } \gamma_k = \frac{1}{2}\left(a_k - i b_k\right) \tag{3.89}$$

The two-point function is given by

$$G_{ij}^A = \langle i\chi_i \chi_j \rangle = \tanh \frac{K^A}{2} \tag{3.90}$$

Then we have

$$\mathrm{tr}\left(\rho_A^n\right) = \prod_k \frac{\left(1 + e^{n\lambda_k}\right)}{\left(1 + e^{\lambda_k}\right)^n} \tag{3.91}$$

$$S_A^{(n)} = -\frac{1}{2}\frac{1}{n-1}\mathrm{tr}\log\left[\left(\frac{\mathbb{I}_A + G^A}{2}\right)^n + \left(\frac{\mathbb{I}_A - G^A}{2}\right)^n\right] \tag{3.92}$$

$$S_A = -\frac{1}{2}\mathrm{tr}\left[\frac{\mathbb{I}_A + G^A}{2}\log\frac{\mathbb{I}_A + G^A}{2} + \frac{\mathbb{I}_A - G^A}{2}\log\frac{\mathbb{I}_A - G^A}{2}\right] \tag{3.93}$$

For two regions A, B we can denote the correlation matrix as

$$G^{AB} = \begin{pmatrix} G^A & F^{AB} \\ F^{AB\dagger} & G^B \end{pmatrix} \equiv G_0^{AB} + \delta G^{AB} \tag{3.94}$$

with

$$G_0^{AB} = \begin{pmatrix} G^A & 0 \\ 0 & G^B \end{pmatrix} \tag{3.95}$$

$$\delta G^{AB} = \begin{pmatrix} 0 & F^{AB} \\ F^{AB\dagger} & 0 \end{pmatrix} \tag{3.96}$$

such that

$$S_{AB} = \frac{1}{2}h\left(G^{AB}\right),\ S_A + S_B = \frac{1}{2}h\left(G_0^{AB}\right) \tag{3.97}$$

with $h(x) = -\frac{1+x}{2}\log\frac{1+x}{2} - \frac{1-x}{2}\log\frac{1-x}{2}$. If we consider two regions that are weakly correlated, with a small off-diagonal term F^{AB}, we can expand the mutual information in F^{AB}. The leading order term is second-order:

$$\begin{aligned} I\left(A:B\right) &\simeq \frac{1}{2}\mathrm{tr}\left[\left(\mathbb{I}+G_0^{AB}\right)^{-1}\delta G^{AB}\left(\mathbb{I}-G_0^{AB}\right)^{-1}\delta G^{AB}\right] \\ &= \mathrm{tr}\left[\left(\mathbb{I}^A+G^A\right)^{-1}F^{AB}\left(\mathbb{I}^B-G^B\right)^{-1}F^{AB\dagger}\right]+\mathrm{tr}\left[\left(\mathbb{I}^A-G^A\right)^{-1}F^{AB}\left(\mathbb{I}^B+G^B\right)^{-1}F^{AB\dagger}\right] \end{aligned} \tag{3.98}$$

If we consider three regions A, B, C and assume the off-diagonal term F^{AB}, F^{BC} and F^{AC} are small, expanding to the leading order of the off-diagonal term will lead to

$$I(A:BC) \simeq I(A:B) + I(A:C) \tag{3.99}$$

or $I_3(A:B:C) \simeq 0$. This is a consequence of the fact that two-point functions determines all higher point functions, and confirms that entanglement in Gaussian states are mostly like EPR pairs (but not exactly).

For fermion with charge conservation, we can simplify the formula. Define

$$C^A = \left\langle c_i c_j^\dagger \right\rangle \tag{3.100}$$

The entanglement Hamiltonian is

$$H_E^A = -\log\rho_A = -c_A^\dagger \log\left(C^{A^{-1}} - \mathbb{I}\right) c_A \tag{3.101}$$

$$S_A = -\mathrm{tr}\left[C^A \log C^A + \left(\mathbb{I}^A - C^A\right)\log\left(\mathbb{I}^A - C^A\right)\right] \tag{3.102}$$

As an example let's consider a two-leg ladder with the Hamiltonian

$$\begin{aligned} H &= H_1 + H_2 + H_{12} \\ H_1 &= -t\sum_n \left(c_{n1}^\dagger c_{n+1,1} + h.c.\right) && (3.103) \\ H_2 &= t\sum_n \left(c_{n2}^\dagger c_{n+1,2} + h.c.\right) && (3.104) \\ H_{12} &= v\sum_n \left(c_{n1}^\dagger c_{n2} + h.c.\right) && (3.105) \end{aligned}$$

We want to study the entanglement between the two chains in the ground state. In k space we have

$$H = \sum_k \begin{pmatrix} c_{k1}^\dagger & c_{k2}^\dagger \end{pmatrix} \begin{pmatrix} -2t\cos k & v \\ v & 2t\cos k \end{pmatrix} \begin{pmatrix} c_{k1} \\ c_{k2} \end{pmatrix} \tag{3.106}$$

The single particle energy eigenvalues are $\pm E_k \equiv \pm\sqrt{4t^2\cos k^2 + v^2}$. The two-point function matrix is

$$C_k = \frac{1}{2}\left(\mathbb{I} + \left(\cos\theta_k\sigma_z + \sin\theta_k\sigma_x\right)\right) \tag{3.107}$$

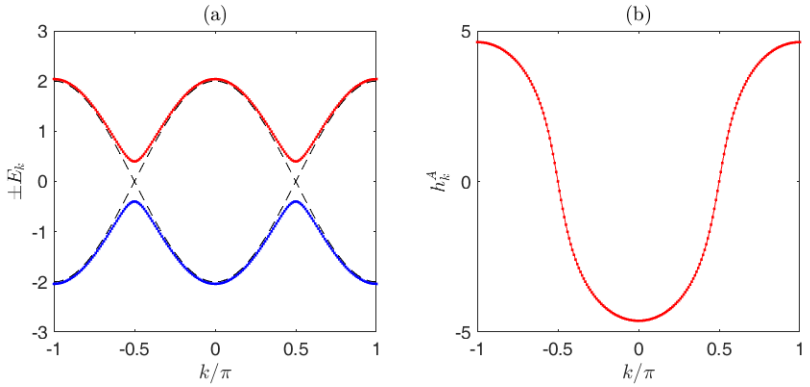

Figure 12. Entanglement Hamiltonian eigenvalue for the chain 1 in a two-leg ladder model.

with

$$\cos\theta_k = \frac{-2t\cos k}{\sqrt{4t^2\cos^2 k + v^2}}, \quad \sin\theta_k = \frac{v}{\sqrt{4t^2\cos^2 k + v^2}} \tag{3.108}$$

The two-point correlator in the upper chain c_{k1} is

$$C_{1k} = \left\langle c_{k1} c_{k1}^\dagger \right\rangle = \frac{1+\cos\theta_k}{2} \tag{3.109}$$

Thus the entanglement Hamiltonian of chain 1 is

$$H = \sum_k c_{1k}^\dagger h_k^A c_{1k} \tag{3.110}$$

$$h_k^A e = -\log\left(\frac{2}{1+\cos\theta_k} - 1\right) = -2\log\tan\frac{\theta_k}{2} \tag{3.111}$$

Near the Fermi point (for $v = 0$) $k_F = \frac{\pi}{2}$, we have

$$\begin{aligned} \cos\theta_k &\simeq \frac{2t}{v}(k - k_F) \\ \theta_k &\simeq -\frac{\pi}{2} + \frac{2t}{v}(k - k_F) \end{aligned} \tag{3.112}$$

Thus

$$h_k \simeq \frac{4t}{v}(k - k_F) \tag{3.113}$$

This simple model illustrates a feature that we often see: when two gapless systems are coupled, leading to a gap opening, then there is often a gapless spectrum in the entanglement Hamiltonian. This is because the lowest energy states (here near the Fermi points $\pm\frac{\pi}{2}$ for the $v = 0$ model) entangles most when we turn on a coupling v.

In the following we will discuss some properties of free Dirac fermions, without rigorous proof. For a $1+1$-d massless Dirac fermion, the two-point function is

$$C^A(x, y) \propto \frac{1}{x - y} \tag{3.114}$$

The entanglement entropy of a single interval with length ℓ is

$$S_\ell = \frac{1}{3}\log\frac{\ell}{a} \tag{3.115}$$

Here a is a UV cutoff scale. (For this result and multi-interval generalizations, see Ref.[11].) For a massive fermion, the $\log\ell$ behavior is replaced by a finite entanglement entropy

$$S_\ell = \frac{1}{3}\log\frac{\xi}{a} \tag{3.116}$$

with $\xi \sim m^{-1}$ the correlation length. For higher dimensional Dirac fermion, we can take a strip region A, which has translation symmetry along the strip direction. In this case momentum $k_\parallel$ are good quantum numbers. With fixed $k_\parallel$ the vertical direction fermion looks like a massive fermion with mass $\left|k_\parallel\right|$. The entanglement entropy is a sum

$$S = \left(\frac{L}{2\pi}\right)^{d-1}\int d^{d-1}k_\parallel \frac{1}{3}\log\frac{1}{\left|k_\parallel\right| a} \tag{3.117}$$

For $d > 1$, this integration is finite in IR (small $k_\parallel$). There is still a UV divergence but that is expected. This formula suggests that the entanglement entropy satisfies area law ($\propto L^{d-1}$) in dimension $d > 1$ even if the fermion is gapless. This is generally true for gapless systems in $d > 1$ but there is an interesting exception. Fermi liquid has a super-area law entanglement entropy[12–14]:

$$S = \left(\frac{L}{2\pi}\right)^{d-1}\log L\frac{1}{12}\int_{FS} d^{d-1}k\int_{\partial A} d^{d-1}r\,|\mathbf{n}_r\cdot\mathbf{n}_k| \tag{3.118}$$

Here $\mathbf{n}_r$ and $\mathbf{n}_k$ are normal vectors at the boundary of the spatial region A and that at the boundary of Fermi sea in k space, respectively. The region is a cube with edge length L. For $1+1$-d fermion there are 2 fermi points and 2 end points of the interval, which reproduces the 1-dimensional formula. If we consider decoupled 1-dimensional wires along different directions in a higher-dimensional system, their contribution to the cube entropy just adds, reproducing the formula above. Fermi gas (and Fermi liquid) is a rare example of a ground state of local Hamiltonian which has super-area-law entropy.

3.4 Conformal field theory

The logarithmic dependence of entropy in $1+1$-dimensional gapless fermion is actually a generic behavior of $1+1$-dimensional conformal field theory.[15, 16]

For a generic quantum field theory, the ground state wavefunction is prepared by an imaginary time path integral. Schematically, if the theory has an action $S[\phi]$ which depends on a field $\phi(x,\tau)$, the wavefunctional at time $\tau = 0$ is

$$\Psi\left[\psi(x)\right] = \int_{\phi(x,\tau=0)=\psi(x)} D\phi e^{-S[\phi]} \tag{3.119}$$

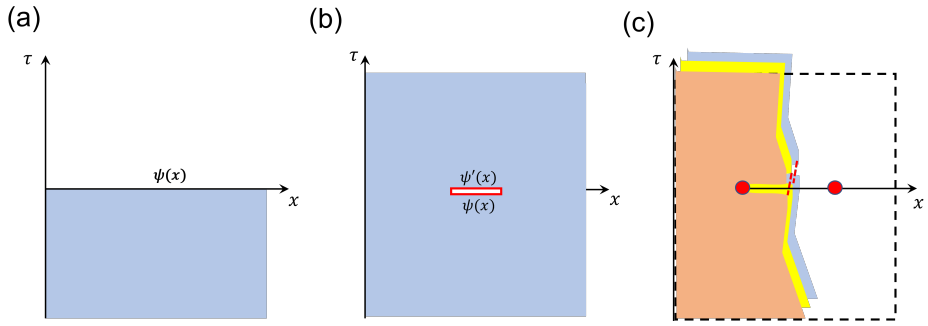

Figure 13. (a) A QFT state prepared by a path integral of the lower half plane. (b) A reduced density operator of region A (red interval) prepared by a path integral with boundary condition at upper and lower boundary of A. (c) $\mathrm{tr}\,(\rho_A^n)$ corresponds to a partition function on the Riemann surface, with branching points at the end of A.

The path integral is carried at the lower half plane $\tau \leq 0$ with a boundary condition at $\tau = 0$. Now we consider the n-th Renyi entropy of a region A which is an interval $x \in [z_1, z_2]$. The reduced density operator is

$$\rho\left[\psi(x), \psi'(x)\right] = \int_{\phi(x,\tau=0^+)=\psi'(x), \phi(x,\tau=0^-)=\psi(x)} D\phi e^{-S[\phi]} \tag{3.120}$$

The path integral is carried on a full plane, with an opening at region A (so that the field can be discontinuous at $\tau = 0$ for $x \in [z_1, z_2]$). Now if we compute the n-th Renyi entropy of A, this corresponds to a path integral on a Riemann surface, which is made by n copies of the plane connected across region A, as is shown in Fig. 13 (c). This is like the Riemann surface for function $\left(\frac{z-z_1}{z-z_2}\right)^{1/n}$. Thus we get

$$\mathrm{tr}\,(\rho_A^n) = \frac{Z_A^{(n)}}{Z_\emptyset^{(n)}} \tag{3.121}$$

Here $Z_A^{(n)}$ is the path integral on the Riemann surface defined by A, while $Z_\emptyset^{(n)}$ is that for trivial region, which is simply the original partition function Z_0 to the power n.

Interestingly, this replica formula (3.121) relates Renyi entropy with geometrical response. If the theory is defined on a generic curved space, with the action $S[\phi, g_{\mu\nu}]$, then $Z_A^{(n)}$ is the partition function on a particular background geometry. This fact plays a key role in the relation between entanglement entropy and spacetime geometry in holographic duality, which will be a topic in the next section.

Notice that Eq. (3.121) does not depend on the location of the branchcut, but only depend on the location of the branching points z_1, z_2. Thus we can view this as a two-point function

$$\mathrm{tr}\,(\rho_A^n) = \left\langle X_n^\dagger(z_1) X_n(z_2) \right\rangle \tag{3.122}$$

where $X_n(z)$ is a twist operator which creates a branching point (also called a conical singularity). $X_n(z)$ acts in the Hilbert space of the n-copied theory. It is not really a local operator. If we gauge the permutation symmetry between replica, this operator $X_n^\dagger(z_1)X_n(z_2)$

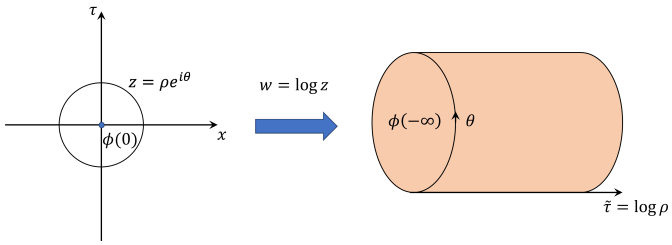

Figure 14. The log map between plane and infinite cylinder. The radial direction is mapped to the infinite direction. An operator at the origin of the plane is mapped to an operator at past infinity of the cylinder system.

is actually a t'Hooft line operator, which inserts a pair of flux, similar to how a Wilson line operator inserts a pair of charge.

The discussion so far applies to all quantum field theories. Now we consider conformal field theories. The conformal transformations are diffeomorphism which transforms the metric only by a prefactor. A general diff transformation is

$$x^\mu \to \tilde{x}^\mu\left(x^\nu\right),\ g_{\mu\nu} \to \tilde{g}_{\mu\nu}\left(\tilde{x}\right) = g_{\sigma\tau}(x)\frac{\partial x^\sigma}{\partial \tilde{x}^\mu}\frac{\partial x^\tau}{\partial \tilde{x}^\nu} \tag{3.123}$$

Conformal transformations for flat space are coordinate transformations that satisfy

$$\eta_{\sigma\tau}\frac{\partial x^\sigma}{\partial \tilde{x}^\mu}\frac{\partial x^\tau}{\partial \tilde{x}^\nu} = e^{\phi(\tilde{x})}\eta_{\mu\nu} \tag{3.124}$$

with $\eta_{\mu\nu}$ the Lorenzian metric. A conformal field theory is a theory that is invariant under conformal transformations. Since dilation $x_\mu \to \lambda x_\mu$ is part of conformal group, conformal invariance implies scale invariance. As a consequence, correlation functions follow power law in the ground state of conformal field theory.

$$\langle\phi(x)\phi(y)\rangle \propto (x-y)^{-2\Delta} \tag{3.125}$$

where Δ is the scaling dimension of ϕ field. If we define the transformation $\phi(x) \to \phi(\lambda(x))\lambda^\Delta$, then the correlation function remains invariant. For example, in a free scalar theory $S\left[\phi\right] = \int d^2x\partial_\mu\phi\partial^\mu\phi$, the scaling dimension of $\partial_\mu\phi$ is 1. More precisely, the operators which are (Virasoro) primary fields follow the power law correlation, with Δ their scaling dimension. A generic operator could be a superposition of different primary fields.

Two-point function of the twist operator also follows the power law

$$\left\langle X_n^\dagger(z_1)X_n(z_2)\right\rangle \propto \frac{1}{\left|z_1-z_2\right|^{2\Delta_n}} \tag{3.126}$$

We need to decide the scaling dimension. Scaling dimension of an operator is related to energy of a ground state on a circle. As is shown in Fig. (14), a plane and an infinite cylinder are related by a log map $w = \log(z)$, which is a local conformal transformation (in this example, it is a conformal transformation everywhere except $z = 0$ and $z = \infty$). w lives on a cylinder, with period 2π. Different operator at $z = 0$ are mapped to different

boundary conditions on the cylinder. If we view the infinite direction of the cylinder (Rew as imaginary time, and the periodic direction as space, each operator $\phi(z=0)$ is mapped to a ground state on the circle. Different operators ϕ corresponds to ground states in different sectors. The dilation $z \to \lambda z$ is mapped to $w \to w + \log\lambda$ with real λ, so that the eigenvalue of dilation, *i.e.* the scaling dimension Δ, is mapped to that of time translation, *i.e.* energy. If we consider a cylinder with general perimeter L, the energy of the ground state is

$$E_\Delta = \left(\Delta - \frac{c}{12}\right)\frac{2\pi}{L} \tag{3.127}$$

c is a constant called the central charge. This is a consequence of conformal anomaly, which means local conformal transformation such as the log map does not really preserves the partition function. Thus even if operator ϕ is trivial, the ground state energy of corresponding sector on cylinder is still nonzero.

[XLQ:To help understanding this result, we can compute the ground state energy $-\frac{c}{12}\frac{2\pi}{L}$ for free fermions with $c=1$ for complex fermion and $c=\frac{1}{2}$ for Majorana fermion.]

Now we consider the state in n copies of the CFT corresponding to the twist operator. The ground state energy without twist is n copies of the ground state energy

$$E_{n0} = -n\frac{c}{12}\frac{2\pi}{L} \tag{3.128}$$

With the twist, the system becomes a cylinder with perimeter nL:

$$E_n = -\frac{c}{12}\frac{2\pi}{nL} \tag{3.129}$$

The energy difference is

$$\Delta E_n = E_n - E_{n0} = \frac{c}{12}\left(n - \frac{1}{n}\right)\frac{2\pi}{L} \tag{3.130}$$

This corresponds to the scaling dimension

$$\Delta_n = \frac{c}{12}\left(n - \frac{1}{n}\right) \tag{3.131}$$

Therefore

$$S_A^{(n)} = -\frac{1}{n-1}\log\left\langle X_n^\dagger(z_1)X_n(z_2)\right\rangle = \frac{c}{6}\left(1+\frac{1}{n}\right)\log|z_1 - z_2| \tag{3.132}$$

The entropy of multi-interval regions in a CFT can also be computed by partition function on Riemann surfaces, or twist operator multipoint functions, but the result will depend on more details of the CFT, and is not entirely determined by c.

3.5 Random state

In the sections above, we have discussed several families of quantum states that are special enough so that we can compute its entanglement entropy. Another kind of states we often come across are more difficult to describe. For example consider a system of many spins

in an interacting model, with an initial state of $|000...0\rangle$. After a long time evolution by a Hamiltonian we obtain the state $|\Psi\rangle = e^{-iHt}|000...0\rangle$. In general it is difficult to determine and characterize the state $|\Psi\rangle$. Interestingly, we can also consider the opposite limit: pick a completely random state in the Hilbert space and study its entanglement property. What does "completely random" mean? There is a way to make that precise by the symmetry of the ensemble. We want to define an ensemble of pure quantum states $|\Psi\rangle$ with probability density $P(|\Psi\rangle)$. Then we can require the probability density to satisfy

$$P(U|\Psi\rangle) = P(|\Psi\rangle), \ \forall \ U \tag{3.133}$$

Since a unitary operator U can take any state to any other state, this condition completely determines the probability distribution. This is the uniform distribution on the manifold of quantum states.

Now we are interested in the entropy of a subsystem. Denote the subsystem Hilbert space dimension as D_A, and the complement dimension as D_B, so that the total Hilbert space dimension is $D = D_A D_B$. It is of course not possible to study the entropy of a particular state if we don't know more information about it, but we can study the ensemble average of entanglement quantities. The von Neumann entropy is difficult, and as usual we start with Renyi entropies of integer n. Let's start with the second Renyi entropy $n = 2$. We want to compute the ensemble average of $\mathrm{tr}\left(\rho_A^2\right)$:

$$\mathrm{tr}\left(\rho_A^2\right) = \mathrm{tr}\left[X_A \overline{|\Psi\rangle\langle\Psi| \otimes |\Psi\rangle\langle\Psi|}\right] \tag{3.134}$$

Here X_A is a swap operator that permutes between two replica of A region. For example if we define a basis $|na\rangle, \ n = 1, 2, ..., D_A, \ a = 1, 2, ..., D_B$, then X_A is defined by

$$X_A |na\rangle |mb\rangle = |ma\rangle |nb\rangle \tag{3.135}$$

The average of state can be determined by the symmetry of the ensemble.

$$\pi_2 \equiv \overline{|\Psi\rangle\langle\Psi| \otimes |\Psi\rangle\langle\Psi|} \tag{3.136}$$

$$U \otimes U \pi_2 U^\dagger \otimes U^\dagger = \pi_2 \ , \forall U \tag{3.137}$$

The only operators in the doubled Hilbert space that commute with all $U \otimes U$ are identity operator and swap operator $X = X_A X_B$ which permutes the two replica of both A and B. Thus we have

$$\pi_2 = a\mathbb{I} + bX \tag{3.138}$$

The coefficients a, b can be determined by another condition:

$$X\pi_2 = \pi_2 \tag{3.139}$$

This is a consequence that the two replicas are identical pure states. Using this condition we determine

$$\pi_2 = \frac{1}{D + D^2}(\mathbb{I} + X) \tag{3.140}$$

This discussion can be further generalized to n copies:

$$\mathrm{tr}\left(\rho_A^n\right) = \mathrm{tr}\left(X_{An}\pi_n\right) \tag{3.141}$$

$$\pi_n \equiv \overline{|\Psi\rangle\langle\Psi|^{\otimes n}} \tag{3.142}$$

$$\left[\pi_n, U^{\otimes} n\right] = 0 \tag{3.143}$$

$$X_g\pi_n = 0 \ \forall g \in S_n \tag{3.144}$$

X_g is a permutation of different replicas, labeled by permutation element g. More precisely

$$X_g\left|m_1a_1\right\rangle\left|m_2a_2\right\rangle ... \left|m_na_n\right\rangle = \left|m_{g(1)}a_1\right\rangle\left|m_{g(2)}a_2\right\rangle ... \left|m_{g(n)}a_n\right\rangle \tag{3.145}$$

X_{An} denote the same kind of permutation operator acting on A, for the cyclic permutation.

These conditions determine

$$\pi_n = \frac{1}{\Omega_n(D)}\sum_{g\in S_n} X_g \tag{3.146}$$

The normalization constant

$$\Omega_n(D) = \sum_{g\in S_n}\mathrm{tr}\left(X_g\right) = \sum_{g\in S_n} D^{l(g)} = \frac{(D-1+n)!}{(D-1)!} \tag{3.147}$$

Here $l(g)$ is the number of cycles in g. For example for the cyclic permutation $l(g) = 1$ and for identity $l(g) = n$.

Now we use this result to compute $\mathrm{tr}\left(\rho^n\right)$. For $n = 2$, we have

$$\begin{aligned}\overline{\mathrm{tr}\left(\rho_A^2\right)} &= \mathrm{tr}\left(\pi_2X_A\right)\\ &= \frac{1}{D(D+1)}\left(D_AD_B^2 + D_A^2D_B\right) = \frac{D_B+D_A}{D_AD_B+1}\end{aligned} \tag{3.148}$$

If A contains $|A|$ qubits and B contains $|B| = N - |A|$ qubits, we have

$$\overline{\mathrm{tr}\left(\rho_A^2\right)} = \frac{2^{|A|}+2^{N-|A|}}{2^N+1} \tag{3.149}$$

For $|A| < N/2$, $N \gg 1$, the second term dominants and we have $\overline{\mathrm{tr}\left(\rho_A^2\right)} \simeq 2^{-|A|}$. The right-hand side is actually the minimum possible value of $\mathrm{tr}\left(\rho_A^2\right)$, which is the purity of the maximally mixed state. Therefore if the average value is approaching this number, it suggests that almost all states have $\mathrm{tr}\left(\rho_A^2\right)$ close to this minimal value. In other words, most quantum states in this ensemble has almost maximal entropy.

To make this more precise, we consider a random variable $X(s)$ which has minimal value X_m and average value $X_a > X_m$. Then

$$\begin{aligned}X_a &= \int dsP(s)X(s)\\ \int dsP(s)\theta(X_a - X(s))\left|X(s) - X_a\right| &= \int dsP(s)\theta(X(s) - X_a)\left|X(s) - X_a\right|\end{aligned} \tag{3.150}$$

Since on the left-hand side $X_a \geq X(s) \geq X_m$, we have

$$\int dsP(s)\theta(X(s)-X_a)\left|X(s)-X_a\right| \leq \int dsP(s)\theta(X_a-X(s))(X_a-X_m) \leq X_a - X_m \tag{3.151}$$

Thus for any positive ϵ,

$$\int dsP(s)\theta(X(s)-X_a-\epsilon) \leq \frac{1}{\epsilon}\int dsP(s)\theta(X(s)-X_a)\left|X(s)-X_a\right| \leq \frac{X_a-X_m}{\epsilon} \tag{3.152}$$

This equation provides an upper bound of the probability that $X(s)$ is bigger than X_a by ϵ. Returning to the purity, $X_m = 2^{-|A|}$, and $X_a = \frac{2^{|A|}+2^{N-|A|}}{2^N+1}$.

$$X_a - X_m = \frac{2^{|A|}-2^{-|A|}}{2^N+1} \simeq 2^{|A|-N} \tag{3.153}$$

Thus if $|A| < N/2$, and we take $\epsilon = \Omega(2^{|A|-N})$ (ϵ decays slower than the right-hand side), we have

$$\text{Prob}\left(X(s) \geq X_a + \epsilon\right) \leq \epsilon^{-1}2^{|A|-N} \tag{3.154}$$

with the right-hand side vanishes in large N.

For $|A| = N - |A|$, when A is exactly half system, we obtain

$$\overline{\text{tr}\left(\rho_A^2\right)} = \frac{2^{|A|+1}}{2^N+1} \simeq 2^{|A|+1} \tag{3.155}$$

The entropy is $\log 2$ smaller than the maximal value.

This discussion can be generalized to higher Renyi entropy. For general n:

$$\text{tr}\left(X_{An}\pi_n\right) = \frac{1}{\Omega_n(D)}\sum_{g\in S_n} D_A^{l(rg)}D_B^{l(g)} \tag{3.156}$$

Here r is the cyclic permutation. If $D_A = 2^{|A|},\ D_B = 2^{N-|A|}$, we get

$$\text{tr}\left(X_{An}\pi_n\right) = \frac{\sum_{g\in S_n} 2^{|A|(l(rg)-l(g))}2^{Nl(g)}}{\sum_{g\in S_n} 2^{Nl(g)}} \tag{3.157}$$

If $|A| < N/2$, the sum is dominated by the identity term $l(g) = n$. For $|A| = N/2$, the numerator in the equation above is

$$2^{\frac{N}{2}(l(rg)+l(g))} \tag{3.158}$$

We can define $d(g,h) = n - l(g^{-1}h)$ as the distance between two group elements g, h, which is equal to the number of pair permutations needed to take h to g. Thus

$$l(rg)+l(g) = 2n - \left(d\left(g,r^{-1}\right)+d\left(g,e\right)\right) \tag{3.159}$$

The dominant term has minimal $d\left(g,r^{-1}\right)+d\left(g,e\right)$. Due to triangle inequality, in general we have $d\left(g,r^{-1}\right)+d\left(g,e\right) \geq d(r^{-1},e) = n-1$. This inequality takes the equal sign for

some number of g. The number is known to be the Catalan number. (See e.g.[17], also discussed in an appendix of [18].)

$$C_n = \frac{1}{n+1}\begin{pmatrix} 2n \\ n \end{pmatrix} \tag{3.160}$$

Thus we have

$$\mathrm{tr}\,(X_{An}\pi_n) \simeq C_n 2^{-\frac{N}{2}(n-1)} \tag{3.161}$$

which leads to

$$-\frac{1}{n-1}\log \mathrm{tr}\,(X_{An}\pi_n) \simeq -\frac{1}{n-1}\log C_n + \frac{N}{2} \tag{3.162}$$

Thus the entropy is a constant $\frac{1}{n-1}\log C_n$ below maximal value. If we take $n \to 1$ limit, we get

$$\begin{aligned} \Delta S &= \lim_{n\to 1}\frac{\log C_n}{n-1} \\ &= \left[\log\Gamma(2x+1) - 2\log(\Gamma(x+1)) - \log(x+1)\right]'|_{x=1} = \Gamma'(3) - 2\Gamma'(2) - \frac{1}{2} = \frac{1}{2} \end{aligned} \tag{3.163}$$

In summary we see that a random state is almost maximally entangled. For a system with size N and subsystem size xN, the entropy is approximately $S(x) = \min\{x, 1-x\} N \log 2$. The correction to this formula is of order $2^{-N|1-2x|}$, which is exponentially small, except at $x = 1/2$. At $x = 1/2$ the correction is a universal number $\frac{1}{2}$. This behavior of $S(x)$ vs x is usually called the Page curve in high energy physics community.[19]

This result suggests that the Page curve is very general, since it is the behavior of a typical state in the Hilbert space. If we consider a generic Hamiltonian H, and evolve a low entropy reference state $|0\rangle = |000...0\rangle$ by a long time, we obtain $|\Psi(t)\rangle = e^{-iHt}|0\rangle$. If there were no reason to believe $|\Psi(t)\rangle$ is special, we would have expected that it has the entropy of a typical state. However, that is clearly not true. For example consider a Hamiltonian that is a spin (qubit) model with two-body interactions:

$$H = \sum_{i,j,\alpha,\beta} J_{ij}^{\alpha\beta}\sigma_{i\alpha}\sigma_{j\beta} + \sum_{i\alpha} h_{i\alpha}\sigma_{i\alpha} \tag{3.164}$$

with $i, j = 1, 2, ..., N$ and $\alpha, \beta = x, y, z$. For a typical state, the energy expectation value is

$$\langle\Psi|H|\Psi\rangle = \sum_{i,j,\alpha,\beta} J_{ij}^{\alpha\beta}\langle\Psi|\sigma_{i\alpha}\sigma_{j\beta}|\Psi\rangle + \sum_{i\alpha} h_{i\alpha}\langle\Psi|\sigma_{i\alpha}|\Psi\rangle \tag{3.165}$$

The expectation value in each term only involves at most two sites, which therefore is determined by the two-site reduce density operator ρ_{ij}. For a typical state, we have $\rho_{ij} \simeq \frac{1}{4}\mathbb{I}_{ij}$, so that $\langle\Psi|H|\Psi\rangle = 2^{-N}\mathrm{tr}(H) = 0$. Since the time evolution preserves energy, the energy expectation value for $|\Psi(t)\rangle$ will be determined by that at time 0, which is generically nonzero. Thus the state $|\Psi(t)\rangle$ cannot be a typical state. However, it could be a "typical

state at that energy". If we know a state has energy average value E, we can try to compute its maximal possible entropy. Adding two Lagrangian multipliers for $\mathrm{tr}\,(\rho H) = E$ and $\mathrm{tr}\,(\rho) = 1$, we have

$$S(\rho) = -\mathrm{tr}\,(\rho \log \rho) - \beta\,(\mathrm{tr}\,(\rho H) - E) - \alpha\,(\mathrm{tr}\rho - 1) \tag{3.166}$$

$\delta S(\rho) = 0$ leads to

$$-\log\rho - \beta H - \alpha\mathbb{I} = 0 \Rightarrow \rho = e^{-\alpha-\beta H} \tag{3.167}$$

This derivation confirms that the canonical ensemble is the maximal entropy state at a given energy eigenvalue. This state generically has a volume law entropy $S_A \propto |A|$. The analog of the random state result at finite temperature is that a typical state with this energy has a reduced density operator

$$\mathrm{tr}_{\overline{A}}\left(|\Psi\rangle\langle\Psi|\right) \simeq \mathrm{tr}_{\overline{A}}\left(\frac{1}{Z}e^{-\beta H}\right),\ \text{for}\ |A| < \frac{N}{2} \tag{3.168}$$

We cannot prove this rigorously as in the infinite temperature case. There is a conjecture that not only typical states but the energy eigenstates look typical, which is known as the eigenstate thermalization hypothesis (ETH)[20–22]. The ETH contains the following two hypothesis for energy eigenstates $|n\rangle$:

1. For a local operator O, $\langle n|\,O\,|n\rangle$ for $|E_n - E| \leq O(\sqrt{N})$ has a small fluctuation.

$$O_{nn} = \langle n|\,O\,|n\rangle = \langle O\rangle_{\mathrm{micro}} + \Delta_i \tag{3.169}$$

with $\langle O\rangle_{\mathrm{micro}} = \frac{1}{N_E}\sum_{|E_n-E|\leq\alpha\sqrt{N}}\langle m|\,O\,|m\rangle$ the microcanonical ensemble average. Δ_i is random with zero mean and $\Delta_i^2 \propto e^{-S(E)}$. This is an analog of the statement that a random state has a reduced density matrix close to maximally mixed state (if the size of subsystem is smaller than half system size). If the operator O is supported on a region A, then

$$|O_{nn} - \langle O\rangle_{\mathrm{micro}}| \leq \parallel \rho_A - \rho_A^{\mathrm{micro}} \parallel_1 \leq S\left(\rho_A|\rho_A^{\mathrm{micro}}\right) \tag{3.170}$$

If ρ_A^{micro} is a thermal state, $\rho_A^{\mathrm{micro}} \simeq Z_A^{-1}e^{-\beta H_A} \equiv \rho_A^{th}$ (which by itself requires locality of Hamiltonian), then

$$S\left(\rho_A|\rho_A^{\mathrm{micro}}\right) \simeq \beta\left(F(\rho_A) - F(\rho_A^{th})\right) \tag{3.171}$$

with F the free energy. $\beta F(\rho_A) = \beta\mathrm{tr}\,(\rho_A H_A) - S(\rho_A)$. The assumption is that

$$S\left(\rho_A|\rho_A^{\mathrm{micro}}\right) \sim e^{-S(E)} \tag{3.172}$$

For large region (when $|A|$ is a finite portion of the system), the right-hand side should be probably increasing with A size, such as $e^{-S_{\overline{A}}}$.

2. The off-diagonal matrix element

$$O_{nm} = \langle n| O |m\rangle = \Delta_{nm} \tag{3.173}$$

is random, with zero mean and the average

$$\left\langle |O_{nm}|^2 \right\rangle_{\text{micro}} \propto f\left(\frac{E_n + E_m}{2}, E_n - E_m\right) e^{-S((E_n+E_m)/2)/2} \tag{3.174}$$

The function f decays quickly with $E_n - E_m$. The off-diagonal matrix elements are responsible for time-dependence of correlation function. For example the connected two-point function

$$\begin{aligned}\langle n| O(t)O(0) |n\rangle - O_{nn}^2 &= \sum_{m\neq n} O_{nm}O_{mn}e^{-it(E_m-E_n)} \\ &\sim \sum_m f\left(\frac{E_n + E_m}{2}, E_n - E_m\right) e^{-S((E_n+E_m)/2)/2} e^{-it(E_m-E_n)}\end{aligned} \tag{3.175}$$

f function guarantees that the long time decay of the two-point function.

We can also consider the thermalization of a pure state.

$$|\Psi(t)\rangle = e^{-iHt} |\Psi\rangle = \sum_n e^{-itE_n} \langle n|\ \Psi\rangle |n\rangle \tag{3.176}$$

The reduced density operator of a small subsystem A is

$$\rho_A(t) = \sum_{n,m} e^{-it(E_n-E_m)} \langle n|\ \Psi\rangle \langle \Psi|\ m\rangle \operatorname{tr}_{\overline{A}} |n\rangle \langle m| \tag{3.177}$$

For any operator O on A,

$$\langle O(t)\rangle = \sum_{n,m} e^{-it(E_n-E_m)} \langle n|\ \Psi\rangle \langle \Psi|\ m\rangle \langle m| O |n\rangle \tag{3.178}$$

By assumption the off-diagonal components O_{mn} are random, so that after the sum we are approximately left with the diagonal term

$$\langle O(t)\rangle \simeq \sum_n \langle n|\ \Psi\rangle \langle \Psi|\ n\rangle O_{nn} \tag{3.179}$$

From this analysis it is not so obvious why at short time, $\langle O(t)\rangle$ is time-dependent. This is because the matrix element of O_{mn} and $\langle n|\ \Psi\rangle \langle \Psi|\ m\rangle$ has correlation with each other (if $\langle \Psi| O |\Psi\rangle \neq 0$). After we add the time-dependent phase, this correlation decays, in the same way as in the two-point function.

It is clear that the discussion above is not rigorous. A lot of questions remain open in the thermalization problem. The ETH remains a hypothesis. One problem with proving ETH is what we have to state the conditions more precisely. It is obvious that not all Hamiltonians satisfy ETH. For example free fermion (Gaussian) Hamiltonians do not satisfy ETH. How to quantify the requirement of "non-integrability"? Or maybe the precise statement is that ETH holds for almost all Hamiltonians.

4 Quantum entanglement and spacetime

4.1 Holographic duality

Basic setup. RT formula. HRT. RT with quantum correction. RT and quantum error correction.

4.2 Random tensor networks

4.3 Sachdev-Ye-Kitaev model

Not sure if we have time to cover SYK.

4.4 Open questions

References

[1] E. Chitambar, D. Leung, L. Mančinska, M. Ozols, and A. Winter, *Everything you always wanted to know about locc (but were afraid to ask)*, *Communications in Mathematical Physics* **328** (2014), no. 1 303–326.

[2] G. Lindblad, *Completely positive maps and entropy inequalities*, *Communications in Mathematical Physics* **40** (1975), no. 2 147–151.

[3] E. H. Lieb and M. B. Ruskai, *Proof of the strong subadditivity of quantum-mechanical entropy*, *Les rencontres physiciens-mathématiciens de Strasbourg-RCP25* **19** (1973) 36–55.

[4] E. Witten, *A mini-introduction to information theory*, *La Rivista del Nuovo Cimento* **43** (2020), no. 4 187–227.

[5] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2010.

[6] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill, *Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence*, *Journal of High Energy Physics* **2015** (2015), no. 6 1–55.

[7] A. Y. Kitaev, *Fault-tolerant quantum computation by anyons*, *Annals of Physics* **303** (2003), no. 1 2–30.

[8] M. Levin and X.-G. Wen, *Detecting topological order in a ground state wave function*, *Physical review letters* **96** (2006), no. 11 110405.

[9] A. Kitaev and J. Preskill, *Topological entanglement entropy*, *Physical review letters* **96** (2006), no. 11 110404.

[10] I. Peschel, *Calculation of reduced density matrices from correlation functions*, *Journal of Physics A: Mathematical and General* **36** (2003), no. 14 L205.

[11] H. Casini and M. Huerta, *Reduced density matrix and internal dynamics for multicomponent regions*, *Classical and quantum gravity* **26** (2009), no. 18 185005.

[12] M. M. Wolf, *Violation of the entropic area law for fermions*, *Physical review letters* **96** (2006), no. 1 010404.

[13] D. Gioev and I. Klich, *Entanglement entropy of fermions in any dimension and the widom conjecture*, *Physical review letters* **96** (2006), no. 10 100503.

[14] B. Swingle, *Entanglement entropy and the fermi surface*, *Physical review letters* **105** (2010), no. 5 050502.

[15] C. Holzhey, F. Larsen, and F. Wilczek, *Geometric and renormalized entropy in conformal field theory*, *Nuclear physics b* **424** (1994), no. 3 443–467.

[16] P. Calabrese and J. Cardy, *Entanglement entropy and quantum field theory*, *Journal of statistical mechanics: theory and experiment* **2004** (2004), no. 06 P06002.

[17] A. Nica and R. Speicher, *Lectures on the combinatorics of free probability*, vol. 13. Cambridge University Press, 2006.

[18] X. Dong, X.-L. Qi, and M. Walter, *Holographic entanglement negativity and replica symmetry breaking*, *Journal of High Energy Physics* **2021** (2021), no. 6 1–41.

[19] D. N. Page, *Average entropy of a subsystem*, *Physical review letters* **71** (1993), no. 9 1291.

[20] J. M. Deutsch, *Quantum statistical mechanics in a closed system*, *Physical review a* **43** (1991), no. 4 2046.

[21] M. Srednicki, *Chaos and quantum thermalization*, *Physical review e* **50** (1994), no. 2 888.

[22] J. M. Deutsch, *Eigenstate thermalization hypothesis*, *Reports on Progress in Physics* **81** (2018), no. 8 082001.