Quantum information theory (20110401)

Lecturer: Jens Eisert

Chapter 5: Entanglement theory

Contents

Chapter 5

Entanglement theory

Entanglement is the key feature of quantum mechanics that renders it different from a classical statistical theory. Bell's theorem that shows that a classical statistical interpretation of quantum mechanics is not compatible with experimental findings resorts to notions of entanglement. Practically speaking, entanglement is the main resource in quantum information theory. Quantum key distribution requires entanglement, quantum computers cannot outperform classical machines without entanglement. In quantum error correction entangled states are of major importance. Sensing protocols can also only outperform classical ones when entanglement is present. Indeed, basically all advantages of protocols in quantum information theory can be traced back to entanglement being available in one way or the other. Hence, it makes a lot of sense to carefully consider notions of entanglement in quantitative terms. This is going to be a comparably long chapter – but again this makes a lot of sense as we are laying out important foundations here.

5.1 Pure state entanglement

5.1.1 Definition of pure state entanglement

Entanglement is a property of composite quantum systems. We need entities that can be separately and locally manipulated. We think of the separate laboratories paradigm, in which one part is held by one party, say, Alice, in a laboratory and the other by another party, say Bob. They can perform local operations with classical communication (LOCC), but cannot directly naturally implement joint quantum operations. The Hilbert space is hence given by

$$\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B. \tag{5.1}$$

Assessing pure state entanglement is rather simple. A pure quantum state reflected by a product state vector

$$|\psi\rangle = |\phi\rangle_A \otimes |\omega\rangle_B \tag{5.2}$$

is referred to as being *not entangled*, all other states are *entangled*. Entanglement is hence defined via its negation. This makes a lot of sense: Product states

$$\rho = \phi_A \otimes \omega_B \tag{5.3}$$

will lead to no correlations in measurement outcomes: All probability distributions resulting from measurements are product distributions. Whenever the state is no product states, quantum correlations will be present. We will assume the local Hilbert spaces to be finite-dimensional and have the same dimension so that $\dim(\mathcal{H}_A) = d$ and $\dim(\mathcal{H}_B) = d$.

Bi-partite pure entangled states: All pure quantum states on $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ that are not product states of the form $\rho = \phi_A \otimes \omega_B$ are called *entangled states*.

Quantum states that are not entangled are also called *separable*.

5.1.2 Entropy of entanglement quantifying pure state entanglement

So far so good. While this definition makes perfect sense, it is also agnostic to the question of how much entanglement is present in a system. One would expect states that are close to being products to contain only "little" entanglement, while maximally entangled states for which the reduction to one part is maximally mixed should have the maximum degree of entanglement. We have already encountered the Schmidt decomposition that should tell us all there is about the entanglement content of a state.

Indeed, there is a quantity that reasonably *quantifies* the degree of pure-state entanglement. This is the *entropy of entanglement*. In fact, we will see that is in a sense the unique measure of entanglement. Let us define it first and then see where it comes from.

Entropy of entanglement: The entanglement content of a bi-partite pure state ρ on $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ is given by its *entropy of entanglement*

$$E(\rho) = S(\mathrm{tr}_B(\rho)), \tag{5.4}$$

where $S : \mathcal{S}(\mathcal{H}) \to [0, \infty)$ is the von-Neumann entropy defined as

$$S(\sigma) = -\mathrm{tr}(\sigma \log_2(\sigma)). \tag{5.5}$$

The latter is a matrix function of σ [1]. As noted above, the resource character of entanglement is stressed by referring to the entanglement content of a maximally

[1] The von-Neumann entropy quantifies the mixedness of a quantum state. Defined as a function $S : \mathcal{S}(\mathbb{C}^d) \to [0, \infty)$, it is zero

$$S(\sigma) = 0 \tag{5.6}$$

entangled pair of qubits on $\mathbb{C}^2 \otimes \mathbb{C}^2$ as an *ebit*. So as a sanity check, let us compute the entropy of entanglement for a maximally entangled pure state on $\mathbb{C}^2 \otimes \mathbb{C}^2$ with

$$|\Omega\rangle = \frac{1}{\sqrt{2}}(|0,0\rangle + |1,1\rangle). \tag{5.9}$$

We find

$$S(\mathrm{tr}_B|\Omega\rangle\langle\Omega|) = S(\mathbb{1}/2) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1. \tag{5.10}$$

So indeed, we find the unit value 1 for the entanglement content of an ebit, as expected. As we have seen earlier, this is precisely the resource that we need for the teleportation of a single qubit, and hence the term ebit is more than justified. At the same time, for a separable quantum state ρ, we find

$$E(\rho) = 0, \tag{5.11}$$

simply because the von-Neumann entropy of pure quantum states is zero.

If the quantum state σ is diagonal with main diagonal elements $p = (p_0, \ldots, p_{d-1})$, so that indeed, these elements form a probability distribution, then

$$S(\sigma) = H(p). \tag{5.12}$$

This is nothing but the famous *Shannon entropy*, the key quantity of classical information theory that quantifies the information content of a source.

Shannon entropy: The Shannon entropy of a probability distribution $p = (p_0, \ldots, p_{d-1})$ is given by

$$H(p) = -\sum_{j=0}^{d-1} p_j \log_2(p_j). \tag{5.13}$$

Once we have diagonalized the reduced quantum state

$$\mathrm{tr}_B(\rho) = \sum_{j=0}^{d-1} p_j |j\rangle\langle j| \tag{5.14}$$

of ρ for a suitable basis $\mathcal{B} = \{|0\rangle, \ldots, |d-1\rangle\}$, the entropy of entanglement is hence nothing but the Shannon entropy of the eigenvalues of the reduced state. It goes

exactly if σ is pure, takes its maximal value $\log_2(d)$ for the maximally mixed state, is additive,

$$S(\sigma_1 \otimes \sigma_2) = S(\sigma_1) + S(\sigma_2) \tag{5.7}$$

for all quantum states σ_1 and σ_2 and sub-additive as

$$S(\sigma) \leq S(\mathrm{tr}_2(\sigma)) + S(\mathrm{tr}_1(\sigma)) \tag{5.8}$$

for all states σ.

without saying that we could also have looked at the reduction $\mathrm{tr}_A(\rho)$ of ρ, as the spectra of the two reductions are identical. Pure states have the spectrum

$$p = (1, 0, \ldots, 0), \tag{5.15}$$

so that we again find that separable states get the value 0 assigned.

Before we turn to the operational interpretation of the entropy of entanglement, let us make the following remarks: The entropy of entanglement meaningfully quantifies the degree of entanglement. The larger the degree of entanglement, the larger the entropy of entanglement. It is also invariant under local basis changes: If we consider for a quantum state ρ the new state

$$(U_A \otimes U_B)\rho(U_A \otimes U_B)^\dagger \tag{5.16}$$

for local unitaries U_A and U_B, then clearly

$$E(\rho) = E((U_A \otimes U_B)\rho(U_A \otimes U_B)^\dagger). \tag{5.17}$$

This immediately follows from the fact that the spectrum of the reduced quantum state will not depend on the basis in which the reduced state is expressed in. Physically, this also makes a lot of sense: The entanglement degree is invariant under local unitary operations.

5.1.3 Typical sequences

We already see that there is a link between notions of entanglement and those of classical information theory, let us more deeply look into that. Indeed, what follows is one of the most profound statements of the field. Let us first make a little detour, then put it all together and find the compelling physical interpretation of the findings. In what follows, we think hence in perfectly classical terms and think of *statistical sources* that generate samples of identically and independently distributed (i.i.d.) random numbers in which each realization is taking the value j with probability p_j for $j = 0, \ldots, d-1$, so that the ensemble is described by the probability distribution

$$p = (p_0, \ldots, p_{d-1}). \tag{5.18}$$

In this way, one obtains strings $(x_1, \ldots, x_n)$ of i.i.d. distributed random variables. That is to say, we think of a source that spits out realizations of random variables in a completely uncorrelated fashion. To keep things simple, we will put an emphasis onto the case of a bit string in which x_i takes the value 1 with probability p and the value 0 with probability $1 - p$. In all what follows, we will be interested in the limit of long strings reflected by the limit $n \to \infty$.

Typical sequences: A *typical sequence* is a string of a source the frequences of $0, \ldots, d-1$ are about as large as the probabilities, normalized by the length of the string. That is to say, a typical sequence $(x_1, \ldots, x_n)$ contains approximately $p_0 n$ many times the value 0, $p_1 n$ many times the value 1, and so on.

An *atypical sequence* would be a highly improbable string such as $(0, 0, 0, \ldots, 0)$. Let us more closely have a look at the case of a random bit string. A typical sequence

here contains about np times the value 1 and $n(1-p)$ times the value 0. This bit string corresponds to a Bernoulli chain. The *expectation value* is given by

$$\bar{x} = np, \tag{5.19}$$

the *variance* as

$$\Delta x = \sqrt{np(1-p)}. \tag{5.20}$$

The *law of large numbers* now tells us that the fluctuations around the mean vanish for large strings as

$$\frac{\Delta x}{\bar{x}} \sim \frac{1}{\sqrt{n}} \to 0 \tag{5.21}$$

as $n \to \infty$. Every typical sequence x will asymptotically have the same probability

$$p(x) \approx p_0^{np_0} p_1^{np_1} \dots p_{d-1}^{np_{d-1}}. \tag{5.22}$$

Taking the binary logarithm thereof, we find

$$\begin{aligned} \log_2(p_0^{np_0} p_1^{np_1} \dots p_{d-1}^{np_{d-1}}) &= n \sum_{j=0}^{d-1} p_j \log_2(p_j) \\ &= -nH(p). \end{aligned} \tag{5.23}$$

This is already an exciting insight: The probability of obtaining a typical sequence is almost uniform over the typical sequences. And we can conclude the following.

Approximately uniform distribution of typical sequences: In the limit $n \to \infty$ of long sequences, the typical sequences are approximately uniformly distributed, and their probability is well approximated by the Shannon entropy of the source

$$p(x) \approx 2^{-nH(p)}. \tag{5.24}$$

There are about $2^{nH(p)}$ many typical strings.

5.1.4 Central limit theorem

This notion can be made more precise (even if we are not fully rigorous here, but it should be clear that all this can be turned into a formal proof). One can define for an $\epsilon > 0$ the ϵ-typical strings that are close to typical sequences in the following sense. We call a string x ϵ-typical if

$$2^{-n(H(p)+\epsilon)} \leq p(x) \leq 2^{-n(H(p)-\epsilon)}. \tag{5.25}$$

This implies, taking the logarithm, that

$$-n(H(p)+\epsilon) \leq \log_2 p(x) \leq -n(H(p)-\epsilon). \tag{5.26}$$

That is, the set of ϵ-typical strings is given by

$$T(\epsilon, n) := \left\{ x \in \{0, \ldots, d-1\}^n : \left| \frac{1}{n} \log_2 p(x) + nH(p) \right| \leq \epsilon \right\}. \tag{5.27}$$

For such strings, we can formulate the central limit theorem.

Central limit theorem:

$$\forall \varepsilon > 0 \forall \delta > 0 \exists n : p(x \in T(\epsilon, n)) = \sum_{x \in T(\epsilon,n)} p(x) > 1 - \delta, \tag{5.28}$$

$$\forall \varepsilon > 0 \forall \delta > 0 \exists n : (1-\delta) 2^{n(H(p)-\epsilon)} \leq |T(\epsilon, n)| \leq 2^{n(H(p)+\epsilon)}. \tag{5.29}$$

This is nothing but a slightly more formal way of putting what we have said before. More specifically, Eq. (5.28) is the central limit theorem. The second can easily be seen as following from the definition of an ϵ-typical sequence and the first line above, as

$$\begin{aligned} 1 &\geq \sum_{x \in T(\epsilon,n)} p(x) \geq \sum_{x \in T(\epsilon,n)} 2^{-n(H(p)+\epsilon)} && (5.30) \\ &= |T(\epsilon, n)| 2^{-n(H(p)+\epsilon)}, \\ 1 - \delta &\leq \sum_{x \in T(\epsilon,n)} p(x) \leq \sum_{x \in T(\epsilon,n)} 2^{-n(H(p)-\epsilon)} && (5.31) \\ &= |T(\epsilon, n)| 2^{-n(H(p)-\epsilon)}, \end{aligned}$$

as we can see from the above considerations.

5.1.5 Putting it all together: Pure state entanglement manipulation

What does this have to do with pure state entanglement manipulation with LOCC? It turns out that in fact, a lot, and we are almost there. We will see that these considerations show how initial pure states associated with state vectors $|\psi\rangle \in \mathcal{H} = \mathbb{C}^d \otimes \mathbb{C}^d$ can be transformed into $|\phi\rangle \in \mathcal{H}$ by means of LOCC. Here we do not mean that the two parties Alice and Bob share a single specimen of this state – this situation we will comment upon later. Instead, the common situation is that Alice and Bob share many specimen, many "copies" of the same state and ask at what rate a transformation is possible. This situation is referred to as the *asymptotic state manipulation*, in contrast to so-called single-shot settings.

The initial situation hence is the one where the parties initially share n copies of the same state, so that

$$|\psi\rangle \otimes \cdots \otimes |\psi\rangle = |\psi\rangle^{\otimes n} \tag{5.32}$$

with

$$|\psi\rangle = \sum_{j=0}^{d-1} \sqrt{p_j} |j, j\rangle. \tag{5.33}$$

Here, we have chosen the basis of the Schmidt decomposition, which we are perfectly free to do, as all quantities we are interested in are invariant under local unitary operations. Indeed, the Schmidt decomposition constitutes a probability distribution, and all we have learnt so far can be applied here. As a first sanity check, we will compute the entropy of entanglement of these n copies. Since the von-Neumann entropy is additive, we find

$$E(|\psi\rangle\langle\psi|^{\otimes n}) = nE(|\psi\rangle\langle\psi|). \tag{5.34}$$

The entanglement content is naturally hence just n times the original degree of entanglement. We also know the spectrum of the reduced state $\mathrm{tr}_B(|\psi\rangle\langle\psi|^{\otimes n})$. It is given by $\{p(x) : x \in \{0, \ldots, d-1\}^n\}$. Taking this idea seriously, we know how to connect this to the central limit theorem and how to proceed.

5.1.6 Pure state distillable entanglement

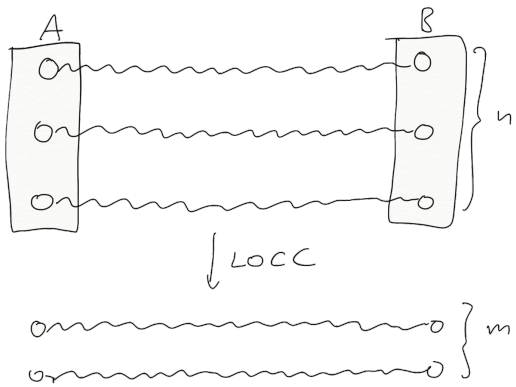
Let us first summarize what is on the desk before we approach the question at hand.

Distillable entanglement (informal): The distillable entanglement $E_D(\rho)$ of a quantum state ρ is the maximum rate at which one can asymptotically distill maximally entangled qubit pairs from many copies of ρ.

In our setting at hand, we ask how many copies m of maximally entangled $|\Omega\rangle$ we can extract from n copies of $|\psi\rangle$. This transformation does not have to be perfect, but only approximately perfect – we will make this more precise later. For such a transformation, we write

$$|\psi\rangle^{\otimes n} \to |\Omega\rangle^{\otimes m} \quad \text{under LOCC}. \tag{5.35}$$

Note also that this is supposed to happen strictly within the separated laboratories paradigm. Alice and Bob share these states. Within each laboratory, Alice and Bob can perform arbitrary transformations, however, even over all n tensor factors. This may be a violent abstraction from an experimental perspective, as it is not easy to perform such asymptotic transformations in practice. It is not even that easy to let two copies interact with little error. If we are interested in the best possible achievable rate, however, we have little choice, and have to resort to this situation. So what is the best achievable rate, what is the distillable entanglement?



For this, we start from n copies of $|\psi\rangle$ in Schmidt form as in Eq. (5.33). We can write

$$\begin{aligned} |\psi\rangle^{\otimes n} &= \sum_{x_1=0}^{d-1} \cdots \sum_{x_n=0}^{d-1} \sqrt{p_{x_1} \cdots p_{x_n}} |x_1, \ldots, x_n\rangle \otimes |x_1, \ldots, x_n\rangle \qquad (5.36) \\ &= \sum_x \sqrt{p(x)} |x\rangle \otimes |x\rangle. \end{aligned}$$

We see the distribution hence in the Schmidt spectrum of the n copies. We will now consider for an $\epsilon, \delta > 0$ the ϵ-typical strings $x \in T(\epsilon, n)$ in this sum. We know that most strings are typical, but not quite all, and hence we have to renormalize the state vector if we consider those typical strings only.

$$|\eta\rangle := \frac{1}{\sqrt{\langle \xi | \xi \rangle}} |\xi\rangle \qquad (5.37)$$

with

$$|\eta\rangle := \sum_{x \in T(\epsilon, n)} \sqrt{p(x)} |x\rangle \otimes |x\rangle. \qquad (5.38)$$

By construction, we have

$$\langle \psi^{\otimes n} | \xi \rangle = \langle \xi | \xi \rangle, \qquad (5.39)$$

since $|\xi\rangle$ is just a part of $|\psi\rangle$. What is more, it is almost the same state vector, as we can easily verify making use of our above findings. We find

$$\begin{aligned} |\langle \psi^{\otimes n} | \eta \rangle|^2 &= \frac{|\langle \psi^{\otimes n} | \xi \rangle|^2}{\langle \xi | \xi \rangle} = \langle \xi | \xi \rangle \qquad (5.40) \\ &= \sum_{x \in T(\epsilon, n)} p(x) \\ &= p(x \in T(\epsilon, n)) \\ &> 1 - \delta \qquad (5.41) \end{aligned}$$

for a sufficiently large n. This means that we can ensure that

$$|\langle\psi^{\otimes n}|\eta\rangle|^2 \to 1 \tag{5.42}$$

for $n \to \infty$. The number of strings in $|\xi\rangle$ is given by

$$|T(\epsilon, n)| \to 2^{nH(p)} = d^{nE(|\psi\rangle\langle\psi|)} \tag{5.43}$$

as $n \to \infty$. The Shannon entropy $H(p)$ of the distribution is precisely the von-Neumann entropy of the reduction of the pure states: Again, the Shannon entropy is precisely the entanglement entropy to the basis 2. All sequences are different and can be locally distinguished with unit probability by means of local measurements.

Asymptotic equivalence: We find that in the asymptotic limit $n \to \infty$ the atypical sequences do not play a role, and hence the effective dimension of the n-copy Hilbert space is $d^{nE(|\psi\rangle\langle\psi|)}$. The states $|\psi\rangle\langle\psi|^{\otimes n}$ and $|\eta\rangle\langle\eta|$ are asymptotically identical.

What does this now mean? So far the arguments have been largely abstract. We now fill this argument with physical meat. In fact, we are largely done and merely need to interpret our findings. The state vector $|\eta\rangle$ that only contains the typical strings of the initial state vector $|\psi\rangle^{\otimes n}$ has in addition the property that all of its Schmidt coefficients are largely identical for large n, i.e.,

$$d^{-n(E(|\psi\rangle\langle\psi|)+\epsilon)} \leq p(x) \leq d^{-n(E(|\psi\rangle\langle\psi|)-\epsilon)} \tag{5.44}$$

for all $x \in T(\epsilon, n)$. In the limit of $n \to \infty$ and suitable $\epsilon \to 0$, we find the following observation. In the asymptotic limit, $|\psi\rangle\langle\psi|^{\otimes n}$ is the same as $\lceil nE(|\psi\rangle\langle\psi|)\rceil$ copies of a maximally entangled state up to relabelings of the basis, i.e., there exist local unitaries U_A and U_B such that

$$(U_A \otimes U_B)|\eta\rangle \approx |\Omega\rangle^{\otimes\lceil nE(|\psi\rangle\langle\psi|)\rceil}, \tag{5.45}$$

to an arbitrarily small error for large n. Here, we have used the integer brackets for mathematical aesthetic considerations, as it will not matter whether we take integer values or not. This means, needless to say, the following.

Distillable entanglement of pure states: The distillable entanglement $E_D(|\psi\rangle\langle\psi|)$ of pure states $|\psi\rangle\langle\psi|$ is given by the entanglement entropy

$$E_D(|\psi\rangle\langle\psi|) = E(|\psi\rangle\langle\psi|). \tag{5.46}$$

We can even give a detailed protocol. For realizing a distillation protocol, one simply has to extract from $|\psi\rangle^{\otimes n}$ the state vector $|\eta\rangle$. This can be obtained by a suitable local measurement. Instead of performing measurements on each of the tensor factors, one performs a global measurement on all of the n copies at the same time, and

checks the relative frequencies of $0, \ldots, d-1$ in $|\eta\rangle$. The entanglement within the respective subspaces is preserved by this. The protocols would look as such. Initially, Alice and Bob share n copies of

$$|\psi\rangle = \sum_{j=0}^{d-1} \sqrt{p_j}|j,j\rangle. \tag{5.47}$$

- Then Alice measures the relative frequencies of $0, \ldots d-1$ in the with $p(x)$ weighted superposition $|\psi\rangle^{\otimes n}$ of all sequences.
- She communicates the measurement outcomes $n_0, \ldots, n_{d-1}$ to Bob. The output is kept if the obtained frequencies are close to $np_0, \ldots, np_{d-1}$. In the asymptotic limit $n \to \infty$ this happens in the overwhelming majority of times. The state obtained is then the equal superposition of all typical sequences with the same Schmidt coefficients. Otherwise, the state is discarded and the protocol aborted (but this is unlikely to happen).
- Since $|\eta\rangle$ is already maximally entangled, Alice and Bob merely need to perform local basis rotations using U_A and U_B to get $\lceil nE(|\psi\rangle\langle\psi|)\rceil$ maximally entangled qubit pairs in the state $|\Omega\rangle\langle\Omega|$.

This is a very exciting result: The von-Neumann entropy has now received an operational meaning. It is the maximum number of maximally entangled qubit pairs one can extract from the input state. Practical protocols will achieve fewer pairs, but this is not so much the point. In principle, the asymptotic entanglement content is that of as many pairs as the entanglement entropy as an optimally achievable rate dictates. But it gets better than that.

5.1.7 Pure state entanglement dilution

The task of entanglement dilution is the converse task. Here, one starts off with maximally entangled states and prepares a given state from these with maximally entangled states under LOCC. This task is referred to as *entanglement dilution* or *entanglement formation*. The optimal rate that is achievable in this task is called the *entanglement cost*. This nomenclature makes a lot of sense, since this is the entanglement one needs to "pay" to get a given state. This may sound less practically relevant than that of entanglement distillation (and it is, to be entirely honest), but we will end up at a remarkable insight.

Entanglement cost (informal): The entanglement cost $E_C(\rho)$ of a quantum state ρ is the maximum rate at which one can prepare the state asymptotically from many copies of maximally entangled qubit pairs.

Nicely prepared as we are now, we can swiftly proceed to the main insight at hand. In fact, we merely need to interpret what we already know. If $|\eta\rangle$ only contains $\lceil nE(|\psi\rangle\langle\psi|)\rceil$ copies of maximally entangled qubit pairs in the state $|\Omega\langle\langle\Omega|$, we can

make use of the following simple LOCC protocol. Again, let us assume that initially, Alice and Bob share $\lceil nE(|\psi\rangle\langle\psi|)\rceil$ maximally entangled qubit pairs.

- Alice prepares $|\psi\rangle^{\otimes n}$ on her side. Since this is a local operation, this is perfectly possible.
- Alice sends per quantum teleportation one half of the tensor factors to Bob. This is possible, since she effectively only needs to send $\lceil nE(|\psi\rangle\langle\psi|)\rceil$ many qubits to Bob, and for this purpose, needs $\lceil nE(|\psi\rangle\langle\psi|)\rceil$ maximally entangled resource states.
- At the end of the protocol, Alice and Bob share n copies of $|\psi\rangle\langle\psi|$.

This protocols shows how quantum teleportation can be used as a primitive in proofs and arguments, as promised. But there is a deeper insight that follows.

Entanglement cost of pure states: The entanglement cost $E_C(|\psi\rangle\langle\psi|)$ of pure states $|\psi\rangle\langle\psi|$ is given by the entanglement entropy

$$E_C(|\psi\rangle\langle\psi|) = E(|\psi\rangle\langle\psi|). \tag{5.48}$$

5.1.8 Asymptotic reversibility

This is precisely the same value as before! This is a remarkable insight. We can extract maximally entangled states at a rate dictated by the entanglement entropy. But using those maximally entangled states we have obtained, we can again at a rate given by the entanglement entropy create new entangled states.

Asymptotic reversibility of pure state entanglement manipulation: Bi-partite pure states can be asymptotically reversibly manipulated with LOCC. The entanglement entropy uniquely quantifies the degree of entanglement of bi-partite pure-state entanglement.

For the actual protocols, we have to measure, communicate results and all that. But the losses we encounter in this fashion are asymptotically negligible. This means, in other words, that we can in principle freely move from one state to another and again to another, without any losses. This also means that the entanglement entropy is a meaningful unique measure of entanglement. The entanglement entropy is in this sense the unique meaningful measure of entanglement of bi-partite pure-state entanglement [2]. Note also that what we have seen refers to asymptotic entanglement manipulation. But in the pure state setting, the single-shot situation is also fully understood [3].

[2] A similar statement does not hold for multi-partite entanglement. In fact, as recently as this year, our group has published further progress on the old question of what the best achievable rates of entanglement manipulation are.

[3] A moment of thought reveals that the optimal transition probability in the single-shot setting is still

5.2 Mixed state entanglement

We have seen that the situation of pure-state entanglement is conceptually remarkable simple, at least in the bi-partite setting. There is a "single number to rule them all", and the entanglement entropy is all we need to know. This is a very clean and nice state of affairs. It turns out that mixed state entanglement is significantly more intricate.

5.2.1 Definition of mixed-state entanglement

Even the definition is intricate. We have seen that product states are not entangled. But should we call states on a bi-partite Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ of the form

$$\rho = \sum_j p_j (\eta_j \otimes \xi_j) \tag{5.49}$$

entangled? Surely not. After all, the state is correlated, yes. But all correlations present we could have inserted by means of simple classical phone calls. On the quantum level, it is a mixture of product states. Such a state is called classically correlated or separable. Let us call the set of separable states

$$\mathcal{P}(\mathcal{H}) \subset \mathcal{S}(\mathcal{H}), \tag{5.50}$$

as a subset of state space constituted by all quantum states.

Mixed state entanglement: Bi-partite quantum states on $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ are called *entangled* if they are not *separable*, i.e., cannot be written in the form

$$\rho = \sum_j p_j (\eta_j \otimes \xi_j), \tag{5.51}$$

where η_j is a state on $\mathcal{H}_A$ and ξ_j acts on $\mathcal{H}_B$, and $\{p_j\}$ is a probability distribution.

Here the emphasis is on "cannot be written", as there are of course many decompositions of given pure states into other quantum states as a convex combinations. So again, we define entanglement as the negation of being not entangled, i.e., separable. But we now no longer have the means to easily check whether a given state is entangled or not. In fact, this turns out to be no detail. The separability problem, precisely formulated as a decision problem asking whether a given quantum state ρ is to a specified accuracy entangled or separable turns out to be an NP-hard problem: There is no polynomial time algorithm known. We might come back to this subtle question later. That is to say, even for a classical supercomputer, the problem of deciding separability is an intractable computational problem [4].

governed by the Schmidt coefficients. But now now longer the single quantity of the entanglement entropy defines the optimal probability of transition, but a notion called majorization. A later version of this script will go into more detail here.

[4] Here more fine print on this. How can one determine whether a mixed quantum state is entangled or

5.2.2 Entanglement criteria

This state of affairs motivates the introduction of entanglement criteria that as one-sided tests would state whether a given state is separable or entangled. We mention two concepts of this kind. This has been an active question of research for a long time, but is less studied now, for reasons that I will explain in the lecture. Still, two concepts are so important and simple to grasp that we discuss them here. The first one is really cute and we have basically seen this already. Let us assume for a moment that a state is separable and takes the form

$$\rho = \sum_j p_j (\eta_j \otimes \xi_j). \tag{5.57}$$

Will

$$\rho^\Gamma = \sum_j p_j (\eta_j \otimes \xi_j^T) \tag{5.58}$$

be a positive operator, i.e., will

$$\rho^\Gamma \geq 0 \tag{5.59}$$

hold true? The symbol ρ^Γ refers here as a *partial transpose*, so "half" a T symbol referring to a transpose. Surely yes, since the transposition is not completely positive, but positive. Therefore, for any separable state ρ, Eq. (5.59) will hold true. It is a quantum state with a *positive partial transpose* (PPT), as one says. Since we have already seen

not? Slightly more formally, the problem at hand is the following membership problem. To make this a little bit more precise, let us define the following two sets: Let us first write $\mathcal{S} := \mathcal{S}(\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B})$ and $\mathcal{S}_{\text{Sep}} := \mathcal{S}_{\text{Sep}}(\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B})$ for brevity. We define

$$S(\mathcal{S}_{\text{Sep}}, \delta) := \left\{\rho \in \mathcal{S} : \exists \sigma \in \mathcal{S}_{\text{Sep}} \text{ with } \|\rho - \sigma\|_2 < \delta\right\} \tag{5.52}$$

as the "deeply separable states" up to an accuracy $\delta > 0$ and

$$S(\mathcal{S}_{\text{Sep}}, -\delta) := \left\{\rho \in \mathcal{S}_{\text{Sep}} : S(\rho, \delta) \subset \mathcal{S}_{\text{Sep}}\right\}. \tag{5.53}$$

Then the separability problem becomes what is called a weak membership problem:

Separability problem: Let $\rho \in \mathcal{S}(\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}) = \mathcal{S}$ and $\delta > 0$ be a a precision parameter.

$$\text{If } \rho \in S(\mathcal{S}_{\text{Sep}}, -\delta) \quad \text{output YES,} \tag{5.54}$$
$$\text{if } \rho \notin S(\mathcal{S}_{\text{Sep}}, \delta) \quad \text{output NO.} \tag{5.55}$$

Here $\|.\|_2$ is the *Frobenius norm*, defined as

$$\|A\|_2 = \text{tr}(A^2) = (A, A) \tag{5.56}$$

for matrices A. However, from the perspective of computational complexity, this turns out to be a computationally hard problem.

Hardness of the separability problem: The separability problem is NP-hard, if δ scales inversely exponentially with respect to the dimensions d_A, d_B.

The proof makes use of a polynomial-time reduction to Edmonds' problem. The original proof shows that this problem is `NP-hard` if ρ is located within an inverse exponential (with respect to dimension) distance from the border of the set of separable quantum states.

that there are states for which the positive partial transpose is not positive, we have actually found an entanglement criterion.

PPT criterion: Every separable bi-partite quantum state has a positive partial transpose, i.e.,

$$\rho^{\Gamma} \geq 0. \tag{5.60}$$

This criterion is easy to check and has at the same time a nice operational interpretation: Partial time reversal detects entanglement. It is the most important entanglement criterion to date. In fact, the idea that the transposition is not completely positive can be generalized to the insight that a state is separable exactly if it is positive semi-definite under the partial application of a general positive map. So the (hard) question of deciding separability versus entanglement is then found to be equivalent to the (hard) problem of classifying all positive maps. Only the following is true.

Entanglement of qubit pairs: A quantum state on $\mathbb{C}^2 \otimes \mathbb{C}^2$ is separable and hence not entangled exactly if it has a positive partial transpose.

The same holds true for systems of a qubit with $d = 2$ and a qutrit with $d = 3$. For higher-dimensional quantum systems, one finds that this criterion is no longer necessary and sufficient for a state being separable. There is an entire zoo of further entanglement criteria we do not go into.

5.2.3 Entanglement witnesses

The second criterion is a criterion of the converse sort: It detects entanglement, not separability. There is an approach to this issue from the perspective of convex analysis, and an approach that is experimentally motivated. Let us first take the convex analysis perspective. A moment of thought reveals that not only state space $\mathcal{S}(\mathcal{H})$, the set of all quantum states is a convex set. The same applies to separable states. Once we have two separable states, mixing them will yield us another separable state. This is actually interesting: Since we already know that the separable states constitute a subset of state space, pictorially speaking, we have the situation that is reminiscent of an egg: There is the bigger egg white surrounding the egg yolk, the yellow stuff. The latter is also convex. So $\mathcal{P}(\mathcal{H})$ is a convex strict subset of $\mathcal{S}(\mathcal{H})$. As such, there is what is called a *separating hyperplane* defined via the Hilbert Schmidt scalar product. This a fanciful way of saying that there exists a $W = W^{\dagger}$ with the property that

$$\mathrm{tr}(W\rho) \geq 0 \tag{5.61}$$

for all separable quantum states ρ. This is a simple geometric insight. But of course, there is a simple interpretation of that W: It is an observable, and $\mathrm{tr}(W\rho)$ is an expectation value. That is to say, one simply needs to measure one observable to detect

entanglement. Now, of course, W is still supported on both tensor factors, so it may require to measure multiple local observables as

$$W = \sum_{j,k} O_j \otimes K_k \tag{5.62}$$

with $\{O_k\}$ acting in $\mathcal{H}_A$ and $\{K_k\}$ acting in $\mathcal{H}_B$, so that one can actually implement W with local operations. This still operationally requires less effort than determining the full quantum state. This simple idea is practically highly important and gives rise to what is called an *entanglement witness*.

Entanglement witness: An entanglement witness in a bi-partite quantum system with Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ is an observable $W = W^\dagger$ with

$$\mathrm{tr}(W\rho) \geq 0 \tag{5.63}$$

for all separable quantum states ρ and for which there exists an entangled quantum state σ for which

$$\mathrm{tr}(W\sigma) < 0. \tag{5.64}$$

The latter quantum state σ is then detected by this entanglement witness, as one says. It is always optimal to choose a tangent hyperplane to the set of separable states. The entanglement witness is then referred to as being optimal. Interestingly, to see how much one needs to shift a given entanglement witness to get an optimal entanglement witness is again an NP-hard problem. Also, it should be clear that not all entanglement witnesses are equally suitable to detect the entanglement content of a given state. But this is not so much the point: Often, one has a-priori knowledge about a quantum state, and only wishes to detect the entanglement in the state. And this can be nicely done employing the concept of an entanglement witness.

5.2.4 Distillable and bound entanglement for mixed states

The idea of entanglement distillation still makes a lot of sense, and the above informal definition still works. In fact, it practically makes a lot of sense, and we will see that one needs to perform entanglement distillation to achieve quantum communication over quantum communication channels. The idea of entanglement being a resource is most manifest here: The *distillable entanglement* $E_D(\rho)$ is the entanglement content one can practically extract from a given quantum state ρ with LOCC, as an optimal rate of maximally entangled states that can be extracted from a collection of copies of a given quantum state. Actually, in order not to be overwhelming in this chapter, we will delegate the rigorous definition of the distillable entanglement to the next chapter. The *entanglement cost* E_C again makes sense as the rate of states that can be prepared from maximally entangled qubit pairs by means of LOCC. Interestingly, the following statement is true.

Asymptotic irreversibility of mixed state entanglement manipulation: In general, for bi-partite quantum states ρ on $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ we have that

$$E_D(\rho) \leq E_C(\rho). \tag{5.65}$$

In practice, one finds that $E_D(\rho) < E_C(\rho)$. This is not that easy to show, however, as there are not so powerful bounds known for the distillable entanglement. It is not even easy to obtain lower bounds for the distillable entanglement. One such lower bound is the *hashing bound*

$$E_D(\rho) \geq -S(\rho) + S(\mathrm{tr}_B(\rho)). \tag{5.66}$$

The right hand side can be negative, however, which renders the bound pointless. There is a most intriguing phenomenon, however.

Bound entanglement: There exist entangled quantum states ρ on $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ for which

$$E_D(\rho) = 0 \tag{5.67}$$

but $E_C(\rho) > 0$.

That is to say, there is *bound entanglement* in nature: There are entangled states that cost entanglement in their preparation, and consequently, their entanglement cost $E_C(\rho)$ is larger than zero. Still, one can get no distillable entanglement out whatsoever with a positive rate, so that $E_D(\rho) = 0$: The entanglement is "bound". PPT states are in fact states for which the distillable entanglement vanishes. The reminiscence of notions of thermodynamics is no accident. It is a bit reminiscent of the second law of thermodynamics, which tells us that there is a form of energy called heat that cannot be used to extract work. Here, bound entanglement takes the role of "heat" in the context of entanglement theory. There is a lot more to be said on entanglement theory, and it may be worth noting that symmetries help a lot when quantifying entanglement. Also, notions of *entanglement monotones*

$$E : \mathcal{S}(\mathcal{H}) \to [0, \infty) \tag{5.68}$$

still take the role of the entanglement entropy for mixed quantum states and meaningfully quantify the degree of entanglement of mixed quantum states, even though they are now much harder to come by. For now, we leave it at that and move to the study of quantum channels and notions of quantum communication.