Topics in Quantum Entropy and Entanglement

Elliott Lieb
Princeton University

Princeton Cond Mat Summer School
Three lectures July 30,31 2014

What is entanglement?

The notion of entanglement dates back to 1935, and was introduced under that name by Erwin Schrödinger in two papers, one published in English in the Proceedings of the Cambridge Philosphical Society and another published in German in Naturwissenschaften. The latter is the famous paper on the 'cat paradox', and both were written in response to the line of debate opened in the paper of Einstein, Podolsky and Rosen. Schrödinger says that entanglement "is not one, but the characteristic trait of quantum mechanics, the one that enforces its entire line of departure from classical lines of thought."

This 'trait' is that two physical systems that are brought into temporary interaction, and then completely separated, can no longer be described, even after the interaction has utterly ceased, by individual wave functions for each of the two systems. Schrödinger wrote that "by their interaction, the two representatives (or functions) have become entangled." The quantitative theory of entanglement has begun to be developed surprisingly recently, largely in the contexts of quantum information, communication and cryptography.

Entanglement is now a very big topic (measured by journal pages and decibels, at least). We shall say something about it and about von Neumann entropy of states in these lectures, and why these topics are related.

First, we shall discuss entropy and its various properties. It is important to understand these, their proofs, and to what extent they are the quantum version of classical entropy notions.

$\boxed{Remark:}$ Entanglement arises from the fact that QM deals with several degrees of freedom in terms of a tensor product of spaces, *not* a sum of spaces (which would be the analog of the classical Cartesian product). More on this later. Put another way, it is a miracle that the early founders of QM, such as Schrödinger, instinctively headed for the tensor product (which resulted in a 'linear' theory for many-body systems, the Schrödinger equation) instead of having a wave function for each particle as in Hartree-Fock theory, which is a very non-linear theory (but much beloved by physicists and chemists).

Some Basic Facts About States (quickly)

(Apologies to everyone who knows this stuff, but this is supposed to be a summer school.)

In QM each 'degree of freedom' is associated with a (separable) Hilbert space of finite or infinite dimension. A d-o-f can be the the state of some subsystem of a larger system, or it can be the momenta of particles in a domain. It can be the 'angles' of a spin, or set of spins. The Hilbert space associated to several degrees of freedom is the tensor product, $\otimes$, of the spaces. For simplicity here we take all Hilbert spaces to be finite dimensional.

$\mathcal{H}_{12\dots} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots$.

The vectors in $\mathcal{H}_{12\dots}$ are linear combinations of multiplets of vectors $v_i, w_j, \cdots_k$ where $v_i, w_j, \cdots_k$ are orthonormal bases for $\mathcal{H}_1,\ \mathcal{H}_2, \cdots_k$. These form an O.N. basis for $\mathcal{H}_{12\dots}$. Note the simple, but important fact that $\mathcal{H}_{12\dots} = \mathcal{H}_1 \otimes (\mathcal{H}_{2\dots})$

The observables are (bounded) linear operators on $\mathcal{H}_{12\dots}$. A state on the observables is a *positive linear functional* (call it ω) whose value on the identity operator is 1. In the finite dimensional case states are given by

$\omega(A) = \mathrm{Tr}\,\rho\,A$ 'positive' *means that* $\omega(A^*A) \geq 0$ *for any* A

where A is an observable and ρ is a positive semi-definite operator with $\mathrm{Tr}\rho = 1$, called a density matrix. Here Tr is the trace. In case there are infinitely many degrees of freedom (as in a spin system in the thermodynamic limit), or when the Hilbert spaces are infinite dimensional, we have to be a bit more cautious in our definitions.

Some references

An excellent reference for matrices, traces, and other things we shall talk about is:

Eric Carlen, *Trace Inequalities and Quantum Entropy: An Introductory Course*, in 'Entropy and the quantum', 73–140, Contemp. Math., **529**, Amer. Math. Soc. (2010).

Also available on www.mathphys.org/AZschool/material/AZ09-carlen.pdf

Another useful (short) reference is: *Matrix and Operator Trace Inequalities* on both Wikipedia and Scholarpedia

Partial Trace

An important concept is partial trace. Given an operator (observable or a density matrix) A_{12} on $\mathcal{H}_{12}$ there is a unique operator (observable or density matrix) on $\mathcal{H}_1$, called

$$A_1 := \mathrm{Tr}_2 A_{12}$$

with the property that for all observables B_1 on $\mathcal{H}_1$

$$\mathrm{Tr}_{\mathcal{H}_{12}} A_{12}(B_1 \otimes \mathbb{1}_2) = \mathrm{Tr}_{\mathcal{H}_1} A_1 B_1$$

The matrix elements of this operator can be written explicitly as

$$\langle v|A_1|v\rangle = \sum_j \langle v \otimes w_j|A_{12}|v \otimes w_j\rangle$$

where the w_j are an O.N. basis for $\mathcal{H}_2$ (the result is basis independent). The symbol $\otimes$ has been extended here to operators. Previously it was defined for vectors. The extension is obvious:

$(A_1 \otimes B_2)(v_1 \otimes w_2) = A_1 v_1 \otimes B_2 w_2$

Note that $\mathrm{Tr}_{\mathcal{H}_{12}} = \mathrm{Tr}_{\mathcal{H}_1}(\mathrm{Tr}_{\mathcal{H}_2})$, so that $\mathrm{Tr}_1 A_1 = \mathrm{Tr}_{12} A_{12}$.

Schmidt Decomposition

A simple example of a (bipartite) state is a rank-one density matrix on $\mathcal{H}_{12}$ given by $\rho_{12} = |\Psi\rangle\langle\Psi|$ with $|\Psi\rangle \in \mathcal{H}_{12}$. This gives us a ρ_1 on $\mathcal{H}_1$ with $\rho_1 = \mathrm{Tr}_2\rho_{12}$, and $\rho_2 = \mathrm{Tr}_1\rho_{12}$.

Choose orthonormal bases for the two spaces and consider the coefficients of $|\Psi\rangle = \psi_{i,j}$ in these bases. Thus, $|\Psi\rangle$ looks like a matrix, which we shall call R. Then the matrix ρ_2 in this basis is simply $R^\dagger R$, while $\rho_1 = R\,R^\dagger$. By a well known theorem of linear algebra, the non-zero eigenvalues of ρ_1 and ρ_2 are identical. Call them $\lambda_j > 0$ and let $\phi_1^j,\ \phi_2^j$ be the corresponding eigenvectors of $\rho_1,\ \ \rho_2$. Then, the usual eigenvector decomposition of $\rho_1,\ \ \rho_2$ tells us that

$$\boxed{|\Psi\rangle = \sum_{j=1}^{\min\{d_1, d_2\}} \sqrt{\lambda_j}\ |\phi_1^j \otimes \phi_2^j\rangle}.$$

The λ_j are called Schmidt numbers.

Purity and Purification

A density matrix can be expanded in its orthonormal eigenvectors $|v_j\rangle$ and its eigenvalues λ_j (which satisfy $\lambda_j \geq 0,\ \sum_j \lambda_j = 1$):

$$\rho = \sum_j \lambda_j |v_j\rangle\langle v_j| \ . \qquad (*)$$

We say that ρ is pure if there is only *one* non-zero term, i.e. $\rho = |v\rangle\langle v|$ with $\|v\| = 1$.

Very Useful Purification Lemma: *For any $\mathcal{H}_1$ and and density matrix ρ_1 on $\mathcal{H}_1$ there is another Hilbert space $\mathcal{H}_2$ and a pure state ρ_{12} on $\mathcal{H}_{12} = \mathcal{H}_1 \otimes \mathcal{H}_2$ such that* $\rho_1 = \mathrm{Tr}_{\mathcal{H}_2}\rho_{12}$.

ρ_{12} is called a purification of ρ_1. One way to construct it is simply to use (*) and think of the index j as vector components in a second Hilbert space $\mathcal{H}_2$ and think of $\sum_j$ as a partial trace over that space and to think of $|v_j\rangle\langle v_j|$ as a vector in $\mathcal{H}_{12}$ that happens to be diagonal in the j basis of $\mathcal{H}_2$.

Purification was introduced in Araki, L, *Entropy Inequalities*, Commun. Math. Phys. **18**, 160-170 (1970).

As we said, whenever ρ_{12} is pure the spectrum of $\rho_1 = \mathrm{Tr}_2\, \rho_{12}$ equals the spectrum of $\rho_2 = \mathrm{Tr}_1 \rho_{12}$. This is the same statement (in another language) of the well known theorem of linear algebra that $\mathrm{spec} X^\dagger X = \mathrm{spec} X X^\dagger$ (except for possibly extra zero eigenvalues to make up the difference of dimensions).

Consequently, $\mathrm{Tr}_1\, F(\rho_1) = \mathrm{Tr}_2\, F(\rho_2)$ for any function, F, when ρ_{12} is pure. This simple fact (together with purification) allows us to get new inequalities from old ones.

EXAMPLE: Suppose We know that $\boxed{\mathrm{Tr}_{12} F(\rho_{12}) \leq \mathrm{Tr}_1 F(\rho_1) + \mathrm{Tr}_2 F(\rho_2)}$ for all ρ_{12}. Now purify $\rho_{12} \to \rho_{123}$ (pure). Then

$\mathrm{Tr}_{12} F(\rho_{12}) = \mathrm{Tr}_3 F(\rho_3), \quad \mathrm{Tr}_2 F(\rho_2) = \mathrm{Tr}_{13} F(\rho_{13})$

and our inequality becomes

$\mathrm{Tr}_3 F(\rho_3) \leq \mathrm{Tr}_1 F(\rho_1) + \mathrm{Tr}_{13} F(\rho_{13})$, equivalently, $\boxed{\mathrm{Tr}_{13} F(\rho_{13}) \geq \mathrm{Tr}_3 F(\rho_3) - \mathrm{Tr}_1 F(\rho_1)}$.

Since the indices are 'dummy', replace 3 by 2 in this inequality and obtain

$$\boxed{\mathrm{Tr}_2 F(\rho_2) - \mathrm{Tr}_1 F(\rho_1) \leq \mathrm{Tr}_{12} F(\rho_{12}) \ \leq \ \mathrm{Tr}_2 F(\rho_2) + \mathrm{Tr}_1 F(\rho_1)}$$ Woweee !!

HOMEWORK PROBLEM: This new inequality, which was obtained from the old one by purification, might seem to be valid only for special matrices $\rho_{12},\ \rho_1,\ \rho_2$ that arise after purification. In fact, it is general. Prove this!

(It is easy if you keep calm.)

Some Facts From the Last Millenium – to be Explained

The (von Neumann) **Entropy** of a density matrix ρ is

$$\boxed{S = S(\rho) := -\mathrm{Tr}_{\mathcal{H}}\, \rho \log \rho\ .}$$

2. *Definitions:* If ρ_{12} is a DM on $\mathcal{H}_{12}$, then we can define $\rho_1 := \mathrm{Tr}_2\, \rho_{12}$ on $\mathcal{H}_1$, etc., We can then define $S_{12},\ S_1,\ S_2$, etc.

3. **Subadditivity**(classical and quantum): $\boxed{S_{12} \leq S_1 + S_2.}$

4. Using purification, as just mentioned, we deduce from 3 the **Araki-L Triangle Inequality:** (No direct proof is known!)

$$\boxed{S_3 \leq S_{23} + S_2} \quad \text{and hence} \quad \boxed{|S_1 - S_2| \leq S_{12} \leq S_1 + S_2}$$

This is the closest one comes in QM to the *classical* monotonicity: $\boxed{S_{12} \geq \max\{S_1,\ S_2\}}$

Some More Ancient Facts

5. **Strong Subadditivity SSA (L, Ruskai) :** $S_{123} - S_{23} \leq S_{12} - S_2.$

I.e., Conditional entropy, $(S_{12} - S_2)$, decreases when a third space (3) is added.

We can also write this as $S_{A\bigcup B} + S_{A\bigcap B} \leq S_A + S_B.$

6. By 'purification' (pure ρ_{1234}) SSA becomes: $S_{14} + S_{12} \geq S_4 + S_2.$

Remarkable!!

7. Recall Golden-Thompson inequality: $\mathrm{Tr}e^{\log A + \log B} \leq \mathrm{Tr}AB \qquad (A, B \geq 0)$

The L, Ruskai proof of SSA was based on a triple matrix inequality (L, 1973 **):

$$\mathrm{Tr}e^{\log A + \log B - \log C} \leq \int_0^\infty \mathrm{Tr}A\frac{1}{C+t}B\frac{1}{C+t}\mathrm{d}t.$$

Note that there is *no* G-T inequality $\mathrm{Tr}e^{\log A + \log B - \log C} \leq \mathrm{Tr}AC^{-1}B.$

** L, *Convex Trace Functions and the Wigner-Yanase-Dyson Conjecture*, Adv. in Math. **11**, 267-288 (1973).

Essential Property of Entropy

Concavity and convexity is what makes thermodynamics work. Nowadays they don't tell you this in a course on stat-mech/thermodynamics, but Maxwell and Gibbs understood it very well. Let's discuss the concavity of entropy.

A function $f(x)$ is concave if $f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda) f(y)$ for all $0 \leq \lambda \leq 1$ and $x,\ y$. It is convex if the inequality goes the other way. Thus, x^4 is convex and $\sqrt{x}$ is concave for $x \geq 0$.

Note that x can stand for two variables (e.g., energy and volume) and the statement that entropy is a jointly concave function of energy and volume is a much stronger statement than that it is concave in energy for fixed volume and concave in volume for fixed energy. It is important to understand this very important principle as being at the heart of thermodynamics and the second law.

For our purposes we want to note that the quantum entropy, $S(\rho)$, is a concave function of ρ. Indeed, if $f(x)$ is a concave function of $x \in \mathbb{R}^+$ then $\mathrm{Tr} f(\rho)$ is a concave function of density matrices. (The function $f(x) = -x \ln x$ is concave, of course.)

HOMEWORK: Prove this by evaluating the traces using the eigenvectors of $\lambda\rho^1 + (1-\lambda\rho^2)$.

Subadditivity, (SA) $\boldsymbol{S_{12} \leq S_1 + S_2}$

Golden-Thompson: $\boxed{\mathrm{Tr}\ \mathrm{e}^{A+B} \leq \mathrm{Tr}\ \mathrm{e}^{A}\,\mathrm{e}^{B}}$ for A and B Hermitean.
This is NOT true without the trace Tr. It can be proved by Trotterizing e^{A+B} and then using the Cauchy-Schwarz inequality for traces: $|\mathrm{Tr}\, XY|^2 \leq \mathrm{Tr}\, X^*X\ \mathrm{Tr}\, Y^*Y$.

Peierls-Bogolubov: $\boxed{\mathrm{Tr}\,\mathrm{e}^{A+B} \geq \left(\mathrm{Tr}\,\mathrm{e}^{A}\right)\ \exp\{\langle B\rangle_A\} = \left(\mathrm{Tr}\,\mathrm{e}^{A}\right)\exp\{\mathrm{Tr} B e^{A}/\mathrm{Tr} e^{A}\}}$
for A and B Hermitean.

We define **Mutual Information** as $\boxed{M(\rho_{12}) = S(\rho_1) + S(\rho_2) - S(\rho_{12})}$.

Proof of SA: Now set $\Delta = -M_{12} = S_{12} - S_1 - S_2$. We want to show that $\mathrm{e}^{\Delta} \leq 1$.
$\Delta = \mathrm{Tr}_{12}\ \rho_{12}\{-\ln\rho_{12} + \ln\rho_1 + \ln\rho_2\}$, whence (by Peierls-Bogolubov),

$$\mathrm{e}^{\Delta} \leq (\mathrm{Tr}_{12}\,\rho_{12})^{-1}\exp\{\ln\rho_{12} - \ln\rho_{12} + \ln\rho_1 + \ln\rho_2\} = \mathrm{Tr}_{12}\ \exp\{\ln\rho_1 + \ln\rho_2\}$$

$= \mathrm{Tr}_{12}\,\rho_1\,\rho_2 = 1$. QED Thus, mutual information is always ≥ 0.

Strong Subadditivity, (SSA) $S_{123} + S_3 \le S_{13} + S_{23}$

Let's try the same idea as for SA. Set $\Delta = S_{123} + S_3 - S_{13} + S_{23}$ and try to prove $\mathrm{e}^{\Delta} \le 1$.

By following exactly the same route we end up with trying to validate the last line, namely, $\mathrm{Tr}_{123} \exp\{\ln \rho_{13} + \ln \rho_{23} - \ln \rho_3\} \le 1$.

If we could say that this is less than $\mathrm{Tr}_{123}\,\{\rho_{13}\,\rho_{23}\,\rho_3^{-1}\}$ we would be done (because doing Tr_2 gives $\mathrm{Tr}_{13}\{\rho_{13}\rho_3\rho_3^{-1}\} = 1.$) Unfortunately we can't say this. In fact it is nonsense because $\mathrm{Tr}_{123}\,\{\rho_{13}\,\rho_{23}\,\rho_3^{-1}\}$ is not necessarily real. One could try to play with

Golden-Thompson inequality: $\boxed{\mathrm{Tr}\ \mathrm{e}^{A+B} \le \mathrm{Tr}\ \mathrm{e}^{A}\,\mathrm{e}^{B}}$ for A, B Hermitean. But this is only 2 matrices, A, B, while we have 3, namely $\ln \rho_{13}, \ln \rho_{23}, -\ln \rho_3$. The remedy is

Triple matrix inequality:

$$\mathrm{Tr} e^{\log A + \log B - \log C} \le \int_0^\infty \mathrm{Tr} A \frac{1}{C+t} B \frac{1}{C+t} \mathrm{d}t.$$

Inserting $\ln\rho_{13}, \ln\rho_{23}, -\ln\rho_3$ for A, B, C we have

$$e^{\Delta} \leq \int_0^\infty \mathrm{Tr}_{123}\, \rho_{13} \frac{1}{\rho_3+t} \rho_{23} \frac{1}{\rho_3+t} \mathrm{d}t$$

$$= \int_0^\infty \mathrm{Tr}_{23}\, \rho_{3} \frac{1}{\rho_3+t} \rho_{23} \frac{1}{\rho_3+t} \mathrm{d}t$$

$$= \int_0^\infty \mathrm{Tr}_{3}\, \rho_{3} \frac{1}{\rho_3+t} \rho_{3} \frac{1}{\rho_3+t} \mathrm{d}t. = \mathrm{Tr}_3 \rho_3 = 1$$ QED

CONCLUSION: $\Delta \leq 1 \Longrightarrow$ SSA is true. $S_{123} + S_3 \leq S_{13} + S_{23}$

Matrix Concavity Theorem

The triple matrix inequality is useful in more than one context. For example it can be used to improve the celebrated *Maassen–Uffink generalized uncertainty principle*, which we will get to later.

For now, let us note that it is a consequence of the following **Matrix Concavity Theorem**. See Nielsen–Chuang (appendix).

Let A and B be arbitrary positive semidefinite operators (matrices), let $0 \leq p \leq 1$ be fixed, and let K be any fixed matrix. Then the function of A and B given by

$$f(A,\, B) = \mathrm{Tr}\, \left(A^p\, K^\dagger\, B^{1-p}\, K \right)$$

is jointly concave.

(Jointly concave means that when $A = \lambda A_1 + (1-\lambda) A_2, \quad B = \lambda B_1 + (1-\lambda) B_2$ then $f(A, B) \geq \lambda f(A_1, B_1) + (1-\lambda) f(A_2, B_2)$).

This mathematical inequality is the starting point of strong subadditivity and other theorems that allow the subject of quantum entropy and entanglement to go forward.

Reference: 'Convex Trace Functions and the WYD conjecture'

Matrix Concavity Theorem (continued)

The original proof in *"Convex Trace Functions"* used complex variable techniques which, while simple in principle, is tricky to set up. It took about 5 years for another proof (B. Simon, given in Nielsen-Chuang) and now there are several. They are all tricky and many claim to be the shortest. It all depends on what prior knowledge you take for granted.

The steps from the concavity theorem to the triple matrix theorem are only in *"Convex Trace Functions"*. They are sort of elementary but not obvious. It would be nice to have a more direct proof.

There is another route to SSA (discovered by Lindblad) *directly* from the matrix concavity theorem via *monotonicity of relative entropy*. We discuss that next.

Monotonicity of Relative Entropy

If ρ and σ are two density matrices we define (Umegaki) their Relative Entropy to be

$$S(\rho\|\sigma) = \mathrm{Tr}\rho\,(\ln\rho - \ln\sigma)\,.$$

which is a quantum analog of the classical $S(\rho\|\sigma) = \int \rho(x)\ln(\rho(x)/\sigma(x))\mathrm{d}x$.

It measures the closeness of ρ and σ. Note that $S(\rho_{12}\|\rho_1\otimes\rho_2) = M(\rho_{12}) = S_1+S_2-S_{12}$.

First, we note that $S(\rho\|\sigma) \geq 0$, and $= 0$ iff $\rho = \sigma$.

HOMEWORK problem: *Prove this using Golden Thompson and Peierls Bogolubov.*

Second, we note that $(\rho, \sigma) \to S(\rho\|\sigma)$ is jointly *convex*. This follows by differentiating the concave function $f(p) = \mathrm{Tr}\rho^p\,\sigma^{1-p}$ at $p = 1$. The derivative is $-S(\rho\|\sigma)$. Since f(1) is linear in (ρ,σ) this derivative must be concave. HOMEWORK: *Check this out!*

Monotonicity says that relative entropy increases after partial trace $\rho_{12} \to \rho_1$, $\sigma_{12} \to \sigma_1$

$$S(\rho_1\|\sigma_1) \leq S(\rho_{12}\|\sigma_{12})$$

But this follows from the joint convexity of $S(\cdot\|\cdot)$. Why? Because going from ρ_{12} to ρ_1 means taking the trace over $\mathcal{H}_1$ of the sum of the matrix elements (w.r.t. 2) instead of taking the sum after the trace over $\mathcal{H}_1$.

HOMEWORK: *Write out this remark formally.*

Monotonicity of Relative Entropy $\Rightarrow$ SSA

Recall $\quad S(\rho_\alpha \| \sigma_\alpha) \leq S(\rho_{\alpha\,\beta} \| \sigma_{\alpha\,\beta}).$

Now, make the following choices: $\mathcal{H}_\alpha \to \mathcal{H}_{12}, \quad \mathcal{H}_\beta \to \mathcal{H}_3$

and $\quad \rho_{\alpha\,\beta} \to \rho_{123}, \quad \sigma_{\alpha\,\beta} \to \rho_1 \otimes \rho_{23}.$ We get SSA:

$$S_{123} + S_2 \leq S_{12} + S_{23}$$

Cute, isn't it! HOMEWORK: *Check this.*

MORAL: Lots of important inequalities arise from each other, so it is really only important to get one of them in order to enter the club. The Matrix Concavity Theorem is such an entry point.

Useful reference: *'Strong Subadditivity of Quantum Entropy'* in Wikipedia

CPT MAPS

We just proved that tracing out degrees of freedom (which means, physically, ignoring them in a measurement) lowers *relative entropy*. This can be generalized to arbitrary CPT maps, which we first explain. The generalization is sometimes called the Data Processing Inequality and it has something to do with the connection between entropy and entanglement. First, CPT:

Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces and let $B(\mathcal{H})$, $B(\mathcal{K})$ be the bounded linear maps (i.e., matrices) on $\mathcal{H}$, resp. $\mathcal{K}$ (aka "superoperators"). Let Φ be a linear map from $B(\mathcal{H})$ to $B(\mathcal{K})$. We say that Φ is a PT Positive, Trace Preserving map if

1. $\boxed{P}$ $\Phi(A)$ is positive whenever $A \in B(\mathcal{H})$ is positive. (positive $\to$ positive).
2. $\boxed{T}$ Φ preserves traces: $\mathrm{Tr}_{\mathcal{H}} A = \mathrm{Tr}_{\mathcal{K}} \Phi(A)$.

We say that a PT map Φ is a CPT (Completely Positive, Trace Preserving) map if for any other Hilbert space $\mathcal{L}$, $\boxed{\Phi \otimes \mathbb{1}_{B(\mathcal{L})}}$ is PT.

A CPT map is also called a **channel**.

Crazy! But don't give up. Don't head for the exit. Hang in there.
What does this mean?

Quantum communication and information is all about CPT maps. When we do something to a density matrix of a system (such as adding or removing degrees of freedom) we want to be sure that we preserve positivity and the unit trace, of course, but we also want to preserve these properties when we think of our system as being a subsystem of the universe (the $\otimes \mathbb{1}$ is the environment).

If you think about it for a while, you will see that this latter condition is that when we tensor on the identity we preserve positivity. (The trace condition is automatic.)

Stinespring's theorem shows how to view this as ordinary unitary time evolution followed by tracing out the environment.

A simple example of a PT but not CPT map is $\Phi(A) = A^{\rm transpose}$.

HOMEWORK: Check what happens for this transpose map when $\mathcal{H}$, $\mathcal{K}$ and $\mathcal{L}$ are just two-dimensional. (* If you understand this you will have understood a lot.)

Kraus operators and Stinespring's Factorization theorem

K.Kraus figured out that Φ is a CPT map from $B(\mathcal{H}) \to B(\mathcal{K})$ if and only if it can be represented as:

$$\boxed{\Phi(A) = \sum_j F_j \, A \, F_j^\dagger}$$

for some (non-unique) operators F_j, called Kraus operators, from $\mathcal{K} \to \mathcal{H}$ satisfying $\sum_j F_j^\dagger F_j = \mathbb{1}_\mathcal{H}$.

HOMEWORK: What if $\sum_j F_j F_j^\dagger = \mathbb{1}_\mathcal{K}$ instead? What is preserved then? (Hint: CPT maps do not generally preserve the identity $\mathbb{1}$)

Earlier W. Stinespring found a general characterization of CPT as follows: There is another Hilbert space $\mathcal{L}$ and a unitary $U : \mathcal{K} \to \mathcal{H} \otimes \mathcal{L}$, with $\mathrm{Tr}_\mathcal{L} \, U \, U^\dagger = \mathbb{1}_\mathcal{H}$, such that

$$\boxed{\Phi(A) = \mathrm{Tr}_\mathcal{L} \; U^\dagger \, (A \otimes \mathbb{1}_\mathcal{L}) \, U} \quad \text{with } \mathrm{Tr}_\mathcal{L} \, U \, U^\dagger = \mathbb{1}_\mathcal{H}$$

This says that a CPT map can be thought of as first embedding A in a bigger space, then rotating with U, and finally bringing A back with a partial trace. We shall now use this to extend the monotonicity of relative entropy to the DPI for general CPT maps.

The extra space $\mathcal{L}$, which helped us effect the CPT map and then disappeared, is called an **Ancilla**. Such assistants are used often in QIT.

Data Processing Inequality or Complete Monotonicity of Relative Entropy

Recall the relative entropy: $S(\rho\|\sigma) = \mathrm{Tr}\rho\,(\ln\rho - \ln\sigma)$, and its monotonicity under partial traces: $S(\rho_1\|\sigma_1) \leq S(\rho_{12}\|\sigma_{12})$, (*)
where $\rho_1 = \mathrm{Tr}_2\rho_{12}$ and $\sigma_1 = \mathrm{Tr}_2\sigma_{12}$.

Partial trace Tr_2 is, of course, a CPT map from $B(\mathcal{H}_{12}) = B(\mathcal{H}_1 \otimes \mathcal{H}_2)$ to $B(\mathcal{H}_1)$. We want to generalize this to arbitrary CPT maps, which has been called the data processing inequality: ρ_{12} becomes simply ρ and ρ_1 becomes $\Phi(\rho)$, etc.. Thus, monotonicity or DPI is the following for all CPT maps Φ.

$$\boxed{S(\Phi(\rho)\|\Phi(\sigma)) \leq S(\rho\|\sigma)}$$

CPT maps can raise or lower entropy, but they always lower relative entropy. They make it harder to distinguish ρ from σ.

Stinespring says we can write $\Phi(\rho) = \mathrm{Tr}_{\mathcal{L}}\, U^\dagger\,(\rho \otimes \mathbb{1}_{\mathcal{L}})\,U$ and $\Phi(\sigma) = \mathrm{Tr}_{\mathcal{L}}\, U^\dagger\,(\sigma \otimes \mathbb{1}_{\mathcal{L}})\,U$. If we think of $\mathcal{H}_{12} = \mathcal{H} \otimes \mathcal{L}$ and $\rho_{12} = U^\dagger\,(\rho \otimes \mathbb{1}_{\mathcal{L}})U$ (and similarly for σ_{12}), then the $S(\rho\|\sigma) = \mathrm{Tr}\rho\,(\ln\rho - \ln\sigma)$ inequality (*) becomes the DPI inequality. QED.

Entanglement

We will be concerned with two objects (degrees of freedom, particles, systems, etc.) described by states (density matrices) on Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. The combined space is $\mathcal{H}_{12} = \mathcal{H}_1 \otimes \mathcal{H}_2$, and the density matrix on $\mathcal{H}_{12}$ is ρ_{12}

A simple example of a state with no correlations or entanglement is $\rho_{12} = \rho_1 \otimes \rho_2$. Others are possible, however.

We say that ρ_{12} is **not entangled** if it is **separable**, i.e., if it is possible to write

$$\boxed{\rho_{12} = \sum_j \lambda_j \, \rho_1^j \otimes \rho_2^j}$$

with all $\lambda_j > 0$ and ρ_1^j and ρ_2^j are density matrices.

If such a decomposition is not possible we say that ρ_{12} is entangled.

The size of the summation $\sum_j$ is at most $d_1 \, d_2 + 1$. (Note: It is always possible to decompose ρ_{12} into a sum of products, but not always with positive λ_j.)

Notes:

1. Since each ρ_i^j can be written as a sum of pure states (the eigenvalue decomposition) we may as well take the ρ_i^j to be pure.

2. We do not assume that the ρ_i^j are orthogonal for different j (meaning $\mathrm{Tr}\,\rho_i^j\,\rho_i^k \neq c^j\delta_{j,k}$).

3. A pure state is entangled unless it is a simple product:

$$\rho_{12} = |\Psi\rangle\langle\Psi|, \text{ and } \Psi = \psi_1 \otimes \psi_2 \qquad (= |\psi_1\rangle\,|\psi_2\rangle \text{ in some papers and books}).$$

Physicists often pretend that the states of interest are pure, but that is not always justified. Consider the optimum situation, that our laboratory is in a pure state $\rho_L = L\rangle\langle L$. The ρ_S of the system on our lab bench, which is obtained by partial trace of ρ_L, cannot be expected to be pure.

Entanglement Questions

How do we measure entanglement when it occurs?

What does entanglement have to do with measurements or with communication or other physical quantities? This will involve Bell states (discussed later on).

If we have a measure, is it **Faithful**? That is, does the measure give us a positive value if and only if ρ_{12} is entangled.

As one might expect, such good measures turn out to be complicated to evaluate, Can we find useful, simple tests (entanglement witnesses) that tell us incomplete, but useful information about entanglement.

Entanglement of Formation

The following definition of entanglement has something to do with the number of Bell states *needed* to form our given state ρ_{12}. This will be explained later. For now we look at its mathematical properties.

$$E^f(\rho_{12}) = \inf\left\{\sum_{j=1}^n \lambda_j S_1(\mathrm{Tr}_2\omega_j) \ : \ \rho_{12} = \sum_{j=1}^n \lambda_j\omega_j\right\}$$

where the ω_j are density matrices and the λ_j are positive and sum to 1. This means we try to decompose ρ_{12} into lots of pieces and add up the entanglement of each. Since entropy is concave, we can take the ω_j to be pure states. If ρ_{12} itself is pure then there can only be one term in the sum, in which case E^f is just the usual entanglement as normally used (namely $E(\rho_{12}) = S_1(\mathrm{Tr}_2\rho_{12})$). Our definition of E^f *transports this usual definition to density matrices.*

It is easy to see that $E^f(\rho_{12}) = 0$ if and only if ρ_{12} is separable.

Thus, E^f is a faithful measure of entanglement! HOMEWORK: Prove this!
We know that E^f is *not additive*. (meaning $E^f(\rho_{12} \otimes \sigma_{12}) = E^f(\rho_{12}) + E^f(\sigma_{12})$) but no counterexamples are known!

Squashed Entanglement

Another definition of Entanglement is the squashed entanglement, introduced by Tucci and Christandl-Winter. It has something to do with the number of Bell states that one can *extract* from a given state ρ_{12}.

$$E^{sq}(\rho_{12}) = \frac{1}{2} \inf_{\rho_{123} \to \rho_{12}} \left(S_{13} + S_{23} - S_{123} - S_3 \right).$$

The quantity in parenthesis is always positive by Strong Subadditivity. One is asking for an extension of ρ_{12} with minimum SSA difference.

Brandao-Christandl-Yard proved that E^{sq} is faithful. It is zero if and *only if* ρ_{12} is separable. It is also additive (hint: use $\mathcal{H}_3 \otimes \mathcal{H}_3$ in place of $\mathcal{H}_3$.)

HOMEWORK: 1. Find one simple extension of ρ_{12} that will demonstrate that

$$E^{f}(\rho_{12}) \geq E^{sq}(\rho_{12}).$$

2. Show that $E^{sq}(\rho_{12}) = 0$ if ρ_{12} is separable. 3. If ρ_{12} is pure then $E^{sq}(\rho_{12}) = S_1$.

Clearly E^{sq} is even harder to compute than E^{f}. We will next prove a simple lower bound for E^{sq} (and hence E^{f}), which is easier to compute. It is not faithful, unfortunately. It was derived together with Eric Carlen.

Squashed Entanglement Lower Bound

Recall our earlier result, obtained by purification , that SSA is equivalent to:

$$S_{14} + S_{12} \geq S_4 + S_2. \qquad \text{for any } \rho_{124}.$$

Now exchange the dummy indices 4 and 1 to get $S_{14} + S_{24} \geq S_1 + S_2$. Add the two inequalities and get $S_{12} + 2S_{14} + S_{24} \geq S_1 + 2S_2 + S_4$.

Purify this last inequality (pure ρ_{1234}) and get, believe it or not, an extended SSA ! :

$$\frac{1}{2}(S_{13} + S_{23} - S_{123} - S_3) \geq \mathbf{S_1 - S_{12}},$$

which implies that $E^f(\rho_{12}) \geq E^{sq}(\rho_{12}) \geq \max\{S_1 - S_{12},\ S_2 - S_{12},\ 0\}$

(Christandl-Winter had this with just the average $\frac{1}{2}(S_1+S_2)-S_{12}$, which is very different.)

$\mathcal{MORAL}$: Define the Extreme Quantum Regime by *negative* conditional entropy (i.e., $S_1 > S_{12}$ or $S_2 > S_{12}$). This never happens classically. This is the regime of most interest to physicists (e.g., the ground state). In this regime there is always squashed entanglement.

Pinsker's Inequality and Quantitative Subadditivity

We just talked about extended SSA.

While on this topic, we mention (an improved) Pinsker's inequality, heavily used in QIT, and quantitative subadditivity (Carlen, L).

Recall relative entropy: $S(\rho||\sigma) = \mathrm{Tr}\rho(\ln\rho - \ln\sigma)$.

$$S(\rho||\sigma) \geq -2\ln\mathrm{Tr}\left[\rho^{1/2}\sigma^{1/2}\right] \geq \mathrm{Tr}\left[\sqrt{\rho} - \sqrt{\sigma}\right]^2 \ .$$

By taking $\rho = \rho_{12}, \text{ and } \sigma = \rho_1 \otimes \rho_2$ we get quantitative subadditivity:

$$S_1 + S_2 - S_{12} \geq -2\ln\mathrm{Tr}_{12}\sqrt{\rho_{12}}\ \sqrt{\rho_1 \otimes \rho_2}$$

With $\Delta = \frac{1}{2}S(\rho||\sigma)$, by the Peierls-Bogoliubov and Golden-Thompson inequalities,

$$\begin{aligned} e^{-\Delta} &= \exp\left[\mathrm{Tr}\rho\tfrac{1}{2}(\ln\sigma - \ln\rho)\right] \leq \mathrm{Tr}\exp\left[\tfrac{1}{2}(\ln\sigma + \ln\rho)\right] \\ &\leq \mathrm{Tr}\exp\left[\tfrac{1}{2}\ln\sigma\right]\exp\left[\tfrac{1}{2}\ln\rho\right] = \mathrm{Tr}\left[\sigma^{1/2}\rho^{1/2}\right] \ . \end{aligned}$$

Entanglement Witnesses

We have discussed two examples of faithful entanglement measures. They are not so easy to use, however. We also need simpler, partial measures, called entanglement witnesses which can be used to help us decide if a state is entangled or not.

The set of non-entangled (= separable) states on a given product of Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$ is a convex subset of the set of all operators on $\mathcal{H} \otimes \mathcal{K}$ (Why?). (Meaning that if ρ_{12} and σ_{12} are separable then so is $\lambda\rho_{12} + (1-\lambda)\sigma_{12}$, for all $0 \leq \lambda \leq 1$.) Call this convex set Sep_{12}.

Because it is convex, given ρ_{12} that is not in Sep_{12} there is a separating hyperplane (a linear functional) such that ρ_{12} is on the negative side of the plane. More specifically: There is a Hermitean operator A_{12} on $\mathcal{H} \otimes \mathcal{K}$ so that $\boxed{\mathrm{Tr}_{12} A_{12} \rho_{12} < 0}$ and that $\mathrm{Tr}_{12} A_{12} \sigma_{12} > 0$ for all product density matrices $\sigma_{12} = \sigma_1 \otimes \sigma_2$.

An entanglement witness may be described as follows (Horodecki's): Let Φ be a Positive (not completely positive) map on $\mathcal{H}_1$. Then, if ρ_{12} is separable $\Phi_1 \otimes \mathbb{1}_2(\rho_{12})$ is a positive operator (i.e., no negative eigenvalues).

Conversely, if $\Phi_1 \otimes \mathbb{1}_2(\rho_{12})$ is positive for all such Φ_1 then ρ_{12} is separable. Not Trivial!

Bell States

Let's try to relate the definitions of entanglement to familiar objects, namely Bell states. A **qubit** is just a system where the Hilbert space is $\mathcal{H} = \mathbb{C}^2$. The four Bell states are pure bipartite 2-qubit states $|\Psi\rangle\langle\Psi|$ defined by the four vectors in $\mathbb{C}^2 \otimes \mathbb{C}^2$:

$$\Phi_+ = \frac{1}{\sqrt{2}}(|\downarrow\rangle \otimes |\downarrow\rangle + |\uparrow\rangle \otimes |\uparrow\rangle) \qquad \Phi_- = \frac{1}{\sqrt{2}}(|\downarrow\rangle \otimes |\downarrow\rangle - |\uparrow\rangle \otimes |\uparrow\rangle)$$
$$\Psi_+ = \frac{1}{\sqrt{2}}(|\uparrow\rangle \otimes |\downarrow\rangle + |\downarrow\rangle \otimes |\uparrow\rangle) \qquad \Psi_- = \frac{1}{\sqrt{2}}(|\uparrow\rangle \otimes |\downarrow\rangle - |\downarrow\rangle \otimes |\uparrow\rangle)$$

provide an orthonormal basis of $\mathcal{H}_1 \otimes \mathcal{H}_2$ consisting of maximally entangled states.

Definition: A maximally entangled state on $\mathcal{H}_1 \otimes \mathcal{H}_2$ has the form
$\rho_{12} = \frac{1}{\min\{d_1, d_2\}} \sum_{j=1}^{\min\{d_1, d_2\}} |\xi_j \otimes \chi_j\rangle\langle\xi_j \otimes \chi_j|$
where the $|\xi_j\rangle$ are from an orthonormal basis of $\mathcal{H}_1$, and similarly $|\chi_j\rangle$. The reason such an ρ_{12} is maximally entangled is that $E^f(\rho_{12}) = \ln(\min\{d_1, d_2\})$ which is the most it can be.

Warning: In this business one usually uses logarithms to base 2. I have been using natural logs, so that $E^f(|\Phi_+\rangle\langle\Phi_+|) = \ln(2)$. The next definition we need is (LOCC).

Local Operations and Classical Communication = LOCC

Any ρ_{12} on any $\mathcal{H}_1 \otimes \mathcal{H}_2$ and be converted to $\sigma_1 \otimes \sigma_2$ on any $\mathcal{K}_1 \otimes \mathcal{K}_2$ by some CPT map. But we are interested in CPT maps that act on only one factor. That is it maps $B(\mathcal{H}_1) \to B(\mathcal{K}_1)$ or it maps $B(\mathcal{H}_2) \to B(\mathcal{K}_2)$. The first kind of map is $\Phi_1 \otimes \mathbb{1}_2$ and the second is $\mathbb{1}_1 \otimes \Phi_2$, These are "Local Operations". Maps of the form $\Phi_1 \otimes \Phi_2$ are also called local operations.

In quantum communication we are usually interested in $\mathcal{H}_1 = \mathcal{K}_1$ and $\mathcal{H}_2 = \mathcal{K}_2$,

There is a soap opera that goes with this in which **'Alice' and 'Bob'** are actors who are interested in finding out what ρ_{12} is by making measurements on it using local operations. Alice on $\mathcal{H}_1$ and Bob on $\mathcal{H}_2$. They communicate their results to each other by carrier pigeon, for example, or any other *classical communication system*. $\rho_{12} \longrightarrow \rho_{AB}$.

Let us begin by defining a measurement. (What! Does he never stop defining things and get down to business?)

Measurement

The measurement apparatus on a state ρ is defined by a CPT map which, according to Kraus, is

$$\Phi(\rho) = \sum_{m=1}^{\mu} M_m\, \rho M_m^{\dagger} \qquad \text{with} \qquad \sum_{m=1}^{\mu} M_m^{\dagger}\, M_m = \mathbb{1}_{12}.$$

The M_m are the possible 'measurement outcomes' . A measurement outcome is a choice of one m (and the reduction of the state to $(\mathrm{Tr}\, M_{m=1}^{\mu}\, \rho M_m^{\dagger})^{-1} M_m\, \rho M_m^{\dagger}$, but we won't need this.)

For a pure state, this reads $|\Psi\rangle \to \langle\Psi|M_m^{\dagger} M_m|\Psi\rangle^{-1/2}\, M_m|\Psi\rangle$.

(Technically, a local operation is a CPT map to operators on a Hilbert space. To make the measurement apparatus into such a map we can define it as a map to $\mathcal{H} \otimes \mathcal{L}$, where $\mathcal{L} = \mathbb{C}^{\mu}$ is an ancilla, and the map is into the set of *diagonal* $\mu \times \mu$ matrices on $\mathbb{C}^{\mu}$. Thus, a measurement apparatus maps ρ into a block-diagonal matrix, each block belonging to $B(\mathcal{H})$. If you don't get this, forget it.)

Now the commercial is over and we take you back to Alice and Bob.

Measurement and Entanglement

Schrödinger again:

By interactions the two representatives (or ψ-functions) become entangled. To disentangle them we must gather further information by experiment, although we knew as much as anybody could possibly know about all that happened (during the interaction).
Of either system, taken separately, all previous knowledge may be entirely lost, leaving us but one privilege: to restrict the experiments to one only of the two systems. After reestablishing one representative by observation, the other one can be inferred simultaneously. In what follows the whole of this procedure will be called the disentanglement. Its sinister importance is due to its being involved in every measuring process and therefore forming the basis of the quantum theory of measurement, threatening us thereby with at least a *regressus in infinitum*, since it will be noticed that the procedure itself involves measurement.

Another way of expressing the peculiar situation is: the best possible knowledge of a whole does not necessarily include the best possible knowledge of all its parts, even though they may be entirely separated...

An Example

A&B know that ρ_{12} is a pure state whose $|\Psi_{12}\rangle$ is one of the following Bell states. They want to decide which one it is.

$\Phi_+ = \frac{1}{\sqrt{2}}(|\downarrow\rangle \otimes |\downarrow\rangle + |\uparrow\rangle \otimes |\uparrow\rangle)$ $\Psi_+ = \frac{1}{\sqrt{2}}(|\uparrow\rangle \otimes |\downarrow\rangle + |\downarrow\rangle \otimes |\uparrow\rangle)$

Here, $\mathcal{H}_1$ is a qubit in Princeton and $\mathcal{H}_2$ is one in Timbuktu. Classical communication is not easy but that doesn't stop this dauntless pair.

With a local operation Alice measures spin up. She could have measured something else, (such as the spin in the X direction) which would make the story more interesting. However, her measurement, indeed no measurement she could have made, distinguishes the two states.

But now Bob measures spin down and that clinches the matter. Ψ_+ wins!

The state has been reduced to $|\uparrow\rangle \otimes |\downarrow\rangle$.

Nielsen's Theorem

We can ask how powerful an LOCC operation can be. That is, given ρ_{12} and σ_{12} on the same $\mathcal{H}_{12}$, can we find a LOCC that maps ρ_{12} to σ_{12}. The general answer is not known, but the necessary and sufficient condition is known (Nielsen) if these are both pure states. It is stated as follows.

Suppose Λ and Γ are two ordered, finite sequences of d real numbers with the same sum. $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_d$ and $\gamma_1 \geq \gamma_2 \geq ... \geq \gamma_d$ with $\sum \lambda_j = \sum \gamma_j$.
We say that Γ majorizes Λ (in symbols $\Gamma \succ \Lambda$) if the partial sums satisfy:

$$\sum_{j=1}^{k} \gamma_j \geq \sum_{j=1}^{k} \lambda_j \qquad \text{for } \textit{every } k = 1,\, 2,\, ...,\, d.$$

If both bipartite states are pure and if $\rho_1,\ \sigma_1$ are the one-system reduced density matrices, let Λ (respectively Γ) be the ordered sequences of eigenvalues of ρ_1 (resp. σ_1). There is an LOCC that takes $\rho_{12} \to \sigma_{12}$ if and only if $\boxed{\Gamma \succ \Lambda}$. (And hence $S(\rho_1) \geq S(\sigma_1)$.)

This is a very good theorem and hard to prove. A similar majorization theorem relates the diagonals and the eigenvalues of a hermitean matrix. (Did you know this fact?)

Now that we have explored LOCC, let's get back to Bell states and entanglement.

The Operational Meaning of Entanglement

If Alice and Bob have n copies of an entangled bipartite *pure state* Ψ_{12}, i.e., $\Psi_{12}^{\otimes n}$, how many copies of a Bell state, say Ψ_+ can they produce from it using only local operations and classical communication? Conversely, to produce $\Psi_{12}^{\otimes n}$ from a 'stream' of Bell States $\Psi_+^{\otimes m}$, using local operations and classical communication, how large does m have to be?

In the large n limit, one can define two operational measures of entanglement, **distillable entanglement E^d** and **entanglement cost E^c** that quantify, respectively, the rate at which one can 'distill' ebits (maximally entangled states, Bell states) from Ψ_{12}, and how many ebits does it takes to build Ψ_{12}. Both are shown to coincide with $E(\Psi_{12}) = E^f(\rho_{12}) = S_1$ for pure states.

The situation for mixed states is much more complicated. It was conjectured that $E^c = E^f$ but this is now known to be false, although no explicit counterexamples are known. It is easy to prove that $E^c \leq E^f$, and it is conjectured, but not proved, that $E^c = E^f$ in 'most' cases.

Entanglement Distillation and Entanglement Cost

Alice and Bob have a state ρ_{12}. They make (in principle) a large number of copies of this state and form the n-fold tensor product $\rho_{12}^{\otimes n} = \rho_{12} \otimes \rho_{12} \otimes \ldots . \rho_{12}$. They ask if they can find an LOCC that maps n' copies of a Bell state $|\Psi_+\rangle\langle\Psi_+|$ into the state $\rho_{12}^{\otimes n}$ and try to make n' as *small* as possible. (Note: It doesn't matter which Bell state we use.) To be honest, they have to produce the n' copies to within a small tolerance, but we won't quibble about that. Then, in general,

$$E^c(\rho_{12}) = \lim_{n\to\infty} n'/n.$$

It is also

$$E^c(\rho_{12}) = \lim_{n\to\infty} \frac{1}{n} E^f(\rho_{12}^{\otimes n}).$$

As mentioned above, $E^c(\rho_{12}) = E^f(\rho_{12})$ for pure ρ_{12} and $E^c(\rho_{12}) \leq E^f(\rho_{12})$ generally.

For entanglement distillation we do things the other way around. We LOCC map a product of n copies of ρ_{12} into n' Bell states, with n'/n as *large* as possible. Then

$$E^d(\rho_{12}) = \lim_{n\to\infty} n'/n.$$

In general, $E^d(\rho_{12}) \leq E^c(\rho_{12})$

The Connection Between Entropy and Entanglement

You may wonder why entropy plays such a dominant role in the quantification of entanglement. For pure states we can use Nielsen's theorem to make this very clear to anyone familiar with elementary equilibrium statistical mechanics, where the difference between energy and free-energy is TS (by definition) and $S = -\sum p \ln p$.

A Bell state has $\rho_1 = \frac{1}{2}\mathbb{1}_1$. This maximally entangled density matrix is dominated by any other $\mathbb{C}^2$ density matrix since its eigenvalues are $\frac{1}{2}$, $\frac{1}{2}$. The eigenvalues of n' tensor copies of this Bell state (which has dimension $= 2^{n'}$) are easily seen to have only the eigenvalue $2^{-n'}$ and this occurs with multiplicity $2^{n'}$ (thus making the trace $= 1$).

Now suppose Alice & Bob share a bipartite pure state $\rho_{12} = |\Psi\rangle\langle\Psi|$ of dimension d^2. Let λ_j denote the set of Schmidt numbers (eigenvalues of ρ_1) by $\lambda_1, \dots, \lambda_d$. Now look at n tensor copies of ρ_{12}, which is a pure state, of course, with the $\rho_1(\rho_{12}^{\otimes n}) = \mathrm{Tr}_2^{\otimes n} \rho_{12}^{\otimes n}$ density matrix equal to the n-fold tensor product $\rho_1^{\otimes n}$ of ρ_1. What are the eigenvalues of this "ρ_1"? Each eigenvalue is a product of Schmidt numbers, one from each factor, without any restriction. $\lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_n}$.

Connection Between Entropy and Entanglement (cont.)

Now, think of a stat-mech problem of n independent particles with one-body eigenvalues $-\ln\lambda_1, \cdots, -\ln\lambda_d$ at inverse temperature $\beta = 1$. The partition function is $Z = (\sum_{j=1}^{d}\lambda_j)^n = 1$, whence the free energy is $F = \ln Z = 0$. We can compute the entropy per particle, which, as always, is $S = -\sum p\,\ln p = -\sum_{j=1}^{d}\lambda_j\,\ln\lambda_j$.

We used the Boltzmann factor $e^{\ln\,\lambda_j}$, and not $e^{\ln\,(1/d)}$, as you might wish, because we are interested in estimating errors in $\mathrm{Tr}\rho_1$ caused by perturbations, and not in $\mathrm{Tr}\,d^{-1}\mathbb{1}$.

In this problem the average energy/particle is $E = S$, since $F = 0$. Physics (actually the **law of large numbers**, a.k.a. the 'asymptotic equipartition theorem') tells us that the distribution of the n-body energy values is sharply peaked around nE. Except for a small probability (i.e., a small change in the trace) all nonzero eigenvalues of ρ_1 lie in the interval $(e^{n(-S-\epsilon)}, e^{n(-S+\epsilon)})$ for any $\epsilon > 0$ as $n \to \infty$. Since the Tr "ρ_1" $= \mathrm{Tr}\ \rho_1^{\otimes n} = 1$, there are between $e^{n(S-\epsilon)}$ and $e^{n(S+\epsilon)}$ eigenvalues in this interval. The missing eigenvalues are rare or close to zero.

Connection Between Entropy and Entanglement (end)

Let us summarize the situation. If we have n' Bell states the non-zero eigenvalues of $\rho_1^{\otimes n'}$ is $2^{-n'}$ (with $2^{+n'}$ degeneracy).

The non-zero eigenvalues of $\rho_1^{\otimes n}$ for our pure bipartite state lie (except for a vanishingly small number) somewhere between $e^{-n(S-\epsilon)}$ (with $e^{+n(S-\epsilon)}$ degeneracy) and $e^{-n(S+\epsilon)}$ (with $e^{+n(S+\epsilon)}$ degeneracy), for any $\epsilon > 0$ and very large n. The entropy of $\rho_1^{\otimes n}$ is nS with $S = -\sum_{j=1}^{d} \lambda_j \ln \lambda_j$.

Nielsen's theorem tells us that we can find an LOCC map from Bell state to our pure ρ_{12} if spectrum (ρ_{12}) majorizes spectrum (Bell). This occurs if $e^{+n(S-\epsilon)} \leq 2^{n'}$. Conversely, we can LOCC from ρ_{12} to Bell if $e^{+n(S+\epsilon)} \geq 2^{n'}$. Since ϵ is arbitrary (as $n \to \infty$ the two conditions coincide, and we conclude that

$$\boxed{E^c(\rho_{12}) = E^d(\rho_{12}) = \lim_{n\to\infty}(n'/n) = S(\rho_1)/\ln 2 = E^f(\rho_{12})/\ln 2}.$$

Entropy and Quantum Uncertainty

We switch now to a use of entropy as a measure of quantum mechanical uncertainty. It has much more to say than does the Heisenberg uncertainty principle.

Given a density matrix ρ on $\mathcal{H}$ and an O.N. basis of vectors $A = \{a_j\}$ of $\mathcal{H}$, define probabilities $p_j = \langle a_j, \rho\, a_j \rangle$. Then define the 'classical entropy' $H(\rho) = -\sum_j p_j \log p_j$. $H(A)$ can be quite small, even if $S(\rho)$ is large, but Maasen & Uffink (PRL 1988) found an uncertainty principle: Let $B = \{b_j\}$ be another O.N. basis. Then

$$H(A) + H(B) \geq -2 \log \{\sup_{j,k} |(a_j,\ b_k)|\}. \qquad (*)$$

This can be generalized to continuous bases, e.g., delta functions and plane waves:

$$-\int \rho(x,x) \ln \rho(x,x) dx - \int \widehat{\rho}(k,k) \ln \widehat{\rho}(k,k) dk \geq 0,$$

with $\rho(x,x) = \langle x|\rho|x\rangle$, etc.

NOTE: The entropies on the left can be arbitrarily negative!

Following a conjecture and result of Rumin, Rupert Frank and I improved this to:

$$-\int \rho(x,x)\ln\rho(x,x)dx - \int \widehat{\rho}(k,k)\ln\widehat{\rho}(k,k)dk \geq S(\rho).$$

The quantum entropy $S(\rho)$ is bounded above by two classical entropies!.

Even more generally, we can generalize to any two spaces $L^2(X,\,\mu), L^2(Y,\,\nu)$ and with a unitary $\mathcal{U} : X \to Y$ having an integral kernel $\mathcal{U}(x,y)$ that is bounded.

As before, we have $\widehat{\rho} = \mathcal{U}^*\rho\,\mathcal{U}$ (like the Fourier transform), and then

$$-\int_X \rho(x,x)\ln\rho(x,x)d\mu(x) - \int_Y \widehat{\rho}(y,y)\ln\widehat{\rho}(y,y)d\nu(y) \geq S(\rho) - 2\log\{\sup_{x,\,y}|\mathcal{U}(x,y)|\}$$

Note that X and Y can be quite different. E.g., $X = \mathbb{Z}$ (the integers) and $Y = [-\pi,\pi]$ (the Brillouin zone). Thus, the X integral is a sum and the Y integral is continuous. (This is reminiscent of strong subadditivity: $S_1 + S_2 \geq S_{12}$.)

Our proof of this last theorem uses only two simple, well known tools:

1. (usual) Golden-Thompson inequality: $\mathrm{Tr}\, e^{A+B} \leq \mathrm{Tr}\, e^{A}\ e^{B}$.

2. Gibbs' Variational Principal (or Peierls-Bogolubov inequality): For self-adjoint H $\mathrm{Tr}\, \rho H\ +\ \mathrm{Tr}\, \rho \log \rho \geq -\log \mathrm{Tr}\, e^{-H}$.

HOMEWORK: Prove the theorem or look it up in Annales Inst. Henri Poincaré **13** (2012)

Now we turn to more serious matters, which are *truly quantum mechanical* and use strong subadditivity and the 'triple matrix inequality'.

Two Hilbert Space Theorem

By definition, a ***classical*** **measurement on a density matrix** ρ on $\mathcal{H}_1$ is a collection of operators M_m on $\mathcal{H}_1$ such that $\sum_m M_m^* M_m = \mathbb{1}_{\mathcal{H}_1}$; the probability of measurement m is $\mathrm{Tr}_1 \, M_m \rho M_m^*$.

Now we suppose that we have the usual ρ_{12} on $\mathcal{H}_{12} = \mathcal{H}_1 \otimes \mathcal{H}_2$. We also suppose that we have two kinds of measurements, which we shall call A and B, and we want to find some kind of uncertainty principle relating the two. The measurement operators will be denoted by A_m and B_n.

Recall that the usual conditional entropy of ρ_{12} is $S(1|2) := S_{12} - S_2$. In its place we define the classical/quantum conditional entropy for the A measurement by

$$H(1^A|2) := -\sum_m \mathrm{Tr}_2(\mathrm{Tr}_1 A_m \rho_{12} A_m^*) \log(\mathrm{Tr}_1 A_m \rho_{12} A_m^*) - S_2.$$

and similarly for a B measurement.

Note where the Σ_m is! This is not the conditional entropy of $\sum_j A_m \rho_{12} A_m^*$.

Two-Space Theorem: Suppose we have two measurements $\{A_m\}$ and $\{B_n\}$. Then

$$\boxed{H(1^A|2) + H(1^B|2) \geq S(1|2) - \log c_1,} \qquad \text{where} \qquad c_1 = \sup_{m,n} \left(\mathrm{Tr}_1 B_n A_m^* A_m B_n^*\right)$$

Note that if $\rho_{12} = \rho_1 \otimes \rho_2$ then this theorem reduces to the generalized Maassen-Uffink theorem mentioned before.

Two Hilbert Space Theorem (continued)

This theorem is difficult because it brings out the terrors of “entanglement”. It was conjectured by Renes and Boileau in 2009, who realized that the SSA theorem, or an equivalent, would be needed in its proof. Berta et. al. proved the special 'rank-one' case where all $\{A_j\}$, and all $\{B_k\}$ are O.N. rank-one projectors.

Subsequently, Coles, Griffiths, et. al. and Tomamichael and Renner eliminated some of the rank-one conditions. All these proofs were quite long, however. It turns out that the theorem can be proved in a few lines by merely adapting the original proof of SSA! That is, by using the triple matrix inequality.

There is also a 3-space theorem, but better to stop this while we are ahead.

But it is worth mentioning the continuous version of this theorem (Heisenberg like).

Continuous Version (e.g., Fourier Transform)

We can now formulate **the entangled version of the Maassen-Uffink theorem**, which relates the classical entropies in Fourier and in configuration space. This is done by allowing the measurement operators A_m and/or B_n indexed by discrete m, n to become indexed continuously. Sums are replaced by integrals in this case. For the Fourier transform, for example, we apply the 2-space theorem and infer that

$$H(1^A|2) + H(1^B|2) \; \geq \; S_{12} - S_2$$

where

$$H(1^A|2) = -\int_{\mathbb{R}^d} \mathrm{dx}\ \mathrm{Tr}_2 \langle \mathrm{x}|\rho_{12}|\mathrm{x}\rangle_{\mathcal{H}_1} \log \langle \mathrm{x}|\rho_{12}|\mathrm{x}\rangle_{\mathcal{H}_1} - \mathrm{S}_2$$

and

$$H(1^B|2) = -\int_{\mathbb{R}^d} \mathrm{dk}\ \mathrm{Tr}_2 \langle \mathrm{k}|\rho_{12}|\mathrm{k}\rangle_{\mathcal{H}_1} \log \langle \mathrm{k}|\rho_{12}|\mathrm{k}\rangle_{\mathcal{H}_1} - \mathrm{S}_2\,.$$

In this case $c_1 = \sup_{x,k} |e^{2\pi i k\cdot x}| = 1$ and, therefore, $\ln c_1 = 0$.

Fermion Entanglement

Here, *solely for your amusement*, is an exercise (with E. Carlen) in computing entanglement. It is where one is led by just following the rules.

Bosons are sometimes thought to be more complicated than fermions because they can 'condense'. But condensed bosons that are in a product, or 'coherent', state $\Psi = \phi(x_1)\,\phi(x_2)\cdots\phi(x_N)$ are *not* entangled in any way (by usual definitions of entanglement) whereas fermions are *always* entangled by the Pauli principle. Our goal is to quantify the minimum possible entanglement and, as folklore might suggest, show that pure Slater determinant states give the minimum entanglement. If this is the case then

Slaters can be said to be the fermionic analog of boson condensation!

We study the bipartite density matrix of 2 fermions embedded in a sea of N fermions. Some results depend on N, while others do not.

Theorem #1

We return to the quantitative investigation of fermionic entanglement. We naturally assume, going forward, that $\mathcal{H}_1 = \mathcal{H}_2$.

Our first theorem expresses an extremal property of Slater determinants for E_f.

Theorem 1. *Let ρ_{12} be fermionic, i.e., suppose that the range of ρ_{12} is contained in $\mathcal{H} \wedge \mathcal{H}$. Then,*

$$\boxed{E_f(\rho_{12}) \geq \ln(2),}$$

and there is equality if and only if ρ_{12} is a convex combination of pure-state Slater determinants; i.e., the state is fermionic separable.

In other words,

$$E_f^{antisymmetric}(\rho_{12}) := E_f(\rho_{12}) - \ln(2)$$

is a faithful measure of fermionic entanglement.

THEOREM #2

We all know subadditivity of entropy (positivity of mutual information): $S_1 + S_2 - S_{12} \geq 0$, and that equality occurs only when $\rho_{12} = \rho_1 \otimes \rho_2$. This cannot happen for fermions.

Theorem 2 (**Mutual Information of fermionic ρ_{12}:**).

$$S_1 + S_2 - S_{12} \geq \ln\left(\frac{2}{1 - \mathrm{Tr}\rho_1^2}\right) ,$$

and there is equality if and only if the N-particle fermionic state is a pure-state Slater determinant. (not a convex combination of Slaters)

Recall that for N fermions $\rho_1 = \rho_2 \leq \frac{1}{N}\,\mathrm{Id}$, and equality occurs only for an N-particle Slater. For a Slater $S_1 = S_2 = \ln N$.

Do Slaters minimize squashed entanglement?

We cannot find the minimum of E_{sq} over all fermionic states but

We conjecture that the minimum E_{sq} occurs for Slaters

We conjecture that E_{sq} for a Slater is given by:

$$E_{sq}(\rho_{12}) = \begin{cases} \frac{1}{2}\ln\frac{N+2}{N-2} & \text{if } N \text{ is even} \\ \\ \frac{1}{2}\ln\frac{N+3}{N-1} & \text{if } N \text{ is odd} \end{cases}$$

SERIOUS HOMEWORK: **Prove or disprove these conjectures.**

Squashed entanglement continued

At first we thought that the minimum was the same as for the entanglement of formation E_f, namely, $\ln(2)$. This is grossly incorrect! The minimum is at least as small as the value just mentioned, i.e.,

$$E_{sq}(\rho_{12}) = \begin{cases} \frac{1}{2}\ln\frac{N+2}{N-2} & \text{if } N \text{ is even} \\ \frac{1}{2}\ln\frac{N+3}{N-1} & \text{if } N \text{ is odd.} \end{cases}$$

This upper bound shows that E_{sq} for a Slater depends heavily on N, namely $\approx 2/N$.

It is obtained by starting with an N-particle Slater (which is pure and which gives us the required ρ_{12}) and then taking ρ_{123} to be the (mixed) $N/2$-particle reduced density matrix of this Slater state. Thus, $\dim\mathcal{H}_3 = \binom{N}{N/2-\,2}$. Then

$$S_{123} = \ln\binom{N}{N/2}, \quad S_3 = \ln\binom{N}{N/2-2}, \quad S_{13} = S_{23} = \ln\binom{N}{N/2-1}$$

and $\frac{1}{2}(S_{13} + S_{23} - S_{123} - S_3)$ is as above.

Entropy minimizing properties of Slater determinants

Let $\rho_{1\ldots N}$ be an N-particle fermionic density matrix. Let $\rho_1 = \mathrm{Tr}_{2\ldots N}\rho_{1\ldots N}$ denote its single particle density matrix. Let $\rho_{12} = \mathrm{Tr}_{3\ldots N}\rho_{1\ldots N}$ denote its two particle density matrix.

It is well-known that

$$\{\rho\ :\ \rho_1 = \mathrm{Tr}_{2\ldots N}(\rho_{1\ldots N}) \text{ with } \rho_{1\ldots N} \text{ fermionic}\} = \{\rho\ :\ \rho \leq \tfrac{1}{N}\mathbb{1}\}\ .$$

The extreme points of this convex set are the normalized projections onto N-dimensional subspaces, which are precisely the one-particle reduced density matrices of N-particle Slater determinants. Hence $S(\rho_1) \geq \ln N$ with equality if and only if ρ_1 comes from an N-particle Slater determinant.

Entropy minimizing properties of Slater determinants continued

No such simple description of the set of all density matrices on $\mathcal{H} \otimes \mathcal{H}$ of the form $\rho_{12} = \mathrm{Tr}_{3\ldots N}(\rho_{1\ldots N})$ is known. Yang proved a sharp upper bound on the largest eigenvalue of ρ_{12} depending on the dimension of $\mathcal{H}$. As the dimension increases, this approaches $N\binom{N}{2}^{-1} = 2/(N-1)$. If there were $\mathcal{O}(N)$ eigenvalues nearly as large as this, one might have $S(\rho_{12})$ close to $\ln N$. However, producing the single large eigenvalue in the Yang state produces many more very small eigenvalues, resulting in a large entropy. The Yang pairing state (which is at the basis of superconductivity and superfluidity) has an entropy of order $\ln(\dim(\mathcal{H}))$.

A Slater ρ_{12} has $\binom{N}{2}$ eigenvalues $1/\binom{N}{2}$, and thus $S_{12} = \ln\binom{N}{2} \approx 2\ln N - \ln(2)$.

Theorem #3

We *believe* that a Slater minimizes S_{12}. What we *can prove* is:

Theorem 3. *The 2-particle reduced density matrix of any N-particle fermionic state satisfies*

$$S(\rho_{12}) \geq 2\ln N + \mathcal{O}(1) \ .$$

and, therefore, a Slater is at least asymptotically close to the minimum.

FINAL EXAM: **Prove the conjecture**

Acknowledgement: Many thanks to Eric Carlen of Rutgers University for his very valuable criticisms and suggestions that made these slides possible.

Thanks, also, to Chaney Lin of Princeton University for carefully proofreading these slides.

THANKS FOR LISTENING!