Quantum Information Theory

A crash course for Modern Physics

Gemma De las Cuevas

November 28, 2022

My goal today:

To give you some basic notions of quantum information.

My goal is not to advertise the research that is being done here, or to advertise the field of quantum information theory more generally. Instead, I aim at sharing with you some basic concepts.

General sources on Quantum Information Theory:

- Nielsen and Chuang. The classical source. About 20 years old.
- John Preskill's notes. Very well explained. About 20 years old.
- John Watrous's Quantum Information book. Very mathematical and valuable.
- PI repository, for all kinds of talks, also of introductory courses pirsa.org
- David Deutsch's online lectures are excellent (although the quality of the video is dreadful).

More info, bound to our space and time:

- You can learn Quantum Info at the Lecture "Theoretical Quantum Information" given in the Winter term (usually by Hans Briegel, Wolfgang Dür or myself).

Contents

1 What is Quantum Information and Quantum Computation?

Quantum information theory and quantum computation is the study of the *information processing tasks* that can be accomplished using quantum mechanical systems. Since classical physics is a special case of quantum physics, the general idea is that one can do *more things* (i.e. *new things*) with quantum resources than with classical ones. New things could mean genuinely new things (like teleportation, as we will see today), or it could means faster things (like solving problems faster; some examples below).

For example, in *quantum information theory*, quantum systems can show a very strong form of correlation called *entanglement*. This is a purely quantum phenomenon, which is stronger than anything that can be accomplished with classical systems. (This can be proven, and it has also been experimentally demonstrated that entanglement are not classical correlations in some hidden form—and awarded with the Nobel Prize in Physics in 2022). With entanglement one can do some surprising things, that is, some new protocols which are impossible classically. One example is *teleportation*, which we will see today. Another is *superdense coding*, which we will not see today.

In *quantum computation* the idea is that quantum algorithms will be able to solve tasks more rapidly than classical algorithms. The most famous example is the factoring algorithm, i.e. the algorithm that solves the factoring problem. The latter is defined as: given a number, find its factors. (Or a suitable decision problem version thereof). For example, given 21, the algorithm should return 3 and 7. The best known classical algorithms for factoring scale very badly with the system size, that is, their running time grows exponentially with the size of the input (i.e. how many digits the number to be factored has). Yet, there is a quantum algorithm, called *Shor's algorithm* (named after Peter Shor, who invented it in 1995) which runs only *polynomially* with the size of the input. Polynomially is much faster than exponentially, so if one could build a quantum computer that runs Shor's algorithm, one could solve factoring really fast. (See my course on Computational complexity theory, called Mathematics and Computation). This is important because factoring is the basis of many encryption systems that are used nowadays, e.g. on the internet.

More generally, however, it is unclear where precisely the power of quantum computation comes from. It is generally a very subtle issue to prove that a quantum algorithm really performs better than a classical algorithm.

Let's start by defining the main players of quantum information: a qubit and multiple qubits.

2 One and multiple qubits

2.1. What is a qubit?

The bit. The bit is the fundamental concept of classical information and classical computation. It is the simplest non-trivial variable: a degree of freedom with two possible values, which are usually taken to be 0 or 1, or false and true. Variables with two possible values are called Boolean variables.

A degree of freedom in quantum physics: an observable. In quantum physics the closest thing to a degree of freedom is called an observable, O. The labels of the variable correspond to the spectrum of O, i.e. the eigenvalues of O. Each label is a real number. (This is guaranteed by the fact that we will require observables O to be represented by *Hermitian matrices*, i.e. $O = O^\dagger$; as you know, Hermitian matrices have real eigenvalues). If an observable O is represented by an $n \times n$ Hermitian matrix, it can have at most n different eigenvalues. So it would describe a quantum variable with at most n different labels. This is a discrete and finite number of eigenvalues.

A qubit. A qubit is an abstraction of a physical system, each of whose non-trivial observables is Boolean, that is, an observable with 2 different eigenvalues. These observables can be of a very different physical nature (e.g. they could describe the polarization of a photon, or the energy level of the last electron of an atom; some examples below). Since the observable has two different eigenvalues, it must be represented by a Hermitian matrix of size at least 2×2. *Every other Hermitian matrix of the same dimension also represents an observable of the system.* There is no classical analogue of the previous statement.

The state of the system. Mathematically, the state of the system is the function that specifies the expectation value function of any given observable. Thus it is a function from observables to real numbers.[1] Just as a classical bit has a state – either 0 or 1 – a qubit also has a state.

The state of the system (for a pure state) is denoted in Dirac's notation by a so-called ket $|\psi\rangle$. For a qubit, it is given by a normalised vector in a two-dimensional complex vector space, $\mathbb{C}^2$. If our system is d-dimensional, i.e. it is a *qudit*, then $|\psi\rangle \in \mathbb{C}^d$. The important thing this object does is to specify the expectation value function for any observable:

$$E_\psi : \mathcal{M}_d \rightarrow \mathbb{R} \tag{1}$$

$$O \mapsto \mathrm{tr}(|\psi\rangle\langle\psi|O) = \langle\psi|O|\psi\rangle \tag{2}$$

where $\mathcal{M}_d$ is the set of complex matrices of size $d \times d$. Here $\langle\psi|$ is the *dual state* of $|\psi\rangle$ (called a bra, so that together they form a bra(c)ket, $\langle\psi|\psi\rangle$.) If $|\psi\rangle$ is expressed in an orthonormal basis (see below), $|\psi\rangle = \sum_j c_j |j\rangle$, where c_j are complex numbers, then $\langle\psi| = \sum_j \bar{c}_j \langle j|$, where $\bar{c}_j$ is the complex conjugate of c_j.

Note that by construction 'the expectation value function' E_ψ is positive, i.e. it maps positive semidefinite matrices to positive numbers, i.e. $E_\psi(M) \geq 0$ if M is positive semidefinite. A matrix M is positive semidefinite if it is Hermitian and has nonnegative eigenvalues. Positive semidefinite matrices of size $d \times d$ form a cone in $\mathcal{M}_d$.

Additionally, E_ψ satisfies a normalisation condition, namely $E_\psi(I) = 1$ where I is the identity matrix.

Both the positivity condition and the normalisation condition stem from the wish of having well-defined probabilities (i.e. nonnegative numbers that sum to 1). The probabilities will thus be as usual (i.e. as in the classical case) but the mathematical objects that give rise to them (i.e. the wavefunction ψ and the projectors, see below) will be different than the classical ones.

For qubits, we often use a basis of the vector space $\mathbb{C}^2$ called the *computational basis*, denoted $|0\rangle, |1\rangle$. This is an orthonormal basis, i.e. $\langle i|j\rangle = \delta_{i,j}$. A wavefunction, or state of the system, is mathematically an element

[1] Note that I am only defining the state of the system mathematically. The physical interpretation of the state of the system (or wavefunction) ψ is much debated, and I am personally very unclear about it.

in this vector space:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \quad \text{where } \alpha, \beta \in \mathbb{C} \text{ such that } |\alpha|^2 + |\beta|^2 = 1. \tag{3}$$

Measurements. We can examine a bit to determine whether it is in the state 0 or 1. Computers do this all the time when they retrieve the contents of their memory. Rather remarkably, we cannot examine a qubit to determine its quantum state, that is, the values of α and β. Instead, quantum mechanics tells us that we can only acquire much more restricted information about the quantum state.

Namely, consider the observable σ_z, whose eigenstates are precisely denoted $|0\rangle, |1\rangle$ and whose associated eigenvalues are $+1, -1$. That is,

$$\sigma_z = |0\rangle\langle 0| - |1\rangle\langle 1|. \tag{4}$$

Then, using **Born's rule**, if the state of the system is $|\psi\rangle$ and we measure σ_z, the *probability* to obtain $+1$ is given by

$$\langle\psi|0\rangle\langle 0|\psi\rangle = |\alpha|^2 \tag{5}$$

and the probability to obtain -1 is

$$\langle\psi|1\rangle\langle 1|\psi\rangle = |\beta|^2 \tag{6}$$

Therefore $|\alpha|^2 + |\beta|^2$ needs to be 1. Therefore, mathematically, the state of a qubit $|\psi\rangle$ is represented by a unit vector in a 2-dimensional complex vector space.[2]

This can also be seen as follows. 'Doing nothing' must correspond to measuring the 'trivial observable', namely the identity. For a qubit, this is the identity matrix of size 2×2. For a *qudit* (a d-dimensional version of the qubit), this is the identity matrix of size $d \times d$. The identity is a trivial observable because it only has one non-degenerate eigenvalue. According to Born's rule, the probability to obtain the result 1 (i.e. the eigenvalue associated to any eigenvector of I) is given by $\langle\psi|I|\psi\rangle = 1$ which for a qubit in state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ yields precisely the normalization condition $|\alpha|^2 + |\beta|^2 = 1$.

When a system is measured, the state **collapses** to the the eigenstate whose eigenvalue has been obtained. In the example above, where we measured the observable σ_z of a qubit, if we obtain $+1$ then the resulting state is $|0\rangle\langle 0|$ and if we obtain -1 the resulting state is $|1\rangle\langle 1|$.

The most important computational rule is the following:

[2] To be more precise, we also ignore the overall phase of this vector. That is, $|\psi\rangle$ and $e^{i\varphi}|\psi\rangle$ give rise to the same outcome probabilities for all observables. For this reason, $|\psi\rangle$ is sometimes called a *ray* in a vector space.

Box 1: Key rule

If the system is in state $|\psi\rangle$, and you measure an observable O with spectral decomposition

$$O = \sum_j \lambda_j M_j$$

(where λ_j are real eigenvalues and M_j are orthogonal projectors), the probability to obtain the result labeled by j is given by

$$p_j := \mathrm{tr}(M_j |\psi\rangle\langle\psi|), \qquad (7)$$

so that the expectation value of observable O in state $|\psi\rangle$ is given by

$$E_\psi = \mathrm{tr}(O|\psi\rangle\langle\psi|) = \sum_j \lambda_j p_j.$$

Note that if eigenvalue λ_j is not degenerate, then M_j is a rank-1 projector, so that it can be written as $M_j = |\phi_j\rangle\langle\phi_j|$. If λ_j is, e.g., two-fold degenerate, then M_j is a rank-2 projector, i.e. it can be written as $M_j = |\phi_{j,1}\rangle\langle\phi_{j,1}| + |\phi_{j,2}\rangle\langle\phi_{j,2}|$, where $|\phi_{j,i}\rangle$ are orthonormal (but note that $UM_jU^\dagger$, for any unitary U, provides another spectral decomposition of M_j). And so on for higher degenerations of an eigenvalue.

⚠**Warning** ⚠The previous Box implies that: If the system is in state $|\psi\rangle$ and one measures observable O, the resulting state is *not* $O|\psi\rangle$. The resulting state is a projection of $|\psi\rangle$ into one of the eigenstates of O, given by the corresponding probability of each measurement outcome. The transformation $|\psi\rangle \mapsto O|\psi\rangle$ describes a linear transformation of the state of the system that does (generally) *not* describe a measurement.

Physical content of the collapse of the wavefunction. Physically this collapse is not well-understood. Some say that there is a branching into different universes (this is the Everettian or multiverse interpretation of quantum mechanics). Others say that this is not a problem (e.g. the Copenhagen view). Others say that measurement outcomes are not objective, but private experiences of the subject doing the measurement (e.g. QBism). I do not understand what happens, physically, during the collapse of the wavefunction. And I find it unbelievable that we disagree on whether we live in a multiverse, or a non-realist local universe, or a non-local realist universe! If we live in a multiverse, there are infinitely many copies of myself despite the fact that I only (seem to) experience one self. I think this is a very important question, and currently there is very interesting research in this direction.

This collapse is also problematic regardless of the ontology, for the following reason. The collapse rule can only be applied once the system is distinguished from the observer, where the observer is the one executing the measurement. This distinction, called the **Heisenberg cut**, seems very artificial because we are all part of the same physical world, and thus the system and observer seem to deserve the same treatment. One can come up with Gedankenexperiments that exploit the arbitrariness of the placement of the Heisenberg cut in order to produce inconsistencies. These usually go under the name of **Wigner's friend** thought experiments, and there's currently very interesting research on this front. So, regardless of the problem of 'interpretations of quantum mechanics' (that is, their ontology), quantum theory leads to inconsistencies.

The description of the quantum state versus what is accessible. There is a *fundamental tension* between the abstract state of the system (i.e. $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$) and what we can observe, and hence the information that can be obtained (Fig. 1). Namely, to find out α, β we would need infinitely many identical copies of $|\psi\rangle$, we would need to measure each copy in the computational basis, and we would need to collect statistics of the relative frequencies of results 0 and 1. This way we could estimate the probabilities $|\alpha|^2$ and $|\beta|^2$ in the limit $N \to \infty$ of infinitely many measurements. To find out α and β (instead of $|\alpha|$ and $|\beta|$) we would need to repeat the same process in another basis, e.g. in the σ_x basis (see Eq. (42)). This is in agreement with the fact that there is in fact infinite information in $\alpha, \beta \in \mathbb{C}$ (because the reals contain the irrationals, which generally require an infinite amount of information for the description of one of their numbers), but this cannot be accessed with any finite number of measurements.

Figure 1: On the left hand there is quantum physics, i.e. what we observe, which are events and their relative frequencies. On the right hand side there is the *theory* of quantum physics, including the description of quantum systems with non-commutative spaces (such as matrices), the composition rule given by the tensor product, positivity structures (central for the description of quantum systems, which interact in a very interesting way with the tensor product), complex numbers (which involve limits, which in my opinion are not physical), and probabilities (which involve the limit of infinitely many repetitions). The shadow of the theory of quantum physics (i.e. the left hand side) is very special, and different from the shadow of the theory of classical physics. Art by Kumi Yamashita.

In fact, the **Holevo bound** dashes many hopes in this direction, as it puts an upper limit on how much information can be contained in a quantum system. Essentially it says that one qubit can contain at most one bit of information; more precisely, that n qubits can communicate at most n bits of decodable information.

More generally, infinity is not a physical notion (as far as I am concerned), so claims involving exact probabilities cannot be operational.

Expressing lack of knowledge with density matrices. So far we have assumed that we have perfect knowledge of the state of the system, which is represented by the state $|\psi\rangle$, called a *pure state*. If we are unsure about whether the system is in state $|\psi_1\rangle$ or $|\psi_2\rangle$, we need to use the **density matrix** formalism. More precisely, if we want to describe the fact that the system is in state $|\psi_1\rangle$ with probability p_1 and in state $|\psi_2\rangle$ with probability p_2, we use the density matrix

$$\rho = p_1|\psi_1\rangle\langle\psi_1| + p_2|\psi_2\rangle\langle\psi_2| \tag{8}$$

Note that this is a convex combination of the *projectors* onto the corresponding states $|\psi_1\rangle$ and $|\psi_2\rangle$. This convex combination is called an incoherent mixture. This is very different from a coherent mixture, which is a convex combination of the vectors (instead of their projectors),

$$p_1|\psi_1\rangle + p_2|\psi_2\rangle =: |\phi\rangle \tag{9}$$

and which describes *another* pure state $|\phi\rangle$.

So, *the lack of knowledge of the state of a system is expressed as a convex combination of the projectors onto the corresponding states.* This mixture results in a density matrix, which mathematically is a positive semidefinite matrix of trace 1. Positive semidefinite means Hermitian with nonnegative eigenvalues. This density matrix is usually denoted ρ, and their conditions are denoted $\rho \geqslant 0$ and $\mathrm{tr}(\rho) = 1$. *Any density matrix describes a possible state of the system.*

If the density matrix has rank one, i.e. if $\rho = |\psi\rangle\langle\psi|$ for some $|\psi\rangle$, then we are describing a pure state, i.e. we have perfect knowledge of the system. If the density matrix is maximally mixed, i.e. if $\rho = I/d$, where I is the identity matrix of size $d \times d$, we have maximal lack of knowledge of the state of the system. So the density matrix formalism is more expressive than the pure state formalism; the former contains the latter as a special case. Pure states are of zero measure—they are an idealisation. Experimentally, only density matrices are relevant.

The lack of knowledge of the state of the system, expressed as incoherent mixtures, can give rise to *classical correlations*. E.g. if I always wear socks so that my left sock is of the same color as my right sock, there is a classical correlation between the color of my two socks. (And there's nothing mysterious about that). Coherent mixtures can lead to *quantum correlations*, i.e. entanglement. This can be pretty mysterious indeed. More on this in the next section.

Regarding the key computational rule (Box 1): The probability to obtain outcome j given that the state of the system is ρ is given by $\mathrm{tr}(\rho M_j)$, as we expect when comparing it with (7), and the expectation value of observable O if the system is in state ρ is given by $\mathrm{tr}(O\rho)$.

Physical realisations of a qubit So far we have talked about the abstract mathematical description of the qubit, on which we will focus today. This abstract notion of a qubit can be realized on various physical systems. Some important physical realizations of a qubit are:

- as the two distinct polarization states of a photon
- as the alignment of a nuclear spin in a magnetic field (the "spin")
- as two states (e.g. ground and excited) of an electron orbiting an atom. E.g. the ground state would correspond to $|0\rangle$ and the first excited state to $|1\rangle$.

You may know more about this than I do! E.g. via the other lectures of Modern Physics – Innsbruck has a lot to say about this :)

2.2. Multiple qubits

To establish a theory, it is equally important to specify

(a) how to describe single entities (a qubit, see above), and

(b) how to compose these entities to obtain multiple entities (next)

(And it is a misconception of reductionism to overestimate the value of (a)).
In order to describe composite systems, we will use the following **consistency rules**:

(1) The whole system must admit a quantum mechanical description, i.e. the rules above (generalized to d level systems) must apply.

(2) If we ignore part of the system, the remaining subsystem must obey the rules mentioned above.

We will now consider the state of two qubits, also denoted $|\psi\rangle$. We will apply the two consistency rules and will see that we can go pretty far.

(1) **The two qubits must admit a quantum mechanical description**. So imagine that we want to measure the observable σ_z (defined in Eq. (4)) in the first qubit, and the observable σ_z in the second qubit simultaneously. That is, measure the overall observable $\sigma_z \otimes \sigma_z$, since the corresponding vector spaces are composed with a tensor product $\otimes$:

$$\sigma_z \otimes \sigma_z = |00\rangle\langle 00| - |01\rangle\langle 01| - |10\rangle\langle 10| + |11\rangle\langle 11| = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & 1 \end{pmatrix} \tag{10}$$

where $|ij\rangle$ is shorthand for $|i\rangle \otimes |j\rangle$, and where the matrix elements are expressed in the computational basis, $\{|ij\rangle \mid i \in \{0,1\}, j \in \{0,1\}\}$.

In order to compute the probabilities of measurement outcomes, we apply the key rule (Box 1) to the two-qubit system. Observable $\sigma_z \otimes \sigma_z$ has two eigenvalues, +1 and −1, each of which is two-fold degenerate,

$$\begin{aligned} \sigma_z \otimes \sigma_z = M_1 - M_2 \quad \text{where } M_1 &= |00\rangle\langle 00| + |11\rangle\langle 11| \\ M_2 &= |01\rangle\langle 01| + |10\rangle\langle 10| \end{aligned} \tag{11}$$

M_1 is a projection to the even parity subspace, and M_2 a projection to the odd parity subspace. Now we need to express $|\psi\rangle$ in an eigenbasis of $O = \sigma_z \otimes \sigma_z$[3]

$$|\psi\rangle = \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle. \tag{12}$$

According to the key rule, we will obtain either +1 or −1 out of this measurement, and the probability to obtain +1 is given by

$$\langle\psi|M_1|\psi\rangle = |\alpha_{00}|^2 + |\alpha_{11}|^2 \tag{13}$$

whereas the probability to obtain −1 is given by

$$\langle\psi|M_2|\psi\rangle = |\alpha_{01}|^2 + |\alpha_{10}|^2 \tag{14}$$

[3] Note that any unitary rotation on M_1 and on M_2, i.e. $U_1 M_1 U_1^\dagger$ and $U_2 M_2 U_2^\dagger$, also results in an eigenbasis of O.

(2) Instead, assume that **ignore the first qubit** and only measure the second qubit; say, we measure observable σ_z in the second qubit. This is equivalent to measuring the observable $I \otimes \sigma_z$, where I is the identity observable, i.e. the trivial observable because it only has one eigenvalue:

$$I \otimes \sigma_z = N_1 - N_2 \text{ where } N_1 = |00\rangle\langle 00| + |10\rangle\langle 10| \qquad (15)$$
$$N_2 = |01\rangle\langle 01| + |11\rangle\langle 11|$$

This observable also has two eigenvalues, $+1$ and -1, but their eigenspaces are different than before. In particular, N_1 is the projection onto the state 0 for the second qubit and anything on the first qubit, and N_2 is the projection onto the state 1 for the second qubit and anything on the first.

Considering the state of Eq. (12), the probability to obtain -1 is given by

$$p_{-1} = \langle\psi|N_2|\psi\rangle = |\alpha_{01}|^2 + |\alpha_{11}|^2 \qquad (16)$$

and similarly for the probability to obtain $+1$.

This is equivalent to first computing the state of the second qubit, denoted ρ_B, by applying the **partial trace** over system A, and then applying the key rule (Box 1) just to the second system. The partial trace mathematically represents the idea of **ignoring a system**. Namely we first compute the state of the second qubit:

$$\rho_B := \mathrm{tr}_A(|\psi\rangle\langle\psi|) \qquad (17)$$

[The partial trace literally means the trace over part of the system. Namely one can think of the trace of a composite system as the trace of each of its subsystems, e.g. $\mathrm{tr} = \mathrm{tr}_A \mathrm{tr}_B$. That is, $\mathrm{tr}(P) = \sum_{i,j}\langle i,j|P|i,j\rangle$; the sum over i is the trace over subsystem A, and the sum over j the trace over subsystem B.] For the state of $|\psi\rangle$ (Eq. (12)) this gives

$$\rho_B = \sum_{j,l}\left(\sum_i \alpha_{ij}\bar{\alpha}_{il}\right)|j\rangle\langle l| = \begin{pmatrix} |\alpha_{00}|^2 + |\alpha_{10}|^2 & \alpha_{00}\bar{\alpha}_{01} + \alpha_{10}\bar{\alpha}_{11} \\ \alpha_{01}\bar{\alpha}_{00} + \alpha_{11}\bar{\alpha}_{10} & |\alpha_{01}|^2 + |\alpha_{11}|^2 \end{pmatrix} \qquad (18)$$

We now measure σ_z (Eq. (4)) on this state of the second qubit. The probability to obtain result -1 is

$$p_{-1} = \mathrm{tr}(\rho_B|1\rangle\langle 1|) = |\alpha_{01}|^2 + |\alpha_{11}|^2 \qquad (19)$$

which is the same as Eq. (16).

In summary: ignoring part of the system is mathematically accomplished by taking the partial trace over that system.

Entanglement. Two qubit systems can exhibit entanglement. This a form of correlations which is uniquely quantum; it cannot happen in classical systems. Mathematically, a pure state $|\psi\rangle$ is entangled if it *cannot* be written as $|\psi\rangle = |v_1\rangle \otimes |v_2\rangle$ for any states $|v_1\rangle$, $|v_2\rangle$. That is, entangled states always have at least two terms in the sum, $|\psi\rangle = \sum_j |v_{1,j}\rangle \otimes |v_{2,j}\rangle$.[4]

[4] Note that entanglement is defined as a negation. This is the beginning of a very long story around the difficulty of characterising entanglement.

A famous example of an entangled state is given by the **Bell state** (or: EPR pair)

$$|\Phi^+\rangle := \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{20}$$

Note that it is a two-qubit state. If we measure σ_z in the first qubit, we obtain

- $+1$ with probability $1/2$, leaving the post-measurement state in $|\phi\rangle = |00\rangle$, and
- -1 with probability $1/2$, leaving the post-measurement state in $|\phi\rangle = |11\rangle$.

As a result, a measurement of σ_z of the second qubit always gives *the same result* as the measurement of the first qubit. That is, the measurement outcomes of the first and second qubit are perfectly **correlated**. (Namely only the results $+1, +1$ or $-1, -1$ can happen. $+1, -1$ or $-1, +1$ cannot happen).[5] As soon as the first party (called Alice) measures her qubit in the σ_z basis and obtains an outcome, she immediately knows what result the second party (called Bob) would obtain if he measured in σ_z. This can be used for teleportation, as we will see below.

These correlations have been the subject of intense interest ever since a famous paper by Einstein, Podolsky and Rosen in 1935 (EPR), in which they first pointed out the strange properties of states like the Bell state. EPR's insights were taken up and greatly improved by John Bell, who in the 1960s proved an amazing result: the measurement correlations in the Bell state are stronger than the correlations that can exist between classical systems, including potential 'hidden variables'. This is the famous **Bell's theorem**, which has been experimentally corroborated multiple times. The Nobel Prize in Physics in 2022 was awarded for the first experimental demonstration of the violation of Bell's inequalities. These results were the first sign that quantum mechanics allows information processing beyond what is possible in the classical world. And these results shed amazing light on the nature of our world! Some call it experimental metaphysics – I like this name.

Bell basis. An important set of states for two qubits is the Bell basis:

$$|\Phi^+\rangle := \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{21}$$

$$|\Phi^-\rangle := \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \tag{22}$$

$$|\Psi^+\rangle := \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \tag{23}$$

$$|\Psi^-\rangle := \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \tag{24}$$

This is an orthonormal basis of $\mathbb{C}^2 \otimes \mathbb{C}^2$. But it is not a product basis; on the contrary, its elements are *maximally entangled*.

Very surprising relation between the parts and the whole. A Bell state (namely, any of the elements of the Bell basis) exhibits a very surprising relation between the parts and the whole. To explain this, imagine that two parties, Alice and Bob, share the state $|\Phi^+\rangle$. Because $|\Phi^+\rangle$ is a pure state, they have perfect knowledge of the total state they share, i.e. they have *perfect knowledge of the whole.* Now, let us compute the state of Alice's

[5] Two systems are correlated if knowing something about one can be used to infer something about the second. For example, if I always wear socks of the same color and you know that my left sock is red, you can infer that my right sock is also red. I.e. the color of my socks is correlated. Correlations can be classical or quantum.

part of the system:

$$\begin{aligned}\rho_A &= \mathrm{tr}_B(|\Phi^+\rangle\langle\Phi^+|) && (25)\\ &= \frac{1}{2}\mathrm{tr}_A(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|) && (26)\\ &= \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) && (27)\\ &= \frac{1}{2}I && (28)\end{aligned}$$

It is the maximally mixed state! That is, Alice has maximal lack of knowledge about her state. The same calculation shows that Bob's state ρ_B is also the normalised identity, i.e. the maximally entangled state. So Alice *has maximal lack of knowledge of her part and maximal knowledge of the whole.* The same holds for Bob. This is very surprising! Usually we have good knowledge of the parts, and less knowledge of the whole (e.g. when we cannot scale a description). Here the relation is maximally inverted. This is a hallmark of entanglement—in fact, this relation between the parts and the whole can be used to define maximal entanglement.

The purification theorem. Every mixed state can be purified, that is, expressed as as a subsystem of pure state. For every (positive semidefinite) ρ_S, there exists a pure state on the system $|\psi\rangle_{SE}$ (where S stands for system and E for extension) such that $\rho_S = \mathrm{tr}_E|\psi\rangle_{SE}\langle\psi|$. The pure state $|\psi\rangle_{SE}$ (or: purification) is highly non-unique, as $I_S \otimes U_E|\psi\rangle_{SE}$, for any unitary on the extension U_E, is another valid purification.[6] Additionally, the dimension of E is at least the rank of ρ_S.

Mathematically, the statement is almost trivial (at least in finite dimensions), as it follows from the spectral decomposition of ρ_S. Yet, physically, it is very interesting. On the one hand, it says that the lack of knowledge expressed in ρ can be attributed to it being part of a larger pure state, that is, entangled with an extension (or environment). On the other hand, recall that a mixed state ρ describes the 'actual' case, whereas a pure state $|\psi\rangle$ describes the ideal case. The purification theorem thus expresses *the relation between the actual and the ideal as a relation between the parts and the whole*. I don't think there's an analogue to this theorem in other parts of physics which also pivot on ideal descriptions.

The exponential scaling of the description with Hilbert spaces. The Hilbert space of n qubits is given by the n-fold tensor product of the Hilbert space of one qubit:

$$\mathcal{H}_{\text{total}} = \mathcal{H}^{(1)} \otimes \mathcal{H}^{(2)} \otimes \cdots \otimes \mathcal{H}^{(n)} \tag{29}$$

If each local Hilbert space has dimension d, the total Hilbert space has dimension d^n, that is, it **grows exponentially with the system size**. If $d = 2$, for $n = 240$ qubits, the Hilbert space dimension is $2^{240} \approx 10^{80}$, which is the estimated number of atoms in the observable Universe. The description with Hilbert spaces is thus *not scalable*, and can be effectively only used for a few qubits. Beyond that, one needs other, scalable tools to describe quantum many-body systems. This is the starting point of the research program of tensor networks (and others), which is very big both numerically and analytically, with many applications across physics, maths, machine learning and beyond (and on which I've been working on in recent years).

[6] In fact, an isometry on the extension results in a valid purification too. Isometries are like unitaries between spaces of different dimensions, i.e. $V : \mathbb{C}^d \to \mathbb{C}^s$ with $s \geq d$ such that $V^\dagger V = I_d$.

3 Teleportation

Let us see one of the surprising things that one can do with entangled states.

Imagine that we have two parties, called Alice and Bob. Alice has a qubit, which we call system A', in some given state $|\chi\rangle_{A'} = \alpha|0\rangle + \beta|1\rangle$. This state may be unknown to Alice. However, Alice wishes to send *the state* of her qubit A' to Bob. To this end, Alice and Bob will use the following resources: first, the Bell state $|\Phi^+\rangle$ between Alice and Bob, and second, classical communication from Alice to Bob. By applying the teleportation protocol (to be described next), at the end the state of Alice's qubit $|\chi\rangle_{A'}$ will "appear" in Bob's qubit (Fig. 2). Let us explain this.

First, Alice and Bob share the entangled state

$$|\Phi^+\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle_{AB} + |11\rangle_{AB}). \tag{30}$$

So that the overall initial state is

$$\begin{align} |\psi\rangle_{A'AB} &= |\chi\rangle_{A'} \otimes |\Phi^+\rangle_{AB} \tag{31} \\ &= \frac{1}{\sqrt{2}}[\alpha|000\rangle + \alpha|011\rangle + \beta|100\rangle + \beta|111\rangle] \nonumber \end{align}$$

where I have ommitted the subscripts in the second line to keep the notation simple. For simplicity sometimes I may also denote $|\psi\rangle_{A'AB}$ by $|\psi\rangle$.

Now Alice performs a measurement to her two qubits, A' and A, **in the Bell basis**. To compute the outcome probabilities on her systems $A'A$, we first need to take the partial trace over system B:

$$\begin{align} \rho_{A'A} &= \mathrm{tr}_B(|\psi\rangle_{A'AB}\langle\psi|) \tag{32} \\ &= \frac{1}{2}[|\alpha|^2(|00\rangle\langle 00| + |01\rangle\langle 01|) + |\beta|^2(|10\rangle\langle 10| + |11\rangle\langle 11|) + \nonumber \\ &\quad \alpha\bar{\beta}(|00\rangle\langle 10| + |01\rangle\langle 11|) + \bar{\alpha}\beta(|10\rangle\langle 00| + |11\rangle\langle 01|)] \tag{33} \end{align}$$

Now she obtains the eigenvalue associated to eigenstates $|\Phi^+\rangle$, $|\Phi^-\rangle$,$|\Psi^+\rangle$ and $|\Psi^-\rangle$ with probabilities

$$\begin{align} p_i &:= \mathrm{tr}(\rho_{A'A}|\Phi^+\rangle_{A'A}\langle\Phi^+|) = \frac{1}{4}(|\alpha|^2 + |\beta|^2) = 1/4 \tag{34} \\ p_{ii} &:= \mathrm{tr}(\rho_{A'A}|\Phi^-\rangle_{A'A}\langle\Phi^-|) = 1/4 \tag{35} \\ p_{iii} &:= \mathrm{tr}(\rho_{A'A}|\Psi^+\rangle_{A'A}\langle\Psi^+|) = 1/4 \tag{36} \\ p_{iv} &:= \mathrm{tr}(\rho_{A'A}|\Psi^-\rangle_{A'A}\langle\Psi^-|) = 1/4 \tag{37} \end{align}$$

respectively. The interesting thing is what happens to Bob's state after the collapse of Alice's qubits:

(i) If she obtains the eigenvalue associated to eigenstate $|\Phi^+\rangle$, the resulting state in Bob's qubit is

$$\frac{((\langle\Phi^+|_{A'A} \otimes I_B)}{p_i}|\psi\rangle_{A'AB} = \alpha|0\rangle_B + \beta|1\rangle_B = |\chi\rangle_B \tag{38}$$

That is, Bob's state is precisely the original state of Alice's qubits A'! The state of Alice's qubit A' has been *teleported* to Bob's qubit state.

(ii) If she obtains the eigenvalue associated to eigenstate $|\Phi^-\rangle$, the resulting state in Bob's qubit is

$$\frac{(\langle\Phi^-|_{A'A}\otimes I_B)}{p_{ii}}|\psi\rangle_{A'AB} = \alpha|0\rangle_B - \beta|1\rangle_B = \sigma_z|\chi\rangle_B \tag{39}$$

(iii) If she obtains the eigenvalue associated to eigenstate $|\Psi^+\rangle$ the resulting state in Bob's qubit is

$$\frac{(\langle\Psi^+|_{A'A}\otimes I_B)}{p_{iii}}|\psi\rangle_{A'AB} = \alpha|1\rangle_B + \beta|0\rangle_B = \sigma_x|\chi\rangle_B \tag{40}$$

(iv) If she obtains the eigenvalue associated to eigenstate $|\Psi^-\rangle$, the resulting state in Bob's qubit is

$$\frac{(\langle\Psi^-|_{A'A}\otimes I_B)}{p_{iv}}|\psi\rangle_{A'AB} = \alpha|1\rangle_B - \beta|0\rangle_B = i\sigma_y|\chi\rangle_B \tag{41}$$

Here we have used the so-called *Pauli matrices*

$$\sigma_x \;:=\; |0\rangle\langle 1| + |1\rangle\langle 0| = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{42}$$

$$\sigma_y \;:=\; -i|0\rangle\langle 1| + i|1\rangle\langle 0| = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{43}$$

$$\sigma_z \;:=\; |0\rangle\langle 0| - |1\rangle\langle 1| = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{44}$$

Now, what happens if Alice obtains results (ii), (iii) or (iv)? In this case, Bob's state is not correct, i.e. it is not $|\chi\rangle$ so the teleportation hasn't 'worked'. Alice needs to *send a classical message*—that is, call by phone, or send a whatsapp, or an email—to Bob and tell him what measurement result she obtained. This way, he will be able to correct his state so that it becomes $|\chi\rangle$.

Alice's message only needs to contain the information of whether she obtained (i), (ii), (iii), (iv). I.e. she only needs to send a number from 1 to 4 (because they have previously agreed on the labeling of the measurement outcomes). This is equivalent to sending two classical bits: for example, 00 for (i), 01 for (ii), 10 for (iii), and 11 for (iv).

When Bob gets this message, he corrects his state accordingly:

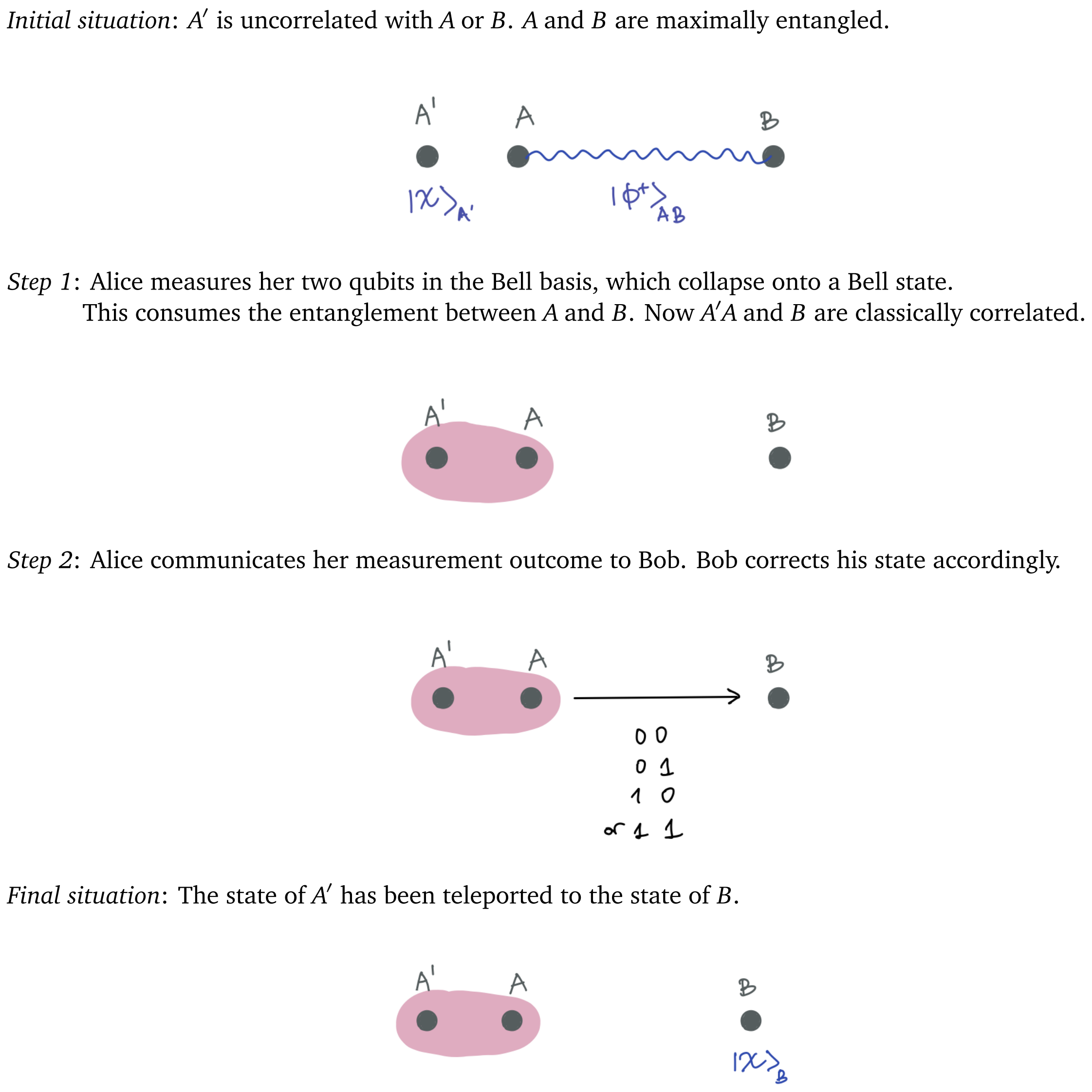
(i) If he gets 00, he applies the identity to his state (i.e. does nothing).

(ii) If he gets 01, he applies the σ_z to his state.

(iii) If he gets 10, he applies the σ_x to his state.

(iii) If he gets 11, he applies the σ_y to his state.

In all of the cases, the resulting state on Bob's side is $|\chi\rangle_B$ (or a global phase, such as -1, times $|\chi\rangle_B$; a global phase is irrelevant). This way Bob ends up having $|\chi\rangle$! Alice's state has been teleported to Bob's.



Initial situation: A' is uncorrelated with A or B. A and B are maximally entangled.

Step 1: Alice measures her two qubits in the Bell basis, which collapse onto a Bell state. This consumes the entanglement between A and B. Now $A'A$ and B are classically correlated.

Step 2: Alice communicates her measurement outcome to Bob. Bob corrects his state accordingly.

Final situation: The state of A' has been teleported to the state of B.

Figure 2: The teleportation protocol, by which the state of A' is teleported to the state of B. It consumes one maximally entangled state and requires the classical communication of two bits.

Some remarks:

- Alice and Bob do not need to meet during the protocol. They can be as far they want (e.g. one on Jupiter and the other in Innsbruck), as long as they share the entangled state, and Alice can call Bob to tell him the measurement result. In fact, they only need to have interacted in the past in order to create the entangled state.

- Alice and Bob do not know the state $|\chi\rangle$ at any stage of the protocol. Nonetheless, if they perform the protocol correctly, they can be sure that this unknown state has been teleported to Bob's qubit.

- Before Alice sends Bob the two classical bits of information, which tell him which correction operator he has to apply, Bob has *no idea* which state he has. Formally, Bob's state is **maximally mixed**, which precisely formalises the idea of his total lack of knowledge of the state. To see this, we use the formalism of density matrices encountered before. Before the two classical bits from Alice, Bob does not know which measurement outcome Alice got, so his density matrix is described by:

$$\begin{aligned}\rho &= \frac{1}{4}[|\chi\rangle\langle\chi| + \sigma_z|\chi\rangle\langle\chi|\sigma_z + \sigma_x|\chi\rangle\langle\chi|\sigma_x + \sigma_y|\chi\rangle\langle\chi|\sigma_y] \\ &= \frac{1}{4}\left\{\begin{pmatrix} |\alpha|^2 & \alpha\bar{\beta} \\ \bar{\alpha}\beta & |\beta|^2 \end{pmatrix} + \begin{pmatrix} |\alpha|^2 & -\alpha\bar{\beta} \\ -\bar{\alpha}\beta & |\beta|^2 \end{pmatrix} + \begin{pmatrix} |\beta|^2 & \bar{\alpha}\beta \\ \alpha\bar{\beta} & |\alpha|^2 \end{pmatrix} + \begin{pmatrix} |\beta|^2 & -\bar{\alpha}\beta \\ -\alpha\bar{\beta} & |\alpha|^2 \end{pmatrix}\right\} \\ &= \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad (45)\end{aligned}$$

  where the matrix entries are written in the computational basis $|0\rangle, |1\rangle$; specifically,

$$\rho = \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} \qquad \text{where } \rho_{i,j} = \langle i|\rho|j\rangle. \qquad (46)$$

  So Bob's state is proportional to the identity, i.e. it is totally mixed, i.e. he has absolutely no idea of which state he has!

- In this protocol **no matter is teleported**. So it is not like in Star Trek, where a body disappears here and appears elsewhere. It's only the *information* of Alice's state that is teleported. Qubits A' and B have some physical realisation (e.g. they are encoded in the state of an atom), and the physical implementation of these qubits hasn't moved.

- When Alice's measurement happens, Bob's state collapses instantaneously, **without any signal travelling from Alice to Bob**. Thus, it may appear that faster-than-light communication is at stake, which would violate the relativity principles. But there is no paradox, i.e. Alice **cannot transmit information faster than light**, because she needs to send classically her 2 bits to Bob so that he can "fix" the state. Before receiving Alice's two bits, Bob has no idea of the state of his qubit, as we have seen. And sending classical information is bound by the laws of relativity, i.e. it can travel at most at the speed of light.

- It may also appear that the state $|\chi\rangle$ is copied during the protocol. This would violate a very central result of quantum information called **the no-cloning theorem**, which states that there cannot exist a copy-machine of quantum states. I.e. unknown quantum states cannot generally be copied (in contrast to classical states, which can clearly be copied). However, **the state $|\chi\rangle$ is not copied** during the protocol. This is because when Alice measures her two qubits A', A, she *destroys* the state $|\chi\rangle$ in A'.