where $\boldsymbol{s} = (s_x, s_y, s_z)$ is a vector. The physical interpretation of $\boldsymbol{s}$ becomes evident from the following relation:

$$s_i = \langle \sigma_i \rangle = \mathrm{tr}(\sigma_i \rho). \tag{2.30}$$

The relation between these parameters and the parametrization in Eq. (2.18) is

$$\begin{aligned} \langle \sigma_x \rangle &= q + q^*, \\ \langle \sigma_y \rangle &= i(q - q^*), \\ \langle \sigma_z \rangle &= p_+ - p_-. \end{aligned}$$

Next we look at the purity of a qubit density matrix. From Eq. (2.29) one also readily finds that

$$\mathrm{tr}(\rho^2) = \frac{1}{2}(1 + \boldsymbol{s}^2). \tag{2.31}$$

Thus, due to Eq. (2.13), it also follows that

$$\boldsymbol{s}^2 = s_x^2 + s_y^2 + s_z^2 \leq 1. \tag{2.32}$$

When $\boldsymbol{s}^2 = 1$ we are in a pure state. In this case the vector $\boldsymbol{s}$ lays on the surface of the Bloch sphere. For mixed states $\boldsymbol{s}^2 < 1$ and the vector is inside the Bloch sphere. Thus, we see that the purity can be directly associated with the radius in Bloch's sphere. The smaller the radius, the more mixed is the state. In particular, the maximally disordered state occurs when $\boldsymbol{s} = 0$ and reads

$$\rho = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{2.33}$$

In this case the state lies in the center of the sphere. A graphical representation of pure and mixed states in the Bloch sphere is shown in Fig. 2.1.

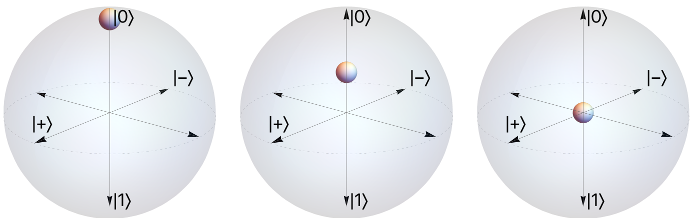

Figure 2.1: Examples of pure and mixed states in the z axis. Left: a pure state. Center: an arbitrary mixed state. Right: the maximally mixed state (2.33).

2.3 Composite systems and the almighty kron

So far we have considered only a single quantum system described by a basis $|i\rangle$. Now we turn to the question of how to describe mathematically a system composed of two or more sub-systems. Suppose we have two sub-systems, which we call A and B. They can be, for instance, two qubits: one on earth and the other on mars. We use a basis $|i\rangle_A$ to describe the sub-system A and a basis $|j\rangle_B$ to describe sub-system B. In general, each sub-system can have different dimensions, $d_A \neq d_B$.

If we now wish to describe the composite system A+B, we could use as a basis set, states labeled as $|i, j\rangle_{AB}$, where i is the quantum number of A and j is the quantum number of B. That makes sense: suppose A and B are spins which can be up or down. Then a state such as $|\uparrow, \downarrow\rangle_{AB}$ means the first is up and the second is down, and so on. But how do we operate with these states? That is, how do we construct operators which act on these states to produce new states?

The intuition is that A and B represent separate universes: things related to A have nothing to do with things related to B. After all, they can be on different planets. Thus, for instance, we know that for one qubit the operator σ_x flips the bit: $\sigma_x|0\rangle = |1\rangle$. Now suppose two qubits are in a state $|0, 0\rangle$. Then we expect that there should be an operator σ_x^A which flips the first qubit and an operator σ_x^B that flips only the second. That is,

$$\sigma_x^A|0, 0\rangle = |1, 0\rangle, \qquad \sigma_x^B|0, 0\rangle = |0, 1\rangle \tag{2.34}$$

The mathematical structure to do this is called the **tensor product** or **Kronecker product**. It is, in essence, a way to glue together two vector spaces to form a larger space. The tensor product between two states $|i\rangle_A$ and $|j\rangle_B$ is written as

$$|i, j\rangle_{AB} = |i\rangle_A \otimes |j\rangle_B. \tag{2.35}$$

The symbol $\otimes$ separates the two universes. Sometimes this is read as "i tens j" or "i kron j". I like the "kron" since it reminds me of a Transformers villain. Sometimes the notation $|i\rangle_A|j\rangle_B$ is also used for convenience, just to avoid using the symbol $\otimes$ over and over again. Let me summarize the many notations we use:

$$\boxed{|i, j\rangle_{AB} = |i\rangle_A \otimes |j\rangle_B = |i\rangle_A|j\rangle_B} \tag{2.36}$$

When is clear from the context, we also sometimes omit the suffix AB and write only $|i, j\rangle$.

Eq. (2.36) is still not very useful since we haven't specified how to operate on a tensor product of states. That is, we haven't yet specified what is the tensor structure of operators. In order to do that, we must have a rule for how objects behave when there is an $\otimes$ around. There is only one rule that you need to remember: *stuff to the left of $\otimes$ only interact with stuff to the left and stuff to the right only interact with stuff to the right.* We write this as

$$\boxed{(A \otimes B)(C \otimes D) = (AC) \otimes (BD),} \tag{2.37}$$

In this rule A, B, C and D can be arbitrary objects. For instance, this rule applies if they are all matrices. Or, they apply if you want to multiply a vector by a matrix. In that case we get instead

$$(A \otimes B)(|\psi\rangle \otimes |\phi\rangle) = (A|\psi\rangle) \otimes (B|\phi\rangle), \tag{2.38}$$

From this we now define operators which act only on A or act only on B as

$$\mathcal{O}_A \otimes 1_B = \text{an operator that acts only on space A} \tag{2.39}$$

$$1_A \otimes \mathcal{O}_B = \text{an operator that acts only on space B} \tag{2.40}$$

where 1_A means the identity operator on system A and similarly for 1_B.

For instance, going back to the example in Eq. (2.34), we can define the Pauli matrices for qubits A and B as

$$\sigma_x^A = \sigma_x \otimes 1_2, \qquad \sigma_x^B = 1_2 \otimes \sigma_x, \tag{2.41}$$

Combining the definition (2.36) with the rule (2.37) we can now repeat the computation in example (2.34) using the $\otimes$ notation:

$$\sigma_x^A |i,j\rangle_{A,B} = (\sigma_x \otimes 1_2)(|i\rangle_A \otimes |j\rangle_B) = (\sigma_x|i\rangle_A) \otimes |j\rangle_B.$$

We can also consider other operators, such as

$$\sigma_x^A \sigma_x^B = \sigma_x \otimes \sigma_x.$$

which is an operator that simultaneously flips both spins:

$$\sigma_x^A \sigma_x^B |0,0\rangle_{AB} = |1,1\rangle_{AB} \tag{2.42}$$

You can also use the $\otimes$ notation to combine weird objects. The only rule is that the combination makes sense in each space separated by the $\otimes$. For instance, the object $\langle 0| \otimes |0\rangle$ is allowed, although it is a bit strange. But if you want to operate with it on something, that operation must make sense. For instance

$$(\langle 0| \otimes |0\rangle)(\sigma_x \otimes \sigma_x)$$

makes no sense because even though $\langle 0|\sigma_x$ makes sense, the operation $|0\rangle\sigma_x$ does not. On the other hand, a weird operation which does make sense is

$$(\langle 0| \otimes |0\rangle)(|0\rangle \otimes \langle 0|) = (\langle 0|0\rangle) \otimes |0\rangle\langle 0| = |0\rangle\langle 0|$$

In particular, in the last equality I used the fact that $\langle 0|0\rangle = 1$ is a number and the tensor product of a number with something else, is just the multiplication of the something else by the number.

I am also obliged to say that everything I said extends naturally to systems composed of more than two parts. For instance, if we have a system of 4 qubits, then we can define $\sigma_x^1 = \sigma_x \otimes 1 \otimes 1 \otimes 1$ or $\sigma_x^3 = 1 \otimes 1 \otimes \sigma_x \otimes 1$, and so on. We will for now focus only on bipartite systems. But you have plenty of opportunities to play with multipartite systems in the future.

Matrix representation of the Kronecker product

If A and B are two matrices, then in order to satisfy Eq. (2.37), the components of the Kronecker product must be given by

$$A \otimes B = \begin{pmatrix} a_{1,1}B & \dots & a_{1,N}B \\ \vdots & \ddots & \vdots \\ a_{M,1}B & \dots & a_{M,N}B \end{pmatrix}. \tag{2.43}$$

This is one of those things that you sort of just have to convince yourself that is true. At each entry $a_{i,j}$ you introduce the full matrix B (and then get rid of the parenthesis lying around). For instance

$$\sigma_x \otimes \sigma_x = \begin{pmatrix} 0\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} & 1\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ 1\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} & 0\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \tag{2.44}$$

This provides an automated way to construct tensor product matrices. Such functionality is implemented in any computer library (Mathematica, Matlab, etc.), which is a very convenient tool to use.

We can also do the same for vectors. For instance

$$|0,0\rangle = |0\rangle \otimes |0\rangle = \begin{pmatrix} 1\begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ 0\begin{pmatrix} 1 \\ 0 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{2.45}$$

You can proceed similarly to find the others basis elements. You will then find

$$|0,0\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \qquad |0,1\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \qquad |1,0\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \qquad |1,1\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \tag{2.46}$$

Thus, from the rule (2.43) we therefore see that the correct order of the basis elements is $|0,0\rangle$, $|0,1\rangle$, $|1,0\rangle$ and $|1,1\rangle$. This is known as *lexicographic* order. As an exercise, try to use the rule (2.43) to compute $\langle 0| \otimes |0\rangle$.

The operation highlighted by Eq. (2.43) is implemented in any numerical library. In MATLAB they call it `kron()` whereas in Mathematica they call it `KroneckerProduct[]`. These functions are really useful. You should really try to play with them a bit.

2.4 Entanglement

If qubit A is on Earth and quibit B is on Mars, it makes sense to attribute to them *local states*. For instance, we could have

$$|\psi\rangle_A = \alpha|0\rangle_A + \beta|1\rangle_A, \qquad |\phi\rangle_B = \gamma|0\rangle_B + \delta|1\rangle_B.$$

Then, the global state of AB will be

$$|\psi\rangle_A \otimes |\phi\rangle_B = \Big[\alpha|0\rangle_A + \beta|1\rangle_A\Big] \otimes \Big[\gamma|0\rangle_B + \delta|1\rangle_B\Big]$$
$$= \alpha\gamma|0,0\rangle_{AB} + \alpha\delta|0,1\rangle_{AB} + \beta\gamma|1,0\rangle_{AB} + \beta\delta|1,1\rangle_{AB}.$$

This state looks just like a linear combination of the global basis $|i,j\rangle_{AB}$. However, it is not an arbitrary linear combination because it contains a very special choice of parameters which are such that you can perfectly factor the state into something related to A times something related to B. This is what we call a **product state**. However, quantum theory also allows us to have more general linear combinations which are not necessarily factorable into a product. Such a general linear combination has the form

$$|\psi\rangle_{AB} = \sum_{i,j} C_{i,j}|i,j\rangle_{AB}, \tag{2.47}$$

where $C_{i,j}$ are a set of coefficients. If it happens that we can write $C_{i,j} = f_i g_j$, then the state (2.47) can be factored into a product and is therefore a product state. Otherwise, it is called an **entangled state**.

An important set of entangled states are the so called **Bell states**:

$$|\Phi_1\rangle = \frac{1}{\sqrt{2}}\Big[|0,0\rangle + |1,1\rangle\Big], \tag{2.48}$$

$$|\Phi_2\rangle = \frac{1}{\sqrt{2}}\Big[|0,0\rangle - |1,1\rangle\Big], \tag{2.49}$$

$$|\Phi_3\rangle = \frac{1}{\sqrt{2}}\Big[|0,1\rangle + |1,0\rangle\Big], \tag{2.50}$$

$$|\Phi_4\rangle = \frac{1}{\sqrt{2}}\Big[|0,1\rangle - |1,0\rangle\Big]. \tag{2.51}$$

These states *cannot* be factored into a product of local states (please try to convince yourself). In fact, they are what is known as **maximally entangled states**: we don't have the tools yet to *quantify* the degree of entanglement, so we are not ready yet to properly define the term "maximally". We will get to that later.

In order to better understand the meaning of entanglement, let us discuss what happens when a composite system is in an entangled state and we measure one of the parts. We have seen in Sec. 1.7 of the previous chapter that in order to measure in a basis we define a projection operator $P_i = |i\rangle\langle i|$ such that, if the system is in a state $|\psi\rangle = \sum_i \psi_i|i\rangle$, then the outcome $|i\rangle$ is obtained with probability

$$p_i = \langle\psi|P_i|\psi\rangle = |\psi_i|^2 \tag{2.52}$$

Moreover, if $|i\rangle$ was found, then the state of the system after the measurement is of course $|i\rangle$. We can write that in a slightly different way as follows:

$$|\psi\rangle \to \frac{P_i|\psi\rangle}{\sqrt{p_i}} \tag{2.53}$$

Since $P_i|\psi\rangle = \psi_i|i\rangle$, this is of course the same as the state $|i\rangle$, up to a global phase which is not important.

We now use this to define what is the operation of making a projective measurement on one of the two sub-systems. The operation performing a projective measurement on B will also be a projection operator, but will have the form

$$P_i^B = 1 \otimes |i\rangle\langle i|, \tag{2.54}$$

You can check that this is a valid projection operator ($P_i^B P_j^B = P_i^B \delta_{i,j}$). Thus, with this definition, the rules (2.52) and (2.53) continue to be valid, provided we use this modified projection operator.

As an example, suppose that AB are prepared in the Bell state (2.48). And suppose Bob measures B in the computational basis $\{|0\rangle, |1\rangle\}$. Then we get outcomes 0 or 1 with probabilities

$$p_0 = \langle\Phi_1|P_0^A|\Phi_1\rangle = \frac{1}{2} = p_1 \tag{2.55}$$

Moreover, if B happened to be found in state $|0\rangle_B$, then the global state after the measurement will be

$$|\Phi_1\rangle \to \frac{P_0^A|\Phi_1}{\sqrt{p_0}} = |0,0\rangle \tag{2.56}$$

whereas if the output $|1\rangle_B$ was found, then the state has collapsed to

$$|\Phi_1\rangle \to \frac{P_1^A|\Phi_1}{\sqrt{p_1}} = |1,1\rangle \tag{2.57}$$

Before the measurement system A could have been found in either 0 or 1. But after the measurement, A will be found with certainty in state 0 or with certainty in state 1. We have changed A even though B could have been 100 light years away. This "spooky action at a distance" is the source of a century of debates and research. Of course, the key question is whether Alice, the person that has system A in her lab, can know whether this happened or not. We will see in a second that she cannot, unless Bob sent her a classical communication (like an e-mail) telling her what he found. Thus, information cannot be transmitted faster than the speed of light. There was definitely a change in the state of A and this change was non-local: it took place during a very short time (the time it took to make the measurement) even if A and B are arbitrarily far apart. But no information was transmitted. We will get back to this discussion over and over again, during this chapter and the next ones.

2.5 Mixed states and entanglement

I want you now to recall our construction of a density matrix in Sec. 2.1. What we did there was mix quantum states with classical uncertainty, which was done by considering a machine which is not very good at producing quantum states. As a result, we found that a density matrix could be written as

$$\rho = \sum_i q_i |\psi_i\rangle\langle\psi_i| \tag{2.58}$$

where the $|\psi_i\rangle$ are arbitrary states and the q_i are arbitrary probabilities. This construction may have left you with the impression that the density matrix is only necessary when we want to mix quantum and classical stuff. That is, that a density matrix is not really a quantum thing. Now I want to show you that this is not the case. I will show you that there is an intimate relation between mixed states and entanglement. And this relation is one the key steps relating quantum mechanics and information theory.

Essentially, the connection is made by the notion of **reduced state** or **reduced density matrix**. When a composite system is in a product state $|\psi\rangle_A \otimes |\phi\rangle_B$, then we can attribute the ket $|\psi\rangle_A$ as representing the state of A and $|\phi\rangle_B$ as the state of B. But if the composite system is in an entangled state, like (2.47), then that is no longer possible. As we will show, *if AB are entangled, the reduced state of A and B are mixed states, described by density matrices.*

Suppose we have a bipartite system and, for simplicity, assume that the two parts are identical. Let $|i\rangle$ denote a basis for any such part and assume that the composite system is in a state of the form

$$|\psi\rangle = \sum_i c_i |i\rangle \otimes |i\rangle \tag{2.59}$$

for certain coefficients c_i.[1] If $c_1 = 1$ and all other $c_i = 0$ then $|\psi\rangle = |i\rangle \otimes |i\rangle$ becomes a **product state**. When more than one c_i is non-zero, then the state can never be written as a product. Whenever a state of a bipartite system cannot be written as a product state, we say it is **entangled**.

Now let A be an operator which acts only on system A. Then, its expectation value in the state (2.59) will be

$$\langle A\rangle = \langle\psi|(A\otimes 1)|\psi\rangle \tag{2.60}$$

[1]This is called the Schmidt form of a bipartite state. We will talk more about this in Sec. 2.8.

Carrying out the calculation we get:

$$\begin{aligned}\langle A\rangle &= \sum_{i,j} c_i^* c_j \langle i,i|(A\otimes 1)|j,j\rangle \\ &= \sum_{i,j} c_i^* c_j \langle i|A|j\rangle\langle i|j\rangle \\ &= \sum_i |c_i|^2 \langle i|A|i\rangle\end{aligned}$$

If we now define the density matrix of system A as

$$\rho_A = \sum_i |c_i|^2 |i\rangle\langle i| \tag{2.61}$$

then the expectation value of A becomes

$$\langle A\rangle = \mathrm{tr}(A\rho_A) \tag{2.62}$$

This result is quite remarkable. Note how Eq. (2.61) has *exactly* the same form as Eq. (2.58), with the classical probabilities q_i replaced by $|c_i|^2$. But there are no classical probabilities at play here: we started with a pure state. Moreover, we also see that in general the state of A will be a mixed state. The only exception is when the original state was a product state. Then one $c_i = 1$ and all other $c_j = 0$, so that $\rho_A = |i\rangle\langle i|$. Thus, we conclude that whenever the global AB state is entangled, the reduced state of a given part will be a mixed state. Eq. (2.61) is what we call a **reduced density matrix**, a concept which is fundamental in the theory of Quantum Information and which we will use throughout this course. In the above calculation I introduced it in a not so formal way. But don't worry, in the next section we will go back to it and see how to define it more generally.

But before we do so, I just want to give one example, which will also connect with our discussion of entanglement in Sec. 2.4, in particular Eq. (2.57). Suppose again that AB is in the Bell state (2.48). This state has the form of Eq. (2.59) with $c_i = 1/\sqrt{2}$. Thus, it is easy to apply Eq. (2.61), which gives

$$\rho_A = \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \tag{2.63}$$

We therefore see that the reduced state of A is actually the maximally mixed state (2.16). This is a feature of all Bell states and it is the reason we call them maximally entangled states: we will learn soon that *the degree of entanglement can be quantified by how mixed the reduced state is.*

Now let us ask what is the state of A after we measure B. As we have seen in Eq. (2.57), the composite state after the measurement can be either $|0,0\rangle$ or $|1,1\rangle$, both occurring with probability $1/2$. Thus, if Alice does not know the outcomes of the measurements that B performed, then best possible guess to the state of A will be a classical probabilistic combination

$$\rho_A = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.64}$$

which is exactly the same state as (2.63). Hence, from the point of view of Alice, it is impossible to know if the state of A is mixed because of entanglement or if it is mixed because Bob performed some measurements. All Alice can know is that the density matrix of A has the form of a maximally mixed state. This is called **the ambiguity of mixtures**. Even though the global AB state is affected by the measurement, from the point of view of Alice, she has no way of knowing. The only way that A would know is if she receives a *classical communication* from B. That is, if Bob sends an e-mail to Alice saying "Hey Alice, are you going to the party tonight? Oh, by the way, I measured my qubit and found it in 0."

2.6 The partial trace

The calculation that led us to Eq. (2.61) is what we call a partial trace. The trace, which we studied in Sec. 1.11, is an operation that receives an operator and spits out a number. The partial trace is an operation which receives a tensor product of operators and spits another operator, but living in a smaller Hilbert space. Why this is the correct procedure to be used in defining a reduced density matrix will be explained shortly.

Consider again a composite system AB. Let $|a\rangle$ and $|b\rangle$ be basis sets for A and B. Then a possible basis for AB is the tensor basis $|a, b\rangle$. What I want to do is investigate the trace operation within the full AB space. To do that, let us consider a general operator of the form $\mathcal{O} = A \otimes B$. After we learn how to deal with this, then we can generalize for an arbitrary operator, since any operator on AB can always be written as

$$\mathcal{O} = \sum_\alpha A_\alpha \otimes B_\alpha \tag{2.65}$$

for some index α and some set of operators A_α and B_α.

Let us then compute the trace of $\mathcal{O} = A \otimes B$ in the $|a, b\rangle$ basis:

$$\begin{aligned}
\text{tr}(\mathcal{O}) &= \sum_{a,b} \langle a, b|\mathcal{O}|a, b\rangle \\
&= \sum_{a,b} (\langle a| \otimes \langle b|)(A \otimes B)(|a\rangle \otimes |b\rangle) \\
&= \sum_{a,b} \langle a|A|a\rangle \otimes \langle b|B|b\rangle \\
&= \sum_{a} \langle a|A|a\rangle \sum_{b} \langle b|B|b\rangle
\end{aligned}$$