FERNANDO DE MELO & RAFAEL CHAVES

SELECTED TOPICS ON QUANTUM FOUNDATIONS

CBPF - NOTAS DE FÍSICA

DISCLAIMER: NOT EVEN THE AUTHORS HAVE READ THESE NOTES. THEY ARE THE PRODUCT OF SOME (VARIOUS BY NOW) YEARS OF DISCUSSIONS AND WORK IN THE AREA OF QUANTUM INFORMATION AND FOUNDATIONS BY BOTH OF US. HOWEVER, THESE NOTES WERE WRITTEN WAY TOO FAST. WAY FASTER THAN A SUBJECT LIKE FOUNDATIONS OF QUANTUM MECHANICS DESERVES AND REQUIRES. IT IS JUST ITS FIRST VERSION, AND IT IS GOING TO BE UPDATED IN THE NEXT YEARS, DECADES... EVENTUALLY IT WILL BECOME SOMETHING A BIT MORE POLISHED. USE IT WITH MODERATION!

FERNANDO DE MELO AND RAFAEL CHAVES, JULY 29, 2019.

Copyright © 2019 Fernando de Melo & Rafael Chaves

PUBLISHED BY CBPF - NOTAS DE FÍSICA

First printing, July 2019

Contents

1 Text-book Quantum Mechanics

Here we quickly review what it is sometimes called "text-book quantum mechanics". It does not raise to the status of an fully fledged interpretation of quantum mechanics, but it is what most physicists have in mind when discuss and/or apply the theory of quantum mechanics.

1.1 Postulates: Old-fashioned text-books

Quantum mechanics was crafted to be a theory that describes atomic and subatomic systems, their structure and time evolution. In a sense, quantum mechanics should play the same role of classical mechanics, but within the realm of microscopic systems – at least that is the way it was created. Therefore, it is reasonable to follow the structure of classical mechanics in order to define quantum mechanics. The questions we must address are then:

i) How to mathematically describe the state of a system at a given time?

ii) Given this state, how can we determine the value of the various physical quantities?

iii) How can we find the state of the system at an arbitrary time t, given the state at some other time t'?

These questions were answered within classical mechanics long time ago, in the XIX century[1]. Our aim here to address these questions within quantum mechanics, i.e., to describe the postulates of quantum mechanics. The postulates of a theory are the premises, the starting point from which everything else must be derived.

[1] Do you know how to answer these questions within classical mechanics? If not, please get back to your favorite classical mechanics book and distill these answers from there.

1.1.1 1^{st} Postulate: Quantum states

Physics is a model of Nature. As such, it is reasonable (mandatory?) to take experiments as the guiding processes to define the main aspects of the theory. From the well-known two-slit experiment, we can apprehend that whatever we use to describe quantum systems it must allow for interference, thus some kind of superposition, and also for situations where this interference is not observed.

The two-slit experiment, and many others, led to the idea of states as unit vectors in a Hilbert space.

First Postulate

To every physical system we assign a Hilbert space. The state of quantum system is described by a unit vector in this Hilbert space.

The reason why we take unit vectors will be clear with the other postulates.

The first postulate then establishes the following correspondence:

$$\begin{array}{lcl} \text{Quantum system} & \longrightarrow & \mathscr{H} \text{ (Hilbert space)} \\ \text{Quantum state} & \longrightarrow & \psi \in \mathscr{H} \text{ s.t. } \|\psi\| = 1 \end{array}$$

As Hilbert spaces are equipped with a scalar product, a vector $\psi \in \mathscr{H}$ is a unit vector, i.e., it is normalized if $\|\psi\| = \sqrt{\langle \psi, \psi \rangle} = 1$.

The fact the 1st postulate identifies quantum states with vectors immediately allows for superpositions: Given $\psi, \phi \in \mathscr{H}$, then $\psi + \phi \in \mathscr{H}$. Note, however, that arbitrary linear combinations do not preserve the norm: Given $\psi, \phi \in \mathscr{H}$, with $\|\psi\| = \|\phi\| = 1$, in general $\|\alpha\psi + \beta\phi\| \neq 1$ for $\alpha, \beta \in \mathbb{C}$. Therefore, more rigorously, given the associated Hilbert space, the space of states is the projective space over that Hilbert space, where we normalize the vectors and ignore global phases.

$$|\alpha| e^{i\theta} \psi \sim \psi, \text{ for } \|\psi\| = 1.$$

For the normalized vectors $\psi \in \mathscr{H}$, for which we ignore a global phase, we reserve the "ket" symbol:

$$\psi \longrightarrow |\psi\rangle.$$

With these restrictions, the set of quantum states, strictly speaking, is not a vector space, but superpositions are still allowed.

1.1.2 2nd Postulate: Quantum measurements

A fundamental aspect of any physical theory is the measurement of a system's properties. In this respect, quantum mechanics departs from the classical intuition that *all* the properties of a system are well defined given the system's state. Within quantum mechanics the value associated with a physical property and the property itself are detached concepts. This detachment can be readily observed with a Stern-Gerlach experiment (see Fig. 1.1).

This non-commutativity of quantum measurements goes hand-in-hand with the vectorial structure dictated by the first postulate. To start with, consider the situation where the quantum system is an state $|\psi\rangle \in \mathscr{H}$, and we want to determine "how much" of another state, say $|\phi\rangle \in \mathscr{H}$, is in $|\psi\rangle$. This is given by the projection of $|\psi\rangle$ onto $|\phi\rangle$, i.e., by:

$$\langle \phi | \psi \rangle | \phi \rangle = (|\phi\rangle\langle\phi|)|\psi\rangle = P_\phi |\psi\rangle,$$

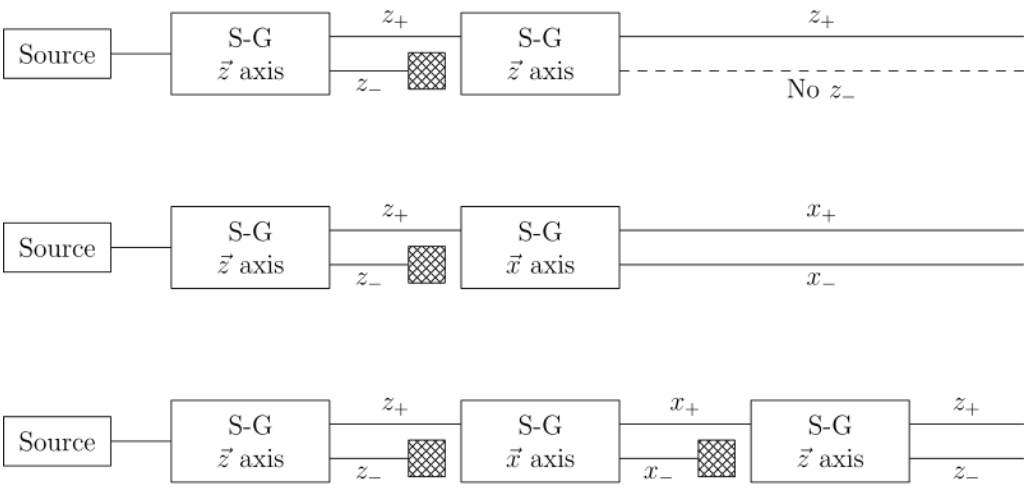

Figure 1.1: Sequential Stern-Gerlach setups. Quantum properties are detached from their values. Public domain picture by Francesco Versaci, taken from Wikipedia entry on Stern-Gerlach experiment.

where we defined the projector onto $|\phi\rangle$ as $P_\phi = |\phi\rangle\langle\phi|$. See Fig. 1.2. A projector is any Hermitian linear operator P such that $P^2 = P$. This is related, for instance, with the Stern-Gerlach experiment in which the first measurement produces the state $|+_z\rangle$, and we want to know what is the chance of a second measurement to project this state onto the states $|\pm_x\rangle := (|+_z\rangle \pm |-_z\rangle)/\sqrt{2}$, where $\langle+_z|+_z\rangle = \langle-_z|-_z\rangle = 1$ and $\langle+_z|-_z\rangle = \langle-_z|+_z\rangle = 0$ (that guarantees the normalization of $|\pm_x\rangle$).

As we are using unit vectors to describe quantum states, we are going to associate the norm square of the projection with the probability of performing such a projection. Going back to our abstract example shown in Fig.1.2, we have

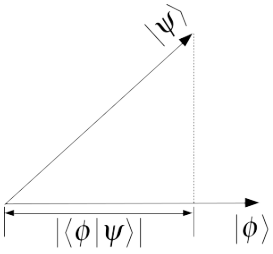

Figure 1.2: Projecting $|\psi\rangle$ on $|\phi\rangle$.

$$\begin{aligned}
\Pr(\phi|\psi) &= \|P_\phi|\psi\rangle\|^2, \\
&= \||\phi\rangle\langle\phi|\psi\rangle\|^2, \\
&= \left(\sqrt{(\langle\phi|\psi\rangle^*\langle\phi|)(|\phi\rangle\langle\phi|\psi\rangle)}\right)^2, \\
&= \left(\sqrt{|\langle\phi|\psi\rangle|^2\langle\phi|\phi\rangle}\right)^2, \\
&= |\langle\phi|\psi\rangle|^2.
\end{aligned}$$

In this way, the probability of not projecting onto $|\phi\rangle$, or equivalently, to project onto the orthogonal subspace to $|\phi\rangle$ which is related to the projector $P_{\phi^\perp} = \mathbb{1} - P_\phi$. [2]

[2] Note that

$$P_{\phi^\perp}^2 = (\mathbb{1} - P_\phi)^2 = \mathbb{1} - P_\phi = P_{\phi^\perp},$$

and

$$P_{\phi^\perp}|\phi\rangle = |\phi\rangle - |\phi\rangle = 0,$$

and thus$P_{\phi^\perp}$ is indeed a projector onto the orthogonal subspace of $|\phi\rangle$.

$$\begin{aligned}\Pr(\phi_\perp|\psi) &= \|P_{\phi^\perp}|\psi\rangle\|^2,\\ &= \left(\sqrt{\langle\psi|(\mathbb{1}-P_\phi)^\dagger(\mathbb{1}-P_\phi)|\psi\rangle}\right)^2,\\ &= \left(\langle\psi|(\mathbb{1}-P_\phi)|\psi\rangle\right),\\ &= 1-\Pr(\phi|\psi).\end{aligned}$$

The sum of probabilities of all possible outcomes is then equal to one. Moreover, note that the (Hilbert-Schmidt) scalar product [3] between the two projectors is zero,

$$\mathrm{Tr}\left(P_{\phi^\perp}^\dagger P_\phi\right) = \mathrm{Tr}\left((\mathbb{1}-P_\phi)^\dagger P_\phi\right) = \mathrm{Tr}\left(P_\phi - P_\phi^2\right) = 0,$$

that is, the projectors are orthogonal.

[3] The Hilbert-Schmidt inner product between two operators is defined as

$$\langle A,B\rangle = \mathrm{Tr}(A^\dagger B).$$

Another thing the we can apprehend from the Stern-Gerlach is that if right after we measure, say $|+_x\rangle$, we repeat such a measurement we get again $|+_x\rangle$. This is in fact expected from any process that we can call a measurement. The system's state after the measurement of the projector P_ϕ is the normalized projected state:

$$\frac{P_\phi|\psi\rangle}{\|P_\phi|\psi\rangle\|} = \frac{|\phi\rangle\langle\phi|\psi\rangle}{\||\phi\rangle\langle\phi|\psi\rangle\|} = \frac{\langle\phi|\psi\rangle}{|\langle\phi|\psi\rangle|}|\phi\rangle \sim |\phi\rangle.$$

Suppose now that the measurement we are performing has more than two outcomes, say with n outcomes. Think for instance in a Stern-Gerlach experiment with a spin-1 particle, or the measurement of angular moment. In these situations we might make the correspondence of each outcome $i \in [n]$ with a projector P_{ϕ_i}, in such a way that $\langle P_{\phi_i}, P_{\phi_j}\rangle = d_i\delta_{ij}$. Here d_i is the dimension of the subspace $\mathscr{H}_i$ associated with the projector P_{ϕ_i}. If the system is prepared in the state $|\psi\rangle$ before the measurement, then the conservation of probabilities reads:

$$\begin{aligned}1 &= \sum_{i=1}^n \Pr(i|\psi)\\ &= \sum_{i=1}^n \|P_{\phi_i}|\psi\rangle\|^2\\ &= \sum_{i=1}^n \sqrt{\langle\psi|P_{\phi_i}^\dagger P_{\phi_i}|\psi\rangle}^2\\ &= \langle\psi|\left(\sum_{i=1}^n P_{\phi_i}\right)|\psi\rangle.\end{aligned}$$

As the sum of the probabilities of all possible outcomes must be equal to one independently of the state of the system, then we must have:

$$\sum_{i=1}^n P_{\phi_i} = \mathbb{1}.$$

The projectors of a given measurement then split the Hilbert space $\mathscr{H}$, associated with the system, into orthogonal subspaces: $\mathscr{H} = \bigoplus_{i=1}^n \mathscr{H}_i$ with $\dim(\mathscr{H}) = \sum_{i=1}^n d_i$.

Up to now we were only interested on the outcomes of measurements. Consider now that a physical property φ takes the value ϕ_i when the state is measured in the subspace associated to the projector P_{ϕ_i}, i.e., when we get the outcome i. The mean/average value of this property when the state of the system is described by $|\psi\rangle$ is given by:

$$\begin{aligned}\sum_{i=1}^{n} \phi_i \Pr(i|\psi) &= \sum_{i=1}^{n} \phi_i \|P_{\phi_i}|\psi\rangle\|^2 \\ &= \langle\psi| \left(\sum_{i=1}^{n} \phi_i P_{\phi_i} \right) |\psi\rangle \\ &:= \langle\psi|\Phi|\psi\rangle := \langle\Phi\rangle_\psi.\end{aligned}$$

The linear operator $\Phi := \sum_{i=1}^{n} \phi_i P_{\phi_i}$ holds all the information about the possible values of the physical quantity φ, and also about the projectors associated with these values. Within quantum mechanics , such an operators are called *observables*. More generally, observables are self-adjoint linear operators. The spectral theorem below spells out the main properties of observables.

Theorem 1 (Spectral decomposition)**.** *Let $\Phi : \mathbb{C}^d \mapsto \mathbb{C}^d$ be a self-adjoint (Hermitian) linear operator, i.e., an observable. Then Φ can be diagonalized, with its eigenvectors forming an orthonormal basis for $\mathbb{C}^d$, and with real eigenvalues.*

Observables play then the role of numerical random variables in probability theory. The eigenvalues are related to the possible values of the physical quantity, and as expected, are real numbers. The eigenvectors determine the projectors onto subspaces which lead to the same value of the physical property: $P_{\phi_i} = \sum_{k=1}^{d_i} |\phi_{i,k}\rangle\langle\phi_{i,k}|$.

We are finally in position to write the second postulate of quantum mechanics .

Second Postulate

To every physical property φ we assign an observable Φ, acting on $\mathscr{H}$, whose spectral decomposition can be written as $\sum_{i=1}^{n} \phi_i P_{\phi_i}$, where $\forall i, j \in [n]$ we have $\phi_i \in \mathbb{R}$, $P_{\phi_i}^2 = P_{\phi_i} = P_{\phi_i}^\dagger$, $\sum_{i=1}^{n} P_{\phi_i} = \mathbb{1}$, and $\langle P_{\phi_i}, P_{\phi_j}\rangle = d_i \delta_{ij}$.
Given a state $|\psi\rangle \in \mathscr{H}$ of the system, the probability of measuring the value ϕ_i is given by:

$$\Pr(\phi_i|\psi) = \|P_{\phi_i}|\psi\rangle\|^2, \quad \text{(Born's rule)}$$

and after the measurement the system is left on the state:

$$\frac{P_{\phi_i}|\psi\rangle}{\|P_{\phi_i}|\psi\rangle\|}.$$

1.1.3 3rd Postulate: Quantum dynamics

The last point we must address is about the dynamics of quantum states. Like for classical dynamics, the equation that dictates the evolution of quantum systems cannot be derived from "first principles", but we do have some guiding criteria and experimental results that we must take into account.

The first property the equation of motion for quantum mechanics must obey is that it must be causal. Given the state of the system at a given time, say $|\psi(t)\rangle$, then the state of the system is known, in principle, for all other times. This implies that the equation of motion is a first order differential equation:

$$\frac{d|\psi(t)\rangle}{dt} = A|\psi(t)\rangle,$$

where A is some linear operator, possibly time dependent.

Secondly, A must be related to the generator of time translations. From classical mechanics we know that the Hamiltonian is such a generator. By analogy, we expect A to be a function of the Hamiltonian operator H, and thus:

$$\frac{d|\psi(t)\rangle}{dt} = f(H)|\psi(t)\rangle.$$

Moreover, in the case of a system composed of two non-interacting subsystems, with Hamiltonians H_1 and H_2 respectively, the linearity of f implies that

$$f(H_1 + H_2) = f(H_1) + f(H_2).$$

The only possibility is thus $f(H) = \alpha H$ with $\alpha \in \mathbb{C}$. Which then gets us to:

$$\frac{d|\psi(t)\rangle}{dt} = \alpha H|\psi(t)\rangle. \tag{1.1}$$

Lastly we want the evolution to preserve the total probability, i.e. , we want it to preserve the norm (squared) of state vectors. Therefore we force that

$$\frac{d\||\psi(t)\rangle\|^2}{dt} = \frac{d\langle\psi(t)|\psi(t)\rangle}{dt} = \frac{d\langle\psi(t)|}{dt}|\psi(t)\rangle + \langle\psi(t)|\frac{d|\psi(t)\rangle}{dt} = 0.$$

Using equation (1.1), and its conjugate, we get:

$$\langle\psi(t)|(\alpha^* H + \alpha H)|\psi(t)\rangle = 0.$$

As this above expression must be valid for any quantum state at any moment, it implies that $\alpha = -\alpha^*$. We can then define $\alpha = 1/(\mathrm{i}\gamma)$, with $\gamma \in \mathbb{R}$, to obtain:

$$\mathrm{i}\gamma\frac{d|\psi(t)\rangle}{dt} = H|\psi(t)\rangle.$$

By dimensional analysis we see that γ has units of angular momentum, $[J.s]$. Its value is finally set by experimental results, like the ones where we measure the spectrum of atoms for instance, and we fix $\gamma = \hbar = h/(2\pi)$. With all that we get the third postulate of quantum mechanics.

Third Postulate

The time evolution of a quantum state $|\psi(t)\rangle$ is governed by Schrödinger's equation:

$$i\hbar\frac{d|\psi(t)\rangle}{dt} = H|\psi(t)\rangle,$$

where H is the (possibly time dependent) observable associated the system's total energy, its Hamiltonian.

For time independent Hamiltonians, Schrödinger's equation can be integrated to give:

$$|\psi(t)\rangle = e^{-\frac{iH}{\hbar}(t-t')}|\psi(t')\rangle,$$

which then connects the state at time t with the state at time t'. The operator that translates in time the quantum state, the evolution operator

$$U(t,t') = e^{-\frac{iH}{\hbar}(t-t')}$$

is a unitary matrix, as $U(t,t')^\dagger U(t,t') = U(t,t')U(t,t')^\dagger = \mathbb{1}$. More generally, even for time dependent Hamiltonians, the evolution operator is also a unitary matrix as such matrices preserve the norm of quantum states:

$$\||\psi(t)\rangle\|^2 = \langle\psi(t)|\psi(t)\rangle = \langle\psi(t')|U(t,t')^\dagger U(t,t')|\psi(t')\rangle = \langle\psi(t')|\psi(t')\rangle = \||\psi(t')\rangle\|^2.$$

1.1.4 Composite quantum systems

To finish this brief review about the structure of quantum mechanics, we must spell out how to deal with composite systems.

Suppose that, employing the first postulate, to a system A we assign the Hilbert space $\mathscr{H}_A$, and to a system B we assign the Hilbert space $\mathscr{H}_B$. Which space should we assign to the composite system AB? To begin with, assume the two systems do not interact with each other, possibly because they are very apart. In this situation, if we prepare system A in the state $|\psi\rangle \in \mathscr{H}_A$ and measure an observable $\Phi = \sum_i \phi_i|\phi_i\rangle\langle\phi_i|$, and similarly prepare the state $|\chi\rangle \in \mathscr{H}_B$ and measure the observable $\Xi = \sum_j \xi_j|\xi_j\rangle\langle\xi_j|$, we expect no correlation among them, and thus the probability of obtaining the outcomes (ϕ_i, ξ_j) must be given by:

$$\Pr(\phi_i, \xi_j|\psi,\chi) = \Pr(\phi_i|\psi)\Pr(\xi_j|\chi).$$

This result for uncorrelated systems suggests some kind of product structure between the individual spaces $\mathscr{H}_A$ and $\mathscr{H}_B$. For classical random variables we employ the Cartesian product. However, as for evaluating the probabilities in quantum mechanics we use the scalar product, the Cartesian product between the individual spaces would imply the sum of the probabilities. The mathematical structure that correctly reproduces the above expression is the *tensor product* of the individual spaces. To the combined system AB we then

assign the Hilbert space $\mathscr{H}_{AB} = \mathscr{H}_A \otimes \mathscr{H}_B$. In the example above, the state of the total system would then be given by $|\Psi\rangle = |\psi\rangle \otimes |\chi\rangle$, and the observable being measured on the combined system is $\mathscr{O} = \Phi \otimes \Xi$, and therefore:

$$\begin{aligned}\Pr(\phi_i, \xi_j | \psi, \chi) &= \|(|\phi_i\rangle\langle\phi_i| \otimes |\xi_j\rangle\langle\xi_j|)(|\psi\rangle \otimes |\chi\rangle)\|^2 \\ &= |\langle\phi_i|\psi\rangle|^2 |\langle\xi_j|\chi\rangle|^2 \\ &= \Pr(\phi_i|\psi)\Pr(\xi_j|\chi),\end{aligned}$$

as required.

There are many consequences in using the tensor product instead of the Cartesian product. First, as we saw above, we get the correct description for the probabilities of uncorrelated systems. Secondly, the dimension of the combined space grows with the product of the individual dimensions: $\dim(\mathscr{H}_{AB}) = \dim(\mathscr{H}_A)\dim(\mathscr{H}_B)$. As such, it grows way faster than the space associated with combined classical random variables. [4] This means that we have "space" for many more different configurations in the quantum domain, when compared to the classical states. These "extra" states come exactly as the correlated states, the *quantum entangled* states that we are going to look at below.

[4] The dimension of a Cartesian product of spaces is the sum of the individual dimensions.

1.2 Postulates reloaded: modern text-books

In order to discuss fundamental aspects of quantum mechanics we must employ the most general form of the its postulates. In the recent years, say last 50 years, the area of quantum information, the more general form of quantum postulates has become standard. Here we will only touch in some aspects of this generalization, and some further details can be seen for instance in the book by Nielsen and Chuang. [5]

[5] Quantum Computation and Quantum Information. Michael A. Nielsen and Issac L. Chuang, Cambridge University Press.

1.2.1 1st Postulate: Density matrix

Consider a situation in which when we push a button on a machine it prepares with probability p_1 a quantum system in the state $|\psi_1\rangle$, with probability p_2 it prepares the state $|\psi_2\rangle$, and so on up to the n-th preparation $|\psi_n\rangle$ which happens with probability p_n. As this is a probabilistic preparation, we have $p_i \geq 0$, $\forall i \in [n]$, and $\sum_{i=1}^{n} p_i = 1$. The expectation value of an observable Φ being measured in this system is then given by:

$$\langle\Phi\rangle_\rho = p_1\langle\psi_1|\Phi|\psi_1\rangle + p_2\langle\psi_2|\Phi|\psi_2\rangle + \ldots + p_n\langle\psi_n|\Phi|\psi_n\rangle,$$

where the symbol ρ identifies the system preparation in the present scenario.

Using the linear operator known as trace [6], $\mathrm{Tr} : \mathscr{L}(\mathscr{H}) \mapsto \mathbb{C}$ (with $\mathscr{L}(\mathscr{H})$ the set of linear operators acting on $\mathscr{H}$), we have

$$\begin{aligned}\langle\Phi\rangle_\rho &= \sum_{i=1}^{n} p_i \mathrm{Tr}(|\psi_i\rangle\langle\psi_i|\Phi), \\ &= \mathrm{Tr}\left((\sum_{i=1}^{n} p_i |\psi_i\rangle\langle\psi_i|)\Phi \right), \\ &:= \mathrm{Tr}\left(\rho\Phi\right).\end{aligned}$$

The linear operator $\rho := \sum_{i=1}^{n} p_i |\psi_i\rangle\langle\psi_i|$ is called the preparation's *density matrix*, and it represents the description of the system given the knowledge we have about it. It contains all the information about the preparation of the system's ensemble, and as such it grants us the possibility to evaluate the probabilities of any measurements we perform on the system. The density matrix allows us to describe situations, like the one we described above, where we have some ignorance about the preparation of the quantum system.

Note that the situation above is not the same as the one where we prepare the state

$$|\psi\rangle = \sum_{i=1}^{n} p_i |\psi_i\rangle.$$

First of all this is not a normalized state (the states $|\psi_i\rangle$ don't even need to be orthonormal). Second, even if we somehow normalize the state, this would represent the situation where with probability one the machine produces always the same state $|\psi\rangle$.

The scenario where we have full knowledge about the preparation of a quantum system (see next chapter), and we are sure that each element of the ensemble is prepared in the same quantum state $|\psi\rangle$, we say we have a *pure state*. The density matrix associated to this vector is then:

$$|\psi\rangle \to \rho_\psi := |\psi\rangle\langle\psi|.$$

Note that within this density matrix formalism the global phase issue is no longer present.

In the case when the ensemble is composed by two or more states, i.e., we don't have full control about the preparation of the system, then we say we have a *mixed state*, which is then described by the density matrix

$$\rho = \sum_{i=1}^{n} p_i |\psi_i\rangle\langle\psi_i| \;\text{ with }\; p_i \geq 0 \forall i \in [n], \text{ and } \sum_{i=1}^{n} p_i = 1.$$

"Mixed" state is an usual abuse of language, as the system in this situation does not have in fact a well defined state.

Given the definition of the density matrix, it is easy to prove its defining properties:

i) Normalization: $\mathrm{Tr}(\rho) = 1$;

[6] The trace of an $n \times n$ square matrix A is defined as,

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii},$$

where a_{ii} denotes the entry on the ith row and ith column of A. The defining properties of the trace are:

$$\begin{aligned}\mathrm{Tr}(A+B) &= \mathrm{Tr}(A) + \mathrm{Tr}(B) \\ \mathrm{Tr}(cA) &= c\mathrm{Tr}(A) \\ \mathrm{Tr}(DE) &= \mathrm{Tr}(ED),\end{aligned}$$

where A, B are $n \times n$ matrices, $c \in \mathbb{C}$, D is a $m \times n$ matrix, and E is a $n \times m$ matrix.

ii) Hermiticity: $\rho = \rho^\dagger$

iii) Positive semi-defitness: $\rho \geq 0$ which means $\forall |\psi\rangle \in \mathscr{H}$, $\langle\psi|\rho|\psi\rangle \geq 0$;

iv) Purity: $\mathrm{Tr}(\rho^2) \leq 1$, with equality only for pure states.

Let $\mathscr{D}(\mathscr{H}) := \{\rho \in \mathscr{L}(\mathscr{H}) | \rho \geq 0, \mathrm{Tr}(\rho) = 1\}$ be the set of density matrices acting on $\mathscr{H}$. The first postulate can be restated as follows:

First Postulate

To every physical system we assign a Hilbert space $\mathscr{H}$. The quantum system is described by a density matrix $\rho \in \mathscr{D}(\mathscr{H})$.

1.2.2 2nd Postulate: POVM's

Like the description of states, measurements can also be generalized. In the description of the second postulate we gave above, we employed projectors as measurement operators. But what are the fundamental requirements of a measuring process?

Let the set of linear operators $\{M_x\}_{x=1}^m$ be associated with a measurement process with m outcomes, i.e., the operator M_x is associated with obtaining the outcome $x \in [m]$. The two requirements of a measurement are then:

Positivity of probabilities: Taking $\Pr(x|\rho) = \mathrm{Tr}(M_x\rho)$, to require $\Pr(x|\rho) \geq 0$ for all $\rho \in \mathscr{D}(\mathscr{H})$ is to impose $M_x \geq 0$ (positive semi-definiteness) for all $x \in [m]$.

Some outcome must happen: We must require the probabilities to sum up to one for any state, i.e., $\sum_{x=1}^m \Pr(x|\rho) = 1$ for all $\rho \in \mathscr{D}(\mathscr{H})$. This leads to $\sum_{x=1}^m \mathrm{Tr}(M_x\rho) = 1$, and, as it must be true for any system description, it implies $\sum_{x=1}^m M_x = \mathbb{1}$.

Note that we arrive at a construction very similar to the one for projectors, but without needing to require orthogonality among the measurement operators. A set of operators $\{M_x\}_{x=1}^m$, such that $M_x \geq 0$ for all $x \in [m]$, and $\sum_{x=1}^m M_x = \mathbb{1}$, is called a Positive Operator Valued Measure (POVM) and it is associated to a measurement process. Of course projectors fulfill these constraints, and as such are valid POVMs – although we reserve the term *projective measurement* for them.

When using POVMs we usually are not interested in the state after the measurement, but only on the probabilities of outcomes. It is possible to describe the state after the measurement by connecting the POVM measurement with a projective measurement in a larger system. This is the how the so-called *instruments* are constructed. For these notes instruments are not necessary and we refer to the book by Busch, Lahti and Mittlestaedt for furhter reading.[7]

With this generalization, the second postulate reads as follows.

[7] The Quantum Theory of Measurement. Paul Busch, Pekka J. Lahti, and Peter Mittlestaedt (Springer).

Second Postulate

To every measurement process we assign a POVM, i.e., a set of linear operators $\{M_x\}_{x=1}^m$, with $M_x \geq 0$ for all $x \in [m]$ and $\sum_{x=1}^m M_x = \mathbb{1}$. Given a description ρ of the system, the probability of measuring the outcome $x \in [m]$ is given by:

$$\Pr(x|\rho) = \mathrm{Tr}(M_x \rho).$$

1.2.2.1 No go theorem: discrimination of non-orthogonal states

Now that we are equipped with the most general form of quantum measurement, we can address our first important point: we cannot discriminate with certainty, in a single shot experiment, between non-orthogonal quantum states.

For example, consider the scenario where Amary wants to send the outcome of a coin flip to Bacuara.[8] If Amary gets heads, she prepares a photon the horizontal polarization, $|H\rangle$, and sends it to Bacuara. If she gets tails, she prepares a photon in the polarization $(|H\rangle + |V\rangle)/\sqrt{2}$ and sends it to Bacuara. Is possible for Bacuara to know with certainty what was the outcome of the coin flip? It doesn't sound likely. In fact, if Bacuara has at his disposal a polarized beam splitter he can measure the projectors $|H\rangle\langle H|$ and $|V\rangle\langle V|$. If he gets a click in the "V-port", he knows the coin toss gave tails. However, if he gets a click in the "H-port" he cannot know what was the outcome of the coin toss.

[8] Amary and Bacuara are names from Tupi origin. Amery means "leafy tree", while Bacuara means "wise man".

Could he make it better if he had access to POVM's? It turns out that this is not the case.

Theorem 2 (Impossibility of perfect discrimination between two non-orthogonal states.)**.** *Let $|\psi_1\rangle$ and $|\psi_2\rangle$ be two non-orthogonal states in $\mathscr{H}$. There exists no quantum measurement that can distinguish the two states with certainty in a single shot.*

Proof. The proof goes by the way of contradiction. Suppose that there is a POVM, with elements $\{M_1, M_2\}$, such that

$$\langle\psi_1|M_1|\psi_1\rangle = 1, \text{ and } \langle\psi_2|M_2|\psi_2\rangle = 1.$$

That is, the POVM distinguishes the two states with certainty. When we get the outcome 1 we known the state $|\psi_1\rangle$ was prepared, and when we get the outcome 2 we know the state $|\psi_2\rangle$ was prepared. Moreover, as $M_1 + M_2 = \mathbb{1}$ (property of a POVM), then

$$1 = \langle\psi_1|(M_1+M_2)|\psi_1\rangle = \langle\psi_1|M_1|\psi_1\rangle + \langle\psi_1|M_2|\psi_1\rangle = 1 + \langle\psi_1|M_2|\psi_1\rangle.$$

Therefore, $\langle\psi_1|M_2|\psi_1\rangle = 0$. Similarly, $\langle\psi_2|M_1|\psi_2\rangle = 0$. The POVM makes no error in the discrimination.

Since $M_2 \geq 0$ (positive semi-definite), then it has square-root, and thus:

$$0 = \langle\psi_1|M_2|\psi_1\rangle = \langle\psi_1|\sqrt{M_2}\sqrt{M_2}|\psi_1\rangle = \|\sqrt{M_2}|\psi\rangle\|^2.$$

As the only vector with null norm is the zero vector, then $\sqrt{M_2}|\psi\rangle = 0$. Hold this result for a moment.

Now, as $|\psi_2\rangle$ by hypothesis is not orthogonal to $|\psi_1\rangle$, we can write the first as

$$|\psi_2\rangle = \alpha|\psi_1\rangle + \beta|\phi\rangle,$$

where $|\phi\rangle$ is orthogonal to $|\psi_1\rangle$, and $|\alpha|^2 + |\beta|^2 = 1$ with $|\beta| \in [0,1[$. Note that we do not allow $|\beta| = 1$, because in that case we would have $\langle\psi_1|\psi_2\rangle = 0$, i.e., the states would be orthogonal.

Using that $\sqrt{M_2}|\psi\rangle = 0$, we evaluate

$$\begin{aligned}\sqrt{M_2}|\psi_2\rangle &= \alpha\sqrt{M_2}|\psi_1\rangle + \beta\sqrt{M_2}|\phi\rangle,\\ &= \beta\sqrt{M_2}|\phi\rangle.\end{aligned}$$

Which is a contradiction with the hypothesis that the POVM discriminates the two states with certainty, because

$$\langle\psi_2|M_2|\psi_2\rangle = \langle\psi_2|\sqrt{M_2}\sqrt{M_2}|\psi_2\rangle = |\beta|^2\langle\phi|M_2|\phi\rangle \leq |\beta|^2 < 1.$$

□

The above theorem proves that within quantum mechanics it is impossible to discriminate two non-orthogonal states with 100% of certainty.

If we relax our certainty requirement and allow to be wrong sometimes, can we do something? One of the first results in quantum information is what is known as Helstrom bound.[9] It states that given two density matrices ρ_1 and ρ_2 with probabilities p_1 and p_2, respectively, then the best POVM discriminates these descriptions with success probability given by:

$$\Pr(\text{success}) = \frac{1}{2}\left(1 + \text{Tr}|p_1\rho_1 - p_2\rho_2|\right).$$

If the density matrices are different, there is thus always a POVM that discriminates the states with probability greater than random guess. It is easy to check that if $\rho_1 = |\psi_1\rangle\langle\psi_1|$, $\rho_2 = |\psi_2\rangle\langle\psi_2|$ with $\langle\psi_1|\psi_2\rangle = 0$, i.e., orthogonal states, then $\Pr(\text{success}) = 1$.

[9] C. W. Helstrom, "Quantum Detection and Estimation Theory" (Academic Press, New York, 1976).

1.2.3 3rd Postulate: Quantum channels

As you are probably already expecting, the third postulate of quantum mechanics can also be generalized. In a sense. We saw that a general deterministic transformation on quantum states is described by unitary operators: $U : \mathscr{H} \mapsto \mathscr{H}$ with $U^\dagger U = UU^\dagger = \mathbb{1}$.

In realistic scenarios, however, we have errors, losses, and various forms of loosing information about the system, and as such its evolution is no

longer deterministic. How to describe such evolutions within quantum mechanics? The main idea is to model the degrees of freedom that we don't have access to, and as such can cause errors, as another quantum system. We then get a deterministic unitary evolution for the composite system, but a stochastic evolution for the our system of interest.

In a general way, consider that we have a system composed of two parts, i.e., we assign to it a Hilbert space $\mathscr{H}_{AB} = \mathscr{H}_A \otimes \mathscr{H}_B$. The full system undergoes some unitary transformation $U : \mathscr{H}_{AB} \mapsto \mathscr{H}_{AB}$. Suppose that the system is initially described by the density matrix $\rho_{AB} = \rho_A \otimes |0\rangle\langle 0|$. This density matrix represents, for instance, the situation in which we prepare the system of interest in the state ρ_A and couple it to an environment, that without loss of generality, we can take as pure state (as we can always choose the dimension big enough as to purify it). The evolved state is then given by $U\left(\rho_A \otimes |0\rangle\langle 0|\right) U^\dagger$.

What is then the dynamics induced on the state of the system A only? To obtain the induced transformation on system A, that we denote by $\varepsilon : \mathscr{D}(\mathscr{H}_A) \mapsto \mathscr{D}(\mathscr{H}_A)$, we must "get rid" of the B system. This is done similarly to the case of joint probability of random variables, i.e., by summing over all possible configurations of the unwanted variable. Within quantum mechanics, this "summing over" is given by the partial trace.[10] Therefore, if we are interested only on the evolution of system A, the induced dynamics is given by

$$\varepsilon(\rho_A) = \mathrm{Tr}_B\left(U(\rho_A \otimes |0\rangle\langle 0|)U^\dagger\right).$$

Using the definition of partial trace we get

$$\begin{aligned}
\varepsilon(\rho_A) &= \sum_i (\mathbb{1} \otimes \langle i|)U(\rho_A \otimes |0\rangle\langle 0|)U^\dagger(\mathbb{1} \otimes |i\rangle), \\
&= \sum_i \left((\mathbb{1} \otimes \langle i|)U(\mathbb{1} \otimes |0\rangle)\right)\rho_A\left((\mathbb{1} \otimes \langle 0|)U^\dagger(\mathbb{1} \otimes |i\rangle)\right), \\
&:= \sum_i K_i \rho_A K_i^\dagger,
\end{aligned}$$

where $K_i := (\mathbb{1} \otimes \langle i|)U(\mathbb{1} \otimes |0\rangle)$ is an operator acting on system A, $K_i : \mathscr{H}_A \mapsto \mathscr{H}_A$, known as Kraus operator. Note that the Kraus operators are in general not unitary matrices, however

$$\begin{aligned}
\sum_i K_i^\dagger K_i &= \left((\mathbb{1} \otimes \langle 0|)U^\dagger(\mathbb{1} \otimes |i\rangle)\right)\left((\mathbb{1} \otimes \langle i|)U(\mathbb{1} \otimes |0\rangle)\right) \\
&= (\mathbb{1} \otimes \langle 0|)U^\dagger(\mathbb{1} \otimes \sum_i |i\rangle\langle i|)U(\mathbb{1} \otimes |0\rangle) \\
&= (\mathbb{1} \otimes \langle 0|)U^\dagger U(\mathbb{1} \otimes |0\rangle) \\
&= (\mathbb{1} \otimes \langle 0|)(\mathbb{1} \otimes |0\rangle) \\
&= \mathbb{1}_A.
\end{aligned}$$

A linear map $\varepsilon : \mathscr{L}(\mathscr{H}_A) \mapsto \mathscr{L}(\mathscr{H})_A$, defined by its Kraus operators $\{K_i\}$ which abide by $\sum_i K_i^\dagger K_i = \mathbb{1}$, observe a series of highly desirable properties of a physically allowed transformation:

[10] The partial trace over the space B is a linear opertator

$$\mathrm{Tr}_B : \mathscr{L}(\mathscr{H}_A \otimes \mathscr{H}_B) \mapsto \mathscr{L}(\mathscr{H}_A),$$

defined by

$$\begin{aligned}
\mathrm{Tr}_B(A \otimes B) &= \sum_{i=1}^{d_b} \mathbb{1} \otimes \langle i|(A \otimes B)\mathbb{1} \otimes |i\rangle \\
&= A\mathrm{Tr}(B),
\end{aligned}$$

for any $A \in \mathscr{L}(\mathscr{H}_A)$ and $B \in \mathscr{L}(\mathscr{H}_B)$. Clearly, we can define the partial trace over A in a similar fashion.

Trace preservation:

$$\begin{aligned}\mathrm{Tr}(\varepsilon(\rho_A)) &= \mathrm{Tr}(\sum_i K_i \rho_A K_i^\dagger),\\ &= \mathrm{Tr}(\sum_i K_i^\dagger K_i \rho_A),\\ &= \mathrm{Tr}(\rho_A).\end{aligned}$$

Hermiticity preservation:

$$\begin{aligned}(\varepsilon(\rho_A))^\dagger &= (\sum_i K_i \rho_A K_i^\dagger)^\dagger,\\ &= \sum_i K_i \rho_A^\dagger K_i^\dagger,\\ &= \varepsilon(\rho_A^\dagger).\end{aligned}$$

Positivity preservation: If $\rho_A \geq 0$, then for all $|\psi\rangle \in \mathscr{H}_A$ we have

$$\begin{aligned}\langle\psi|\varepsilon(\rho_A)|\psi\rangle &= \langle\psi|\sum_i K_i \rho_A K_i^\dagger|\psi\rangle,\\ &= (\langle\psi|\sum_i K_i\sqrt{\rho_A})(\sqrt{\rho_A}K_i^\dagger|\psi\rangle),\\ &= \|\sqrt{\rho_A}K_i^\dagger|\psi\rangle\|^2 \geq 0.\end{aligned}$$

Complete Positivity: Even when we are dealing with a subsystem of a larger system, the map ε preserves the positivity of operators. Mathematically, if $\rho \in \mathscr{D}(\mathscr{H}_A \otimes \mathscr{H}_Z)$, with $\mathscr{H}_Z$ the space assigned to the other parts composing the total system, then $\varepsilon \otimes \mathbb{1}_Z(\rho) \geq 0$, as we can easily check. For all $\psi \in \mathscr{H}_A \otimes \mathscr{H}_Z$

$$\begin{aligned}\langle\psi|\varepsilon\otimes\mathbb{1}(\rho)|\psi\rangle &= \langle\psi|\sum_i (K_i\otimes\mathbb{1})\rho(K_i^\dagger\otimes\mathbb{1})|\psi\rangle,\\ &= (\langle\psi|\sum_i (K_i\otimes\mathbb{1})\sqrt{\rho})(\sqrt{\rho}(K_i^\dagger\otimes\mathbb{1})|\psi\rangle),\\ &= \|\sqrt{\rho}(K_i^\dagger\otimes\mathbb{1})|\psi\rangle\|^2 \geq 0.\end{aligned}$$

All these properties show that the map ε maps quantum states into quantum states. Such maps are called *quantum channels* and are the most general physical transformation that we can perform on quantum systems. The generalized version of the third postulate reads as follows.

Third Postulate

Given a quantum system acting on the Hilbert space $\mathscr{H}$, described by the density matrix $\rho \in \mathscr{D}(\mathscr{H})$, the most general transformation that can be performed on the system is a map $\varepsilon : \mathscr{L}(\mathscr{H}) \mapsto \mathscr{L}(\mathscr{H}')$:

$$\varepsilon(\rho) = \sum_i K_i \rho K_i^\dagger,$$

where the operators $K_i : \mathscr{H} \mapsto \mathscr{H}'$ are such that $\sum_i K_i^\dagger K_i = \mathbb{1}_{\mathscr{H}}$.

1.3 Entanglement - formal introduction

We have already seen that for a quantum quantum system composed of two or more parts, we assign a global Hilbert space which is given by the tensor product of the individual Hilbert spaces. This allows for systems configurations that are not possible when we have the Cartesian product between the individual spaces, which is what we usually have between classical systems.[11] Here we are going to see who are these states, and see how they are related to a kind of correlations between particles that is intrinsically a quantum manifestation. In the following chapters we will explore in more depth this correlation, but now we attain to its structure.

[11] Note that there classical systems that are constructed with a tensor structure, and thus these "extra states", which we are going to call entangled, also may exist classically.

Let $|\psi\rangle \in \mathscr{H}_{AB}$ describe the state of a system composed of two parts, A and B. If we can write $|\psi\rangle$ as the product of a state in $\mathscr{H}_A$ with a state in $\mathscr{H}_B$,

$$|\psi\rangle = |\phi\rangle \otimes |\chi\rangle,$$

for $|\phi\rangle \in \mathscr{H}_A$ and $|\chi\rangle \in \mathscr{H}_B$, then there is no correlation between measurements performed separately in each particle. We can see this by picking a POVM acting in A, say $\{M_i^A\}$, and a POVM in B, say $\{M_j^B\}$, and evaluating the probability of obtaining the outcomes (i, j):

$$\begin{aligned}\Pr(i,j|\psi) &= \mathrm{Tr}(|\psi\rangle\langle\psi| M_i^A \otimes M_j^B) \\ &= \mathrm{Tr}(|\phi\rangle\langle\phi| M_i^A \otimes |\chi\rangle\langle\chi| M_j^B) \\ &= \Pr(i|\phi)\Pr(j|\chi).\end{aligned}$$

This means that states of the form $|\psi\rangle = |\phi\rangle \otimes |\chi\rangle$ are uncorrelated. Which then suggests the following definition:

Definition 1.3.1 (Separable states)**.** *A state $|\psi\rangle \in \mathscr{H}_A \otimes \mathscr{H}_B$ is said separable if $\exists |\phi\rangle \in \mathscr{H}_A$ and $|\chi\rangle \in \mathscr{H}_B$ such that*

$$|\psi\rangle = |\phi\rangle \otimes |\chi\rangle.$$

For example consider the case of two two-level systems (qubits), i.e., $\mathscr{H}_A = \mathscr{H}_B \sim \mathbb{C}^2$. Is the state

$$|\psi_1\rangle = \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$$

separable? The answer is Yes, as we can write the state as

$$|\psi_1\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle + |1\rangle}{\sqrt{2}}.$$

What about the state

$$|\psi_2\rangle = \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle - |11\rangle),$$

is it separable? In this case we have

$$|\psi_2\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle \otimes \frac{|0\rangle + |1\rangle}{\sqrt{2}} + |1\rangle \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}}\right).$$

As the states in the B part are different (in fact in this case they are orthogonal) we can not factor it out, and the state is not separable.

Definition 1.3.2 (Entangled state)**.** *If a pure state $|\psi\rangle \in \mathscr{H}_A \otimes \mathscr{H}_B$ is not separable, then it is said to be entangled.*

Suppose then that we have an entangled state $|\psi\rangle_{AB} \in \mathscr{H}_A \otimes \mathscr{H}_B$. We want to show that such states allow for correlations between measurements on the individual parts. To do that more explicitly, we write $|\psi\rangle$ in its Schmidt decomposition[12]

$$|\psi\rangle = \sum_i \sqrt{\lambda_i}|u_i\rangle \otimes |v_i\rangle,$$

where the $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$, $\{|u_i\rangle\}$ and $\{|v_i\rangle\}$ form an orthonormal basis for $\mathscr{H}_A$ and $\mathscr{H}_B$, respectively. Note that as we are using orthonormal basis for both spaces, then any entangled state when written in its Schmidt decomposition has at least two non-zero λ_i coefficients. Proceeding as before, we take a POVM acting in A, say $\{M_i^A\}$, and a POVM in B, say $\{M_j^B\}$, and evaluate the probability of obtaining the outcomes (i, j):

$$\begin{aligned}
\Pr(i,j|\psi) &= \mathrm{Tr}(|\psi\rangle\langle\psi| M_i^A \otimes M_j^B) \\
&= \sum_{n,m} \sqrt{\lambda_n \lambda_m}\, \mathrm{Tr}(|u_n v_n\rangle\langle u_m v_m| M_i^A \otimes M_j^B) \\
&= \sum_{n,m} \sqrt{\lambda_n \lambda_m}\, \mathrm{Tr}(|u_n\rangle\langle u_m| M_i^A \otimes |v_n\rangle\langle v_m| M_j^B) \\
&= \sum_{n,m} \sqrt{\lambda_n \lambda_m}\, \mathrm{Tr}(|u_n\rangle\langle u_m| M_i^A)\mathrm{Tr}(|v_n\rangle\langle v_m| M_j^B),
\end{aligned}$$

which cannot be written as the product of probabilities for the local measurements whenever we have at least two non-zero λ_i's, i.e., whenever the state is entangled. Entanglement is thus some sort of correlation.

The above result immediately implies that entangled pure states have mixed local density matrices. To see this take again $|\psi\rangle = \sum_i \sqrt{\lambda_i}|u_i\rangle \otimes |v_i\rangle$, then

$$\begin{aligned}
\rho_A &= \mathrm{Tr}_B |\psi\rangle\langle\psi| \\
&= \mathrm{Tr}_B \left(\sum_{n,m} \sqrt{\lambda_n \lambda_m} (|u_n v_n\rangle\langle u_m v_m| \right) \\
&= \sum_i \lambda_i |u_i\rangle\langle u_i|.
\end{aligned}$$

As for entangled states we have at least two non-zero λ_i's, and $\sum_i \lambda_i = 1$, then $\mathrm{Tr}(\rho_A^2) < 1$. For pure entangled states, only the global system is well defined. In the words of Schrödinger:

> The maximum knowledge of the whole, does not necessarily include the maximum knowledge of its parts.

The scenario in which the global state is already mixed is a bit more complicated, and it goes beyond the scope of these notes. To the interested reader, we recommend the review article by the Horodecki family.[13]

[12] Let $\mathscr{H}_A$ and $\mathscr{H}_B$ be Hilbert spaces of dimensions d_A and d_B respectively. For any vector $w \in \mathscr{H}_A \otimes \mathscr{H}_B$, there exist orthonormal sets $\{u_1, \ldots, u_{d_A}\} \subset \mathscr{H}_A$ and $\{v_1, \ldots, v_{d_B}\} \subset \mathscr{H}_B$ such that $w = \sum_{i=1}^{\min(d_A, d_b)} \alpha_i u_i \otimes v_i$, where the scalars α_i are non-negative reals.

[13] Quantum entanglement, Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. Rev. Mod. Phys. **81**, 865 (2009).

1.3.0.1 No go theorem: signaling

One common misconception about quantum entanglement is that it allows for instantaneous communication at distance. Einstein called this possibility as "spooky action at distance". This, however, is not correct.

It is not difficult to realize that quantum mechanics is a local theory. Consider a quantum system formed by two particles, described by a density matrix $\rho_{AB} \in \mathscr{D}(\mathscr{H}_A \otimes \mathscr{H}_B)$ – possibly an entangled state. Now, for the sake of the argument, imagine that these two particles are very far from each other, and we measure an observable $\mathscr{O}$ in particle A while nothing is done in particle B. It is reasonable to suppose that the expectation value of $\mathscr{O}$ should not depend on particle B, i.e., it should depend only on the reduced state of particle A, $\rho_A = \mathrm{Tr}_B \rho_{AB}$. And that is indeed the case, as

$$\langle \mathscr{O} \rangle = \mathrm{Tr}(\mathscr{O} \otimes \mathbb{1} \rho_{AB}).$$

Writing $\rho_{AB} = \sum_{i,j} \sum_{k,l} \rho_{k,l}^{i,j} |i,j\rangle\langle k,l|$ we have:

$$\begin{aligned}
\langle \mathscr{O} \rangle &= \sum_{i,j} \sum_{k,l} \rho_{k,l}^{i,j} \langle k,l|(\mathscr{O} \otimes \mathbb{1})|i,j\rangle, \\
&= \sum_{i,j} \sum_{k,l} \rho_{k,l}^{i,j} \langle k|\mathscr{O}|i\rangle\langle l|j\rangle, \\
&= \sum_{i,j} \sum_{k} \rho_{k,j}^{i,j} \langle k|\mathscr{O}|i\rangle, \\
&= \mathrm{Tr}\left((\sum_{i,j} \sum_{k} \rho_{k,j}^{i,j} |i\rangle\langle k|) \mathscr{O} \right).
\end{aligned}$$

To conclude, we just need to realize that

$$\begin{aligned}
\rho_A &= \mathrm{Tr}_B \rho_{AB}, \\
&= \mathrm{Tr}_B (\sum_{i,j} \sum_{k,l} \rho_{k,l}^{i,j} |i,j\rangle\langle k,l|), \\
&= \sum_{i,j} \sum_{k} \rho_{k,j}^{i,j} |i\rangle\langle k|.
\end{aligned}$$

And thus $\langle \mathscr{O} \rangle = \mathrm{Tr}(\mathscr{O} \rho_A)$, as we expected. Experiments performed on particle A do not depend on the state of particle B.

This does not mean that there cannot be correlation between the results of local experiments. This correlation, however, cannot be used to transmit information instantaneous. To transmit some information from the side A to side B, the probabilities on one side must depend on experiments performed on the other side. The above calculations can be quickly extended to prove that quantum mechanics is *no-signallig*.

Consider again the scenario of two distant labs. Lab A is run by Amary, and lab B is run by Bacuara. Amary has at her disposal a setup that can perform a set $\mathscr{X} = \{1,2,\ldots\}$ of different experiments (if $x \in \mathscr{X}$ is 1 then she performs the first type of experiment, if $x = 2$ she performs the second

type of experiment, and so it goes). To simplify, assume that independently of the experiment she chooses, the set of possible outcomes is always the same, and we call it $\mathscr{A}$. In this way, an experiment $x \in \mathscr{X}$ can have, with some probability, the outcome $a \in \mathscr{A}$. Similarly, Bacura can make experiments $y \in \mathscr{Y}$ and obtain outcomes $b \in \mathscr{B}$. A theory is said to be non-signalling if it satisfies:

$$\sum_{b\in\mathscr{B}} \Pr(a,b|x,y) = \sum_{b\in\mathscr{B}} \Pr(a,b|x,y') \qquad \forall\, a \in \mathscr{A}, x \in \mathscr{X}, \text{and } y,y' \in \mathscr{Y};$$
$$\sum_{a\in\mathscr{A}} \Pr(a,b|x,y) = \sum_{a\in\mathscr{A}} \Pr(a,b|x',y) \qquad \forall\, b \in \mathscr{B}, y \in \mathscr{Y}, \text{and } x,x' \in \mathscr{X}.$$

What do these equations mean? Take the first set of equations. From probability theory we know that by summing over all outcomes of a measurement we get the marginal distribution:

$$\sum_{b\in\mathscr{B}} \Pr(a,b|x,y) = \Pr(a|x,y).$$

This is a mathematical result. From the other side we get

$$\sum_{b\in\mathscr{B}} \Pr(a,b|x,y') = \Pr(a|x,y').$$

From the equality imposed by the non-signalling constraints, we get

$$\Pr(a|x,y) = \Pr(a|x,y') \qquad \forall\, a \in \mathscr{A}, x \in \mathscr{X}, \text{and } y,y' \in \mathscr{Y}.$$

For the experiment x performed by Amary to output a, it does not matter what Bacuara is measuring. There is no way that she will know then what he is measuring, and as such he cannot use the choice of what experiment to make in order to send an information. We get a similar conclusion from the other set of equations.

Now that we understand what a non-signalling theory must satisfy we want to show that quantum mechanics abides by these rules.

Theorem 3. *Quantum mechanics is a non-signalling theory.*

Proof. The only thing that we have to show is that the way quantum mechanics evaluates probabilities of local measurements satisfies the non-signalling constraints. Let $\{M_a^x\}$ be the family of POVMs associated with Amary's measurements, and $\{M_b^y\}$ be the family of POVMs associated with Bacuara's measurements. Furthermore, assume that they share a bipartite quantum system described by ρ_{AB}, which like before may be entangled or not. Then:

$$\begin{aligned}
\sum_{b\in\mathscr{B}} \Pr(a,b|x,y') &= \sum_{b\in\mathscr{B}} \mathrm{Tr}(\rho_{AB}\, M_a^x \otimes M_b^y), \\
&= \mathrm{Tr}(\rho_{AB}\, M_a^x \otimes \sum_{b\in\mathscr{B}} M_b^y), \\
&= \mathrm{Tr}(\rho_{AB}\, M_a^x \otimes \mathbb{1}), \\
&= \mathrm{Tr}(\rho_A\, M_a^x), \\
&= \Pr(a|x).
\end{aligned}$$

In this calculations we simply used the linearity of the trace, and that for any POVM $\sum_{b\in\mathscr{B}} M_b^y = \mathbb{1}$. We see from the result that there is no dependence on the measurement performed by Bacuara, and as such the first set of no-signalling constraints are fulfilled. An identical reasoning shows the second set of non-signalling constraints. Quantum theory is thus non-signalling, and no instantaneous information can be sent. □

2 Quantum state

Now that we've established the structure of quantum theory, we are going to analyze it more in depth. We start by scrutinizing the idea of a quantum state. What does it mean? Does it have a reality attached to it, or is it related to the knowledge we have about the system?

We are going to address these questions in different ways within this notes. Here we start by establishing some similarities and differences between classical and quantum states. Classical pure states are points in phase space, i.e., probability densities corresponding to delta functions in phase space, with well defined position and momentum for all the particles. Once positions and momenta are know, any other physical property can be evaluated. Classical mixed states are probabilistic mixtures – convex combinations – of these points, leading to stochastic values for the physical properties. This probabilistic aspect of classical mechanics comes only from our ignorance about the system's state or dynamics, and it is not intrinsic to the theory. In the following we see how this classical description compares to the notion of a quantum state.

2.1 No-cloning theorem

Possibly the easiest result that already has some implication on the nature of quantum states is the no-cloning theorem. This result was first obtained by Wootters and Zurek,[1] and later simplified by various authors. As the theorem's name indicates, it states that quantum states cannot in general be copied.

[1] "A Single Quantum Cannot be Cloned", William Wootters and Wojciech Zurek . Nature **299** 802 (1982).

Theorem 4 (No-cloning theorem.)**.** *There is no unitary U acting on $\mathscr{H}\otimes\mathscr{H}$ such that for all $|\phi\rangle\in\mathscr{H}$ we have*

$$U|\phi\rangle\otimes|0\rangle=|\phi\rangle\otimes|\phi\rangle,$$

where $|0\rangle$ is some fixed normalized state in the second copy of $\mathscr{H}$.

Proof. The proof goes by the way of contradiction. Suppose that such cloning unitary does exist. Then for two states $|\phi\rangle$ and $|\psi\rangle$ in $\mathscr{H}$, it must be

the case that:

$$U|\phi\rangle \otimes |0\rangle = |\phi\rangle \otimes |\phi\rangle$$
$$U|\psi\rangle \otimes |0\rangle = |\psi\rangle \otimes |\psi\rangle.$$

Taking the scalar between the two above expressions we get

$$(\langle\phi| \otimes \langle 0|U^\dagger)(U|\psi\rangle \otimes |0\rangle) = (\langle\phi| \otimes \langle\phi|)(|\psi\rangle \otimes |\psi\rangle),$$

which thus leads to

$$\langle\phi|\psi\rangle\langle 0|0\rangle = \langle\phi|\psi\rangle^2.$$

As $\langle 0|0\rangle = 1$, the possible solutions are: either $\langle\phi|\psi\rangle = 1$, and then $|\phi\rangle$ and $|\psi\rangle$ are parallel; or $\langle\phi|\psi\rangle = 0$ and the states must be orthogonal. As by hypothesis the states $|\phi\rangle$ and $|\psi\rangle$ are arbitrary, they are not necessarily parallel or orthogonal, and we then reach a contradiction. Therefore there is no such unitary that clones an arbitrary quantum state. □

Note that the key ingredient in this proof was the linearity of quantum mechanics. Above we used an unitary evolution, but it would make no difference if we had used more general quantum channels, as these are also linear. Given that classical mechanics is also linear, a similar result also exists for classical dynamics.[2] The main difference between the quantum and classical result is that within quantum mechanics we have the possibility of superposing pure states, while in classical mechanics we only have convex mixtures of pure states – within classical mechanics, the pure states have well defined position and momentum for all particles, i.e., delta distributions in the phase space. In this way the deterministic states in classical mechanics are all orthogonal to each other, which can then be cloned. Within quantum mechanics the pure states are not necessarily orthogonal, and that is what makes the no-cloning theorem interesting.

[2] "Classical No-Cloning Theorem", A. Daffertshofer, A. R. Plastino, and A. Plastino. Phys. Rev. Lett. **88**, 210601 (2002).

Even more generally, probability distributions can not be cloned by linear processes. This may suggest that a quantum states should be though as some sort of probabilistic description of the knowledge we have about a system. In fact, if one attaches a reality to quantum states, like in an ontic complete interpretation does, it comes as surprise that quantum states cannot be cloned. Note however that there are realist interpretations of quantum mechanics, like Bohmian mechanics, that although realist they are not ontic complete, and as such this non-clonability may not be that strange.

2.2 Preparation of quantum states

The first postulate of quantum mechanics assigns a Hilbert space to a quantum system, and describes the system state by a density matrix acting on this space. The choice of the Hilbert space is possibly not that complicated, depending the number of orthogonal states allowed to the system. For example, if one has a spin 1/2 particle, with spin-up and spin-down as possible

orthogonal states, it is reasonable to assign to this system a bi-dimensional Hilbert space, $\mathbb{C}^2$. But how to decide which description should we give to the system?

In classical mechanics the state of the system is defined by the positions and momenta of the particles. As these are physical properties, the way to determine them is by measuring. Our description of the system is then updated with the data we obtain about it. If the state is seeing as description of the knowledge we have about the system, this updating process is known as Bayesian inference.

Within quantum mechanics it is the same. We must perform measurements in order to prepare a quantum state. The main difference here is due to the commutation relation among the observable quantities.

For instance, suppose we have a spin 1 particle. The assigned Hilbert space is a three dimensional one, and we take $\{|0\rangle, |1\rangle, |2\rangle\}$ as an orthonormal basis. If a machine sending i.i.d spin one particles, and we have no further information about its inner workings, which state should we assign to the system? At this initial moment we no information and we thus take a unknown description ρ. Imagine now that I can measure the observable

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

written in the basis $\{|0\rangle, |1\rangle, |2\rangle\}$. If we get the outcome $+1$, we know for sure that after the measurement the system is prepared in the state $|0\rangle$. However, as this observable is degenerated, if we get the outcome -1 we can not be sure about the state of the system. If no further information is present, the best description is to assign to the system the density matrix

$$\rho = \frac{1}{2}|1\rangle\langle 1| + \frac{1}{2}|2\rangle\langle 2|.$$

This is the description that corresponds to knowledge so far obtained about the system.

If got the outcome -1, and I still want to prepare a state which has a well defined property A, then I must measure another observable B that commutes with A and that splits the degeneracy in the -1 subspace. This is achieved, for example, by the observable

$$B = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

also written in the basis $\{|0\rangle, |1\rangle, |2\rangle\}$. As both observables are diagonal in the same basis, they commute to each other. It is a well known result in linear algebra, that if two observables commute, $[A,B] := AB - BA = 0$, then there exists a common base in which both operators are diagonal. In the present

example, observable B has different eigenvalues in the subspace spanned by $\{|1\rangle, |2\rangle\}$. As such, if after an outcome -1 for the A measurement, I get now, say, 1, I've prepared the state $|2\rangle$. The table below summarizes this preparation procedure.

A	B	prepared state
1	2	$\|0\rangle$
-1	2	$\|1\rangle$
-1	1	$\|2\rangle$

Therefore, by measuring these two commuting observables we are able to prepare a completely determined state of the system. Such a state has the two properties, A and B, well defined.

The preparation of a quantum state is then given by the measurement of a Complete Set of Commutable Operators (CSCO).

Definition 2.2.1 (CSCO)**.** *A set of observables $A, B, C, \ldots$ acting on $\mathscr{H}$ is called a CSCO if two things happen:*

i) All the observables commute pairwise;

ii) specifying the eigenvalues of all the operators determines a unique (to within a multiplication factor) common eigenvector.

2.3 Uncertainty principles

Once we prepared a quantum state, what physical properties have a well defined value? Remember that within classical mechanics, whenever we prepare a state, all the physical quantities are precisely defined. Within quantum mechanics, however, the value of a physical property is detached from the property itself. That is, to some extent, the content of the famous Heisenberg uncertainty relations.

Remember that a state has a well defined value of a given physical property, if it is an eigenvector of the associated observable. Mathematically speaking, given a property $\mathscr{C}$, with associated observable in its spectral decomposition given by $C = \sum_i c_i |c_i\rangle\langle c_i|$, then $|\psi\rangle$ has a well defined value of $\mathscr{C}$ if $C|\psi\rangle = c_i|\psi\rangle$ for some c_i. If C is non-degenerated, then $|\psi\rangle = |c_i\rangle$. Experimentally, such a situation represents a measurement of $\mathscr{C}$ without any dispersion; we always get the same outcome. Clearly, if we were also able to measure another observable D that commutes with C, no dispersion would be observed as well.

Now, suppose that we prepared the state $|\psi\rangle$ by measuring a CSCO that contains C and D. What can we say about the dispersion on the measurement of two other observables, say, A and B? In general the situation will be like the one we show in Fig. 2.1.

The experimental procedure to obtain these histograms is as follows: First we prepare a large number of copies of $|\psi\rangle\langle\psi|$, say $N_c \gg 1$. We then divide

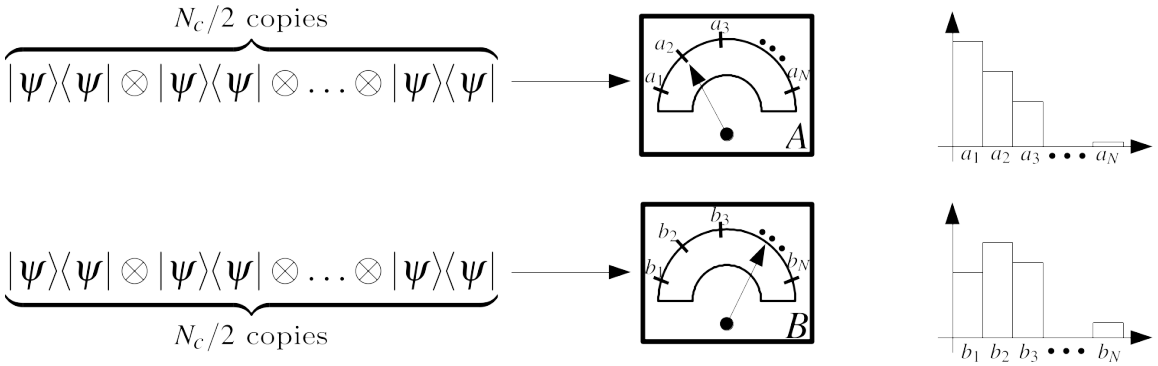

Figure 2.1: Experimental scenario related with the preparation uncertainty relations.

these copies into two sets of about $N_c/2$ copies each. For each copy in the first set we measure A. For each copy in the second set we measure B.

2.3.1 Heisenberg Uncertainty Relation (HUR)

The HUR puts a lower bound on the product of the variances associated with the distribution of outcomes:

$$\sigma_A(\psi)\sigma_B(\psi) \geq \frac{1}{2}|\langle\psi|[A,B]|\psi\rangle|, \tag{2.1}$$

where $\sigma_A(\psi) = \sqrt{\langle\psi|(A-\langle A\rangle_\psi)^2|\psi\rangle}$, and similarly to σ_B. In a sense, the HUR tries to lower bound the variance of one observable, given the variance of another observable.

It is important to notice that the HUR does not say anything about the possibility of simultaneous measurement of two physical properties. It does say something about the possibility of a given state to have well defined values for two different properties. As we can see, a state $|\psi\rangle$ has well defined A and B properties, if and only if, it is an eigenvector of at least one observable – A and B don't even need to commute.[3] The proof of the HUR can be found in any quantum mechanics book. Here, however, we want alert to some issues with the HUR.

i) *The lower bound depends on the state.*
As it is clear from Eq. 2.1, the left-hand-side (l.h.s.) does depend on the state that we are looking at. This means that if $[A,B] \neq 0$ and for a generic $|\phi\rangle$ (not an eigenvector of A neither of B), both σ_A and σ_B are greater than zero and are related by the HUR as

$$\sigma_A \geq \frac{1}{2\sigma_B}|\langle[A,B]\rangle_\phi|.$$

There is thus a relation between the two variances. However, if we change the state $|\phi\rangle$, leading to a decrease of σ_B, that does not necessarily implies an increase of σ_A. The reason is because $|\langle[A,B]\rangle_\phi|$ may also decrease as we change $|\phi\rangle$.

[3] To see that, suppose that $A|\psi\rangle = a_i|\psi\rangle$, then

$$\begin{aligned}\langle\psi|[A,B]|\psi\rangle &= \langle\psi|(AB-BA)|\psi\rangle \\ &= a_i\langle\psi|B|\psi\rangle - \langle\psi|B|\psi\rangle a_i \\ &= 0.\end{aligned}$$

Moreover, as we already pointed out, the l.h.s. can be zero even when the observables do not commute.
The HUR is only meaningful for canonically conjugated variables in infinite dimensions, like for position and momentum. In this case we have $[X,P] = i\hbar$, and the state dependence on the l.h.s. disappears: $\sigma_X(\psi)\sigma_P(\psi) \geq \frac{\hbar}{2}$.

ii) *The lower bound depends on the observables eigenvalues.*
If we make $A \to \alpha A$, the lower bound is going to be multiplied by α. Taking $\alpha \ll 1$ will make the lower bound to almost vanish, despite the fact that the distribution of outcomes did not change.

2.3.2 Entropic Uncertainty Relation (EUR)

To overcome the issues mentioned above, in the recent literature it was introduced another type of uncertainty relation. This time, instead of using the variance of the outcomes' distribution, one uses the Shannon entropy of the outcomes' probability distribution.

Given a probability distribution $\{p_i\}_{i=1}^N$, its Shannon entropy is defined as

$$H(\{p_i\}) = -\sum_{i=1}^{N} p_i \log_2 p_i. \tag{2.2}$$

Using the logarithm in base 2, the unit of Shannon's entropy is the *bit*. Like the variance, the entropy also quantifies the dispersion of the distribution of outcomes.

Now, going back to our observables $A = \sum_{i=1}^N a_i |a_i\rangle\langle a_i|$ and $B = \sum_{i=1}^N b_i |b_i\rangle\langle b_i|$ (assumed non-degenerated), the postulates of quantum mechanics tell us that for a given preparation $|\phi\rangle$

$$\Pr(a_i|\phi) = |\langle a_i|\phi\rangle|^2,$$
$$\Pr(b_i|\phi) = |\langle b_i|\phi\rangle|^2.$$

The Shannon entropy associated to these distributions can be easily evaluated:

$$H(A|\phi) := -\sum_{i=1}^{N} |\langle a_i|\phi\rangle|^2 \log_2(|\langle a_i|\phi\rangle|^2),$$
$$H(B|\phi) := -\sum_{i=1}^{N} |\langle b_i|\phi\rangle|^2 \log_2(|\langle b_i|\phi\rangle|^2).$$

A relation between the two entropies was first suggested by Deutsch (1983) and later proved by Massen and Uffink (1988), and it reads:

$$H(A|\phi) + H(B|\phi) \geq -2\log_2 \max_{i,j} |\langle a_i|b_j\rangle|. \tag{2.3}$$

Like for the HUR, here we also obtain a lower bound for the "dispersions". Here, however, the lower bound does not depend neither on the state, nor

on the observables' eigenvalues. Furthermore, an similar expression can be found for the case of generalized measurements, i.e., for POVMs. A large literature is already available on these entropic relations.[4]

[4] See for instance the review "Entropic uncertainty relations and their applications" by Coles et al. RMP **89**, 015002 (2017).

In conclusion we see that the quantum state cannot be given a realistic complete interpretation. On the other hand, the quantum state does not contain all the information about the system. As such, a epistemological view that assigns to the quantum state the knowledge we have about the system is also not possible. So, how to interpret the quantum state? Don't answer just now. In the next chapters more constraints and information will be given.

3 Quantum Measurement Problem

Measurement is an integral part of science. Despite of that, there is little consensus among philosophers and practitioners about the meaning of measurement.[1] If within classical mechanics we can live with this ambiguity, within quantum mechanics this cloudiness becomes more striking.

[1] See for instance Eran Tal, "Measurement in Science", The Stanford Encyclopedia of Philosophy (Fall 2017 Edition), Edward N. Zalta (ed.).

3.1 Where the problem is, and where it is not

In quantum mechanics measurements are somehow described by its 3rd postulate, but it also gets intertwined with the other postulates.

3.1.1 Small problem: a single outcome happens

Given a preparation $|\psi\rangle$, which can for example describe a single particle just before a position detector, the third postulate tells us that the probability (density) of measuring the particle between x and $x+dx$ is given by $|\langle x|\psi\rangle|^2$. The particle, eventually, is detected in a single position, despite the fact that its wave function was spread all over the detector (possibly even further away). If one gives complete reality to the wave function, it may be difficult to explain how this instantaneous collapse of the wave function happens.

A bit more formally, suppose we are measuring an observable $A = \sum_{i=1}^{d} a_i |a_i\rangle\langle a_i|$, with $\langle a_i|a_j\rangle = \delta_{ij}$ and $a_i \in \mathbb{R}$ for all $i, j \in [d]$ (its spectral decomposition). If the system is described by the state $|\psi\rangle$ just before the measurement, we can write it in the observable eigenbasis as

$$|\psi\rangle = c_1|a_1\rangle + c_2|a_2\rangle + \ldots + c_d|a_d\rangle,$$

with $c_i \in \mathbb{C}$ for all $i \in [d]$, and such that $\sum_{i=1}^{d} |c_i|^2 = 1$.

If we get the value, say a_2, which happens with probability $|c_2|^2$, the system evolves as

$$|\psi\rangle = \sum_{i=1}^{d} c_i|a_i\rangle \longrightarrow |a_2\rangle.$$

If we now measure the system again, we again will obtain the value a_2. This is consistent with the idea of a measurement. Nevertheless, this instantaneous collapse may look strange at first glance.

This is sometimes called the "small problem" of quantum measurement.[2]

[2] Časlav Brukner,"On the quantum measurement problem". Proceedings of the Conference "Quantum UnSpeakables II: 50 Years of Bell's Theorem" (Vienna, 19-22 June 2014).

The reason is because various interpretations do not suffer from this problem. The Bohiam interpretation, although realistic, the wave equation enters as "guiding the particles" in configuration space. The many-worlds interpretation avoids the problem by stating that all possibilities actually happen in some universe. Epistemic interpretations do not take a realistic view for the wave function and thus no problem is present – somehow it is similar to the classical case where we have a probability distribution for a coin toss, but only one result happens.[3] It is then clear that this part of the quantum measurement problem is possibly more related to the notion of quantum state than with the actual measurement dynamics.

[3] For a quick discussion about how the different interpretations deal with the small problem of quantum measurement, see for instance Wayne Myrvold, "Philosophical Issues in Quantum Theory", The Stanford Encyclopedia of Philosophy (Fall 2018 Edition), Edward N. Zalta (ed.).

3.1.2 Big problem: what makes a measurement a measurement?

If the small problem of quantum mechanics mixes the first and third postulates, the big problem ties up the second and the third postulates.

Like it is shown above, the third postulate states that when a measurement happens a state written as the superposition of various eigenstates of the observable being measured "collapses" into a single eigenvector. This is clearly not an unitary dynamics as stated by the second postulate. There is thus an ambiguity in which dynamical rule to use in each case. What makes a beam-splitter an unitary, and a photodiode a measurement apparatus?

For historical reasons, this "big problem" was not really a problem back at the time. Quantum mechanics was first designed around 1900, and at that time there was no question about what was the quantum system and what was the measurement apparatus. Measurement apparatus were described by classical mechanics, and as such they were out of the quantum formalism. Possibly as a free-thinking exercise, Schrödinger's cat thought experiment exposed the bizarre consequences we would face if quantum mechanics is taken as an universal theory, i.e., supposed to describe the micro and the macro worlds.

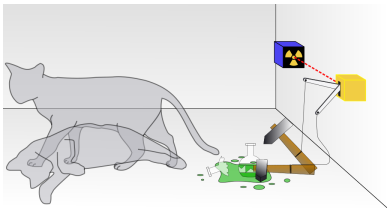

Figure 3.1: Schrödinger's cat thought experiment. Public domain picture from Wikipedia.

In the (in)famous thought experiment, a cat is put inside a box together with a, as Schrödinger said, "diabolical device". This device consists of a Geiger counter, and a small portion of an unstable radioactive substance. If the Geiger counter detects a decay, a hammer breaks a flask with a poisonous substance that kills the cat. See Fig.3.1.

A single atom is most well described by quantum mechanics. From this description, using the third postulate, one expects that atom that starts in a excited state, $|e\rangle$, to evolve at some time to a superposition of excited and ground (decayed), $c_e|e\rangle + c_g|g\rangle$ – with $c_e, c_g \in \mathbb{C}$ and such that $|c_e|^2 + |c_g|^2$. If we extend the quantum mechanical description to the cat, the dynamics of the two systems get correlated, entangled. After some time we expect the system to be described by:

$$c_e|e\rangle|\text{cat alive}\rangle + c_g|g\rangle|\text{cat dead}\rangle.$$

As the state "cat alive" is distinguishable from the state "cat dead", we

assume then orthogonal. For all values of $|c_e|$ different from 0 or 1 we have an entangled state between a microscopic system, the atom, and a macroscopic system, the cat. in our daily experiences we don't get to see cats in a superposition of dead-and-alive. Assuming quantum mechanics to be an universal description takes us to situations that we do not see in Nature. Schrödinger himself downplayed the problem saying that

> It is fair to state that we are not experimenting with single particles, any more than we can raise Ichthyosauria in the zoo.

In Schrödinger's example, the cat plays the role of a measurement apparatus, and the atom is the system to be measured. Bohr and Einstein had many discussions on where, or whether, to put the divide between the quantum and classical description.

It was however just in the 1980's that experiments in the area known as quantum optics were able to control and manipulate single atoms and single photons. In fact, the 2012 Nobel Prize was awarded to Serge Haroche and David J. Wineland for "ground-breaking experimental methods that enabled measuring and manipulation of individual atoms.[4]

In the experiment described in [5] the authors show how a single two-level atom can be measured by a coherent state of light. The state they are able to produce looks like

$$c_e|e\rangle|\alpha\rangle + c_g|g\rangle|-\alpha\rangle.$$

The coherent state has quasi-classical properties, and for $|\alpha| \gg 0$ one can show that $\langle\alpha|-\alpha\rangle \to 0$. In the actual experiment,[6] the average number of photons composing the "measurement apparatus" was $|\alpha| = 3.1$, which is already sufficient to have $\langle\alpha|-\alpha\rangle \approx 0$. It is, however, fair to say that it is not a superposition of what we would call a macroscopic state. It is far from a cat.

In the recent years, with the advance quantum technologies – specially driven by the race to build a quantum computer –, we have been reaching quantum control of fairly large systems. The question "what makes a measurement a measurement?" is more than ever in the agenda.

[4] S. Haroche, Rev. Mod. Phys. 85, 1083 (2013),
D. J. Wineland, Rev. Mod. Phys. 85, 1103 (2013).

[5] "Manipulation of photons in a cavity by dispersive atom-field coupling: Quantum-nondemolition measurements and generation of Schrödinger cat states", M Brune, Serge Haroche, JM Raimond, L Davidovich, N Zagury. PRA **45** 5193, (1992).

[6] Observing the Progressive Decoherence of the "Meter" in a Quantum Measurement M. Brune, E. Hagley, J. Dreyer, X. Maître, A. Maali, C. Wunderlich, J. M. Raimond, and S. Haroche Phys. Rev. Lett. **77**, 4887 (1996).

3.2 Formal treatment

If one is to accept that quantum mechanics is universal, the measuring process should be described by an interaction between the system to be measured and the measuring apparatus. The first model of a measurement within the quantum formalism was put forward by von Neumann,[7] and it is the basis of our description up to today.

Suppose we want to measure the projection of a spin-1/2 in the z direction. To do that we can couple to the system a "macroscopic" apparatus that depending on the particle's spin in the z direction it generates a large angular momentum in a corresponding direction. Mathematically, our toy model has

[7] J. von Neumann. Matematische Grundlagen der Quantenmechanic. Springer, Berlin, 1932.

an interaction Hamiltonian like:

$$H = \hbar g \sigma_z \otimes \frac{J_y}{|J|}.$$

Here g is a coupling constant (units of angular frequency), σ_z is the third Pauli matrix acting on the spin 1/2 system, J_x is the angular momentum operator for the apparatus, and $|J|$ is the maximum value the angular momentum can take – in order to have a bounded Hamiltonian in the limit $|J| \to \infty$.

When we want to measure the spin system we turn on the interaction. Let the spin state at this moment be written as $c_0|0\rangle + c_1|1\rangle$, with $c_0, c_1 \in \mathbb{C}$ such that $|c_0|^2 + |c_1|^2 = 1$, and $\sigma_z|0\rangle = |0\rangle$, $\sigma_z|1\rangle = -|1\rangle$. Moreover, let the initial state of the apparatus be the eigenvector of J_x with eigenvalue zero, that we denote by $|0_x\rangle$. The evolution is then:

$$\begin{aligned} e^{-\mathrm{i}\frac{Ht}{\hbar}}\left[(c_0|0\rangle + c_1|1\rangle)\otimes|0_x\rangle\right] &= e^{-\mathrm{i}\frac{gt}{|J|}\sigma_z\otimes J_y}\left[(c_0|0\rangle + c_1|1\rangle)\otimes|0_x\rangle\right] \\ &= c_0|0\rangle\otimes e^{-\mathrm{i}\frac{gt}{|J|}J_y}|0_x\rangle + c_1|1\rangle\otimes e^{\mathrm{i}\frac{gt}{|J|}J_y}|0_x\rangle, \\ &= c_0|0\rangle\otimes R_y\left(\frac{\hbar g t}{|J|}\right)|0_x\rangle + c_1|1\rangle\otimes R_y\left(-\frac{\hbar g t}{|J|}\right)|0_x\rangle, \end{aligned}$$

where $R_y(\theta)$ is the rotation operator about the y axis. We then see that depending on the particle's spin, the interaction Hamiltonian will generate rotations in opposite directions. See Fig.3.2.

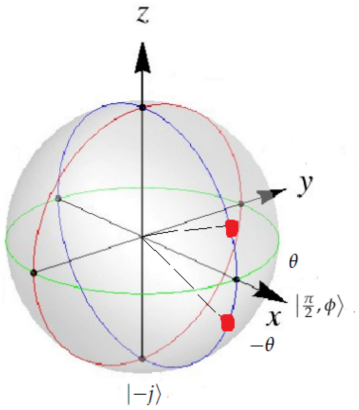

Figure 3.2: Quantum angular momentum space. Spin-dependent rotation of the angular momentum state. Picture taken from Gabriel D. Carvalho PhD thesis: "Emerging dynamics and its application in attempts to model a quantum measurement process".

As time passes by the two alternatives in the measuring apparatus turn distinguishable. Note that the larger $|J|$ is, the longer it takes to go through the same angle. This can be realized by noting that $\theta = \hbar g t/|J|$. This is compensated, however, by the fact that number of orthogonal states also increases with $|J|$. In the limit of $|J| \to \infty$ all the points turn orthogonal to each other, like a classical state space.

Rewriting the state after the interaction as

$$c_0|0\rangle\otimes|\theta\rangle + c_1|1\rangle\otimes|-\theta\rangle,$$

and assuming a time long enough such that $\langle\theta|-\theta\rangle \approx 0$, we see that by measuring the apparatus we get information about the system. The probability of measuring θ is equal to $|c_0|^2$; while the probability of measuring $-\theta$ is equal to $|c_1|^2$.

Like in Schrödinger's cat example we ended up with a macroscopic superposition, which we do not expect to observe in day-to-day life. About this point Bell made the following remark:[8]

> What exactly qualifies some physical systems to play the role of 'measurer'? Was the wave-function of the world waiting to jump for thousands of millions of years until a single-celled living creature appeared? Or did it have to wait a little longer, for some better qualified system ... with a Ph.D.? If the theory is to apply to anything but highly idealised laboratory operations, are we not obliged to admit that more or less 'measurement-like' processes are going on more or less all the time, more or less everywhere? Do we not have [quantum] jumping then all the time?

[8] J.S. Bell, Against 'Measurement', reprinted in Speakable and Unspeakable in Quantum Mechanics, 2nd edn. (Cambridge University Press, Cambridge, 2004)

3.2.1 Decoherence

Following the idea put out in the above sentence by Bell, in the late 1970's H. D. Zeh and W. Zurek included in the dynamical description of quantum systems the unavoidable interaction with degrees of freedom that we don't have access to.[9]

[9] For a readable introduction to the theory of decohere read: W. Zurek, "Decoherence and the Transition from Quantum to Classical", Physics Today **44** 36 (1991). See also the revised version published with the same title in 1993.

In the cat experiment we can imagine that the air molecules inside the box get heated depending whether the cat is dead or alive. The information about the internal temperature leaks out of the box, an thus even without opening the box we can gather information about the cat's fate and the superposition vanishes. The environment is always measuring the system, even if we are not looking at it.

In our description of a quantum measurement process above, this interaction with other degrees of freedom can be taken into account by including a third space. In the above explanation, we have the space for the spin 1/2 particle and the angular momentum measuring apparatus $\mathscr{H}_S \otimes \mathscr{H}_A$. As the measuring apparatus is meant to be macroscopic, it is reasonable to expect that we cannot completely isolate it. We then include a space for the environment, $\mathscr{H}_E$.

Suppose now that the environment starts in the state $|E_0\rangle$, and that the system-apparatus interaction is much faster than their interaction with the environment. After some time, the total quantum description is given by

$$(c_0|0\rangle \otimes |\theta\rangle + c_1|1\rangle \otimes |-\theta\rangle) \otimes |E_0\rangle.$$

This stage is sometimes called as pre-measurement.

Now we take into account the interaction of system-apparatus and the environment:

$$\begin{aligned} |\psi_{SAE}(\text{pre})\rangle &= (c_0|0\rangle \otimes |\theta\rangle + c_1|1\rangle \otimes |-\theta\rangle) \otimes |E_0\rangle \\ &\downarrow \\ |\psi_{SAE}(\text{final})\rangle &= c_0|0\rangle \otimes |\theta\rangle \otimes |E_+\rangle + c_1|1\rangle \otimes |-\theta\rangle \otimes |E_-\rangle, \end{aligned}$$

where $|E_+\rangle, |E_-\rangle$ are states in $\mathscr{H}_E$, such that $\langle E_+|E_-\rangle = 0$.

As by hypothesis we don't have access to the environmental degrees of freedom, we trace them out to get the reduced state in the system+apparatus part:

$$\begin{aligned} \rho_{SA}(\text{final}) &= \text{Tr}_E |\psi_{SAE}(\text{final})\rangle\langle\psi_{SAE}(\text{final})|, \\ &= |c_0|^2 |0\rangle\langle 0| \otimes |\theta\rangle\langle\theta| + |c_1|^2 |1\rangle\langle 1| \otimes |-\theta\rangle\langle -\theta|. \end{aligned}$$

In this final local description, we still have the spin part correlated with the apparatus value: we get θ with probability $|c_0|^2$, and $-\theta$ with probability $|c_1|^2$. Moreover, instead of a quantum superposition, now we have a convex mixture of the possibilities. This is the same one has in classical mechanics.

This emergence of classical aspects due to decoherence was experimentally observed by Haroche's group. As we mentioned above, their "apparatus"

was the electromagnetic field trapped between two mirrors, i.e, a cavity. This cavity of course is not perfect, and as the field escaped from it, the quantum interference decreases and eventually one gets to a classical mixture. See Fig.3.3.

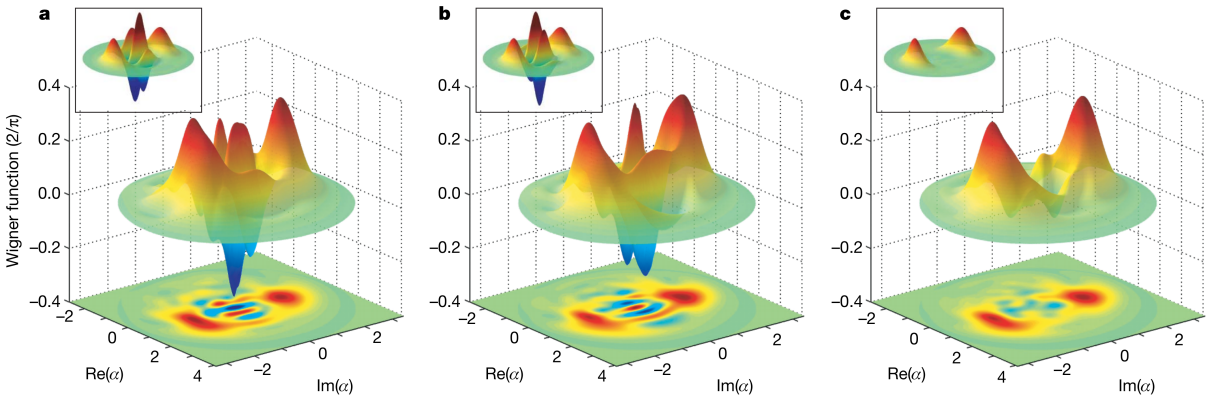

Figure 3.3: Evolution of a cat-like state $\propto (|\alpha\rangle + |-\alpha\rangle)$ due to a decoherence process. Figure taken from S. Deléglise et al. "Reconstruction of non-classical cavity field states with snapshots of their decoherence". Nature **455** 510 (2008).

Does that solve the quantum measurement problem? No one denies that the decoherence process is important in the description of the quantum-to-classical transition. Nevertheless, in a sense, it makes the problem even bigger. The state $|\psi_{SAE}(\text{final})\rangle$ is now an even bigger superposition, in the sense that it takes now also the environment.[10]

Is it even possible to solve the big measurement problem within quantum mechanics? Some new approaches are emerging in the recent year.[11] However, it seems that without a better understanding of what a measurement means, we are always running in circles.

[10] To a discussion about the role of decoherence in the quantum measurement process see for instance: Guido Bacciagaluppi, "The Role of Decoherence in Quantum Mechanics", The Stanford Encyclopedia of Philosophy (Fall 2016 Edition), Edward N. Zalta (ed.).

[11] Časlav Brukner,"On the quantum measurement problem". Proceedings of the Conference "Quantum UnSpeakables II: 50 Years of Bell's Theorem" (Vienna, 19-22 June 2014).

C Duarte, GD Carvalho, NK Bernardes, F de Melo, "Emerging dynamics arising from coarse-grained quantum systems". Phys. Rev. A **96**, 032113 (2017).

4 *Quantum non-Locality and/or Realism*

We start this chapter with a remark: quantum mechanics (non-relativistic) is not only about the Schroedinger equation! Even though most of quantum mechanics courses and books focus on the solution of Schoedinger's equation in a variety of situations (wall potential, harmonic oscillator, hidrogen atom, scattering theory, etc), this often obscures the fact that the unitarity of a quantum mechanical evolution is only one of the postulates of quantum theory. Your condensed matter or particle physics colleagues might get outraged, but trust us, it is not necessary to introduce a Hamiltonian in order to be able to talk deeply about quantum mechanics. The aim of this chapter is to show you why this is the case. We are going first to focus on the Born's rule, the postulate connecting the abstract objects of the theory (state and measurement operators) with what we actually see in our laboratories: cliques in a detector. This will lead us to Bell's theorem, showing that quantum mechanical predictions are incompatible with a very natural and intuitive set of assumptions about the world surrounding us. After discussing some of the practical issues in Bell's theorem and its application in cryptographic protocols we will turn our attention to yet another of the quantum postulates, that one stating that physical systems are described by state vectors (more generally a density operator). What the wave function stands for? Is it just a abstract mathematical tool or it probes something deeper about the physical reality? If after reading this chapter you are not deeply shocked, read it again. Otherwise, adapting Feynman's famous quote, we "can safely say that you have not yet understood quantum mechanics".

4.1 The EPR "Paradox"

In the thirties, quantum mechanics was already a well developed and well tested theory. The successful explanation of a wide range of different phenomena (black-body radiation, photoeletric effect, the hidrogen atom, absortion and emission lines of various elements, etc) left no room for anyone doubting its predictions. Until today, there is no experimental evidence for the need of a post-quantum theory (combining gravity and quantum mechan-

ics into a single coherent picture is a different matter...). However, and in spite of its clear successes, quantum theory was conceptually very different from what physicists have become accustomed. First of all, quantum theory is a probabilistic theory. It does predict the probabilities of what will be the measurement outcomes of a given experiment if that experiment is performed over and over again. However, it can't say what will be the outcome in a specific run of the experiment. This new view of the world, intrinsically probabilistic, was in clear clash with the old way of doing physics. In the Newtonian mechanics, everything boils down to determine the position and momentum of a given physical system. These properties are well defined independently of our act of observing them. The measurement is simply a way to determining a pre-defined property of the physical system.

In the quantum description, however, physical properties and measurements are inseparable. There is nothing in the quantum formalism that allow us to speak about the momentum of a particle if actually the measurement being performed is the one about its position. This matter is famously illustrated via the Heisenberg uncertainty relation

$$\sigma_A \sigma_B \geq \frac{1}{2} |\langle [A,B] \rangle|, \tag{4.1}$$

basically stating that the product of standard deviations (the degree of uncertainty we have about a random variable) of two observables, A and B, is lower bounded by the commutation relation between these observables. For instance, position and momentum, or the spin along two orthogonal directions, do not commute, implying that no quantum state will have both properties well defined (or sharp) at the same time. With "well defined" we mean that we can predict with certainty the result of a measurement of that given observable. Take for instance, a qubit state that has a well defined spin along the z direction: $|\Psi\rangle = |0\rangle$ (we are using the convention that $|0\rangle$ and $|1\rangle$, are the eigenvectors of operator describing the spin along the z direction $S_z = (\hbar/2)\sigma_z$, with $\sigma_z = (|0\rangle\langle 0| - |1\rangle\langle 1|)$ being one of the Pauli matrices). If we measure the spin along the z direction we are certain to obtain spin-up. However, if we just turn our magnets around (thinking here about the Stern-Gerlach setup) and measure the spin direction along the x direction, we will have probability half of obtaining the spin up or down along this direction (maximal uncertainty).

This state of affairs was deeply unsatisfactory to many physicists, including Einstein that, together with Podolsky and Rosen, in 1935 have attempted a fatal blow against quantum theory. In what is now known as the EPR paper [1], they argue that even though quantum theory is correct, it would be an incomplete theory. By describing quantum theory they arrive at the logical conclusion that either (1) the guantum-mechanical description of reality given by the wave function is not complete or (2) when the operators corresponding to two physical quantities do not commute the two quantities cannot have simultaneous reality. To that aim, they first had to introduce a mathematical

[1] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical review*, 47(10):777, 1935

definition of what would be an element of reality:

If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity.

Clearly, any theory to be considered complete, should have in its description all elements of reality. And here enters the core of the EPR argument: by assuming quantum mechanics to be complete and using entangled quantum systems and some apparently innocuous and natural assumptions, they argue that one can associated elements of reality to non-commuting observables. That is, by negating the statement (1) leads to the negation of the only other alternative (2). We are thus forced to conclude that the quantum-mechanical description of physical reality given by wave functions is not complete. If that would really be the case, quantum theory could then be seen as an effective theory of a more fundamental and yet to be discovered theory. The EPR argument is very elegant and correct. Its flaws lie in the apparently innocuous and natural assumptions, something we will see when talking about Bell's theorem. First, however, let's see how the EPR argument goes. We will employ here the neater version of the argument as outlined by David Bohm using spin $1/2$ particles.

Consider a entangled pair of spins, described by the state vector (written in the $S_z = (\hbar/2)(|0\rangle\langle 0| - |1\rangle\langle 1|)$ basis)

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right). \tag{4.2}$$

Clearly, the spins along the z-direction of both particles are correlated: if we measure one of the spins along the z direction and find spin up (down) we know with certainty that the other spin will also be up (down).

But let's now consider that we make a measurement of spin along the x direction. Rewriting the state vector in the in the $S_x = (\hbar/2)(|+\rangle\langle +| - |-\rangle\langle -|)$ basis we have

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}\left(|++\rangle + |--\rangle\right), \tag{4.3}$$

where $|\pm\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle \pm |1\rangle\right)$. Clearly, the spins along the x-direction of both particles are also correlated: if we measure one of the spins along the x direction and find spin up (down) we know with certainty that the other spin will also be up (down).

Now, let's suppose we take these two spins and bring them very far apart, say to opposite sides of the universe. Suppose now, that we measure the first spin (either along the z or the x direction) and obtain a given measurement outcome (either up or down). The quantum mechanical description ensures then that if the same measurement is performed on the second spin (light years away from the first), the same measurement outcome will happen. Furthermore, given their space-like separation, we can safely say that the

measurement result of the measurement on the second spin cannot in any way depend on which of the measurements have been performed in the first, it must be an intrinsic and pre-existing property of this second particle. Apparently then, we have achieved the two conditions to call these observables (spin along orthogonal directions) elements of reality: we can predict with certainty and without disturbing it the spin of the second particle along two orthogonal directions. Quantum theory, however, implies that these two observables do not commute, and thus cannot be assigned well defined values at the same time. Hence, quantum theory is incomplete. Is that so?

4.2 Bell's theorem

If we accept the assumption underlying the EPR argument, there is no way around: quantum mechanics is indeed an incomplete theory. Instead of that, however, we can challenge the EPR assumptions. There are 2 explicit assumptions (locality and realism) and a third implicit one (free-will or measurement independence, to be introduced later on). The locality assumption basically states that only events that might affect a measurement outcome are those in its causal past. In turn, the realism assumption implies that the properties of a physical system are well defined and pre-existing regardless of whether a measurement is made to reveal its value.

Coming back to the EPR argument. Following the notation usual in quantum information, we can say that one of the spin 1/2 particles is sent to Alice and the other to Bob, their labs being very far apart. Every time they receive their share of the physical system, they randomly select one out of two possible measurements to perform. Furthermore, we assume that only two possible outcomes are possible (up or down, or equivalently eigenvalues $+1$ and -1). Quite generally, we can describe this experiment via a conditional probability distribution $p(a,b|x,y)$, a being the measurement outcome of Alice given that she measured an observable labelled by x (similarly to Bob). Let's now see what these two assumptions (usually refereed as local realism), imply to the EPR Gedankenexperiment. It follows that

$$\begin{aligned} p(a,b|x,y) = \quad & \sum_{\lambda} p(a,b,\lambda|x,y) \\ & \sum_{\lambda} p(a,b|\lambda,x,y)p(\lambda|x,y) \\ & \sum_{\lambda} p(a|\lambda,x)p(b|\lambda,y)p(\lambda|x,y) \\ & \sum_{\lambda} p(a|\lambda,x)p(b|\lambda,y)p(\lambda), \end{aligned} \tag{4.4}$$

the so called local hidden variable (LHV) model. The realism assumption is already present when we introduce an auxiliary variable λ (also known as a hidden variable) that supposedly describes the properties of the physical system and governs the probability distribution of the measurement outcomes

given some measured observables. At the first line above, we have simply used the realism assumption together with the law to total probability. In the second line we have simply used Bayes law. At the third one we have used the locality assumption (observe that each of the measurement outcomes is fully specified only by the variables in their causal past). Finally, at the fourth line we have used the third (implicit assumption in the EPR argument), the free-will assumption (or measurement independence assumption), stating that the measurement choices of Alice and Bob (x and y) are independent of how the system (described by λ) has been prepared. Mathematically, $p(x,y,\lambda) = p(x)p(y)p(\lambda)$.

As natural as these 3 assumptions might appear, the quantum mechanical predictions are incompatible with them. This is precisely Bell's theorem [2]. To prove the theorem it is enough to find a quantum mechanical distribution incompatible with the LHV distribution (4.4). We have thus, two new ingredients to consider here: i) how to describe the quantum probability distributions arising in this scenario and ii) how to test whether a given distribution $p(a,b|x,y)$ can or not be decomposed as in the LHV description of (4.4).

[2] John S Bell. On the einstein podolsky rosen paradox. *Physics Physique Fizika*, 1(3):195, 1964

Born's rule, one of the postulates of quantum theory, implies that the quantum mechanical description of the distribution $p(a,b|x,y)$ is given by

$$p(a,b|x,y) = \mathrm{Tr}\left[\left(M_a^x \otimes M_b^y\right)\rho\right], \tag{4.5}$$

where M_a^x and M_b^y are the measurement operators of Alice and Bob and ρ is density operator describing their shared state.

In turn, the LHV description given by (4.4) forms a convex set, more precisely a polytope (a paradigmatic example would be a cube). We will not enter in the geometrical details here but is enough to say that this convex set can be either described in terms of its extremal points (for instance, the 8 corners of a 3D cube) or in terms of its boundary hyperplanes (for instance, the 6 facets of a 3D cube). The facets of the convex set defined in (4.4) are the famous Bell inequalities. Any probability distribution admitting a LHV decomposition should fulfil these inequalities. Otherwise, if a probability distribution violates any of these Bell inequalities, this proves unambiguously that this distribution is incompatible with a LHV model. In the simplest Bell scenario, corresponding to the case where Alice and Bob measure two possible observables (each with two possible outcomes) it is easy to prove that the so called CHSH inequality [3] holds

[3] John F Clauser, Michael A Horne, Abner Shimony, and Richard A Holt. Proposed experiment to test local hidden-variable theories. *Physical review letters*, 23(15):880, 1969

$$\langle \mathrm{CHSH} \rangle = \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle \leq 2 \tag{4.6}$$

where $\langle A_x B_y \rangle = \sum a,b (-1)^{a+b} p(a,b|x,y)$ is the expectation value of the joint measurement outcomes of Alice and Bob. Without loss of generality, let's assume that the measurement outcomes (spin up or down) are labelled by the eigenvalues ± 1 of the corresponding Pauli operators. We can write the CHSH operator as

$$\mathrm{CHSH} = A_0(B_0 + B_1) + A_1(B_0 - B_1). \tag{4.7}$$

Assuming the LHV condition, it is easy to see that B_0 and B_1 should either be equal or different independently of whether their are measured together with A_0 or A_1. If B_0 and B_1 are equal, then CHSH $= \pm 2A_0$. If B_0 and B_1 are different, then CHSH $= \pm 2A_1$. In either case, since $A_0 = \pm 1$ and $A_1 = \pm 1$, its follows that CHSH $= \pm 2$. On average, the maximum possible value of CHSH is 2, then proving the CHSH inequality $\langle \text{CHSH} \rangle \leq 2$.

The final piece to prove Bell's theorem is to find measurement operators for Alice and Bob and a shared joint state achieving a value for $\langle \text{CHSH} \rangle$ beyond the classical bound of 2. If we choose $A_0 = \sigma_z$, $A_1 = \sigma_x$, $B_0 = \frac{\sigma_z + \sigma_x}{\sqrt{2}}$ and $B_1 = \frac{\sigma_z - \sigma_x}{\sqrt{2}}$, a straightforward calculation using Born's rule shows that $\langle A_0 B_0 \rangle = \langle A_0 B_1 \rangle = \langle A_1 B_0 \rangle = -\langle A_1 B_1 \rangle = 1/\sqrt{2}$ thus implying that the quantum prediction gives $\langle \text{CHSH} \rangle = 2\sqrt{2}$, violating the CHSH inequality and proving the incompatibility of quantum mechanics with local realism (alternatively with the three assumptions discussed above).

The violation of a Bell inequality, leads to the phenomenon known as quantum non-locality (or Bell non-locality). As it will be explored in more details later on, this does not imply that quantum mechanics can be used to communicate faster than light. What is implies is that if we want to keep a description where we have free-will and the where the properties of a physical system are well defined independently of any measurement being performed, then necessarily we have to give the locality assumption (that is, even if far apart two systems would behave as parts of unit indivisible system).

4.3 Experimental loopholes

The beauty of Bell's theorem is that is makes possible to test experimentally a very fundamental aspect of our physical reality. An experimental violation of a Bell inequality is a definite proof that whatever theory describes Nature, it needs to be non-local (be it quantum or not). All tests to date have irrefutably proved the non-local aspect of nature [4] (and so far the correctness of quantum mechanics, more on that on the next chapter). However, when dealing with experimental implementations of a Bell scenario, we have to take utmost care to fulfil some crucial requirements. Otherwise, our experiment is open to loopholes, ways to simulate or mimic a non-local behaviour with a local hidden variable model. Over the years, several loopholes have been identified and discussed. The three crucial ones, to be discussed here are: i) free-will loophole, ii) locality loophole and iii) the detection efficiency loophole.

4.3.1 Free-will Loophole

In the derivation of (4.4) we have explicitly used that $p(x,y,\lambda) = p(x)p(y)p(\lambda)$, basically stating that the choices of which measurements to perform can be assumed independent of the variable governing the state of the physical sys-

[4] Bas Hensen, Hannes Bernien, Anaïs E Dréau, Andreas Reiserer, Norbert Kalb, Machiel S Blok, Just Ruitenberg, Raymond FL Vermeulen, Raymond N Schouten, Carlos Abellán, et al. Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres. *Nature*, 526 (7575):682, 2015; Marissa Giustina, Marijn AM Versteegh, Soeren Wengerowsky, Johannes Handsteiner, Armin Hochrainer, Kevin Phelan, Fabian Steinlechner, Johannes Kofler, Carlos Larsson, et al. Significant-loophole-free test of bellâĂŹs theorem with entangled photons. *Physical review letters*, 115(25):250401, 2015; and Lynden K Shalm, Evan Meyer-Scott, Bradley G Christensen, Peter Bierhorst, Michael A Wayne, Martin J Stevens, Thomas Gerrits, Scott Glancy, Deny R Hamel, Michael S Allman, et al. Strong loophole-free test of local realism. *Physical review letters*, 115(25):250402, 2015

tem (the one to be measured). This is the so called free-will assumption (also called measurement independence) [5]. If this condition does not hold, then hidden variable models can easily be shown to mimic any quantum and even post-quantum correlations. But how can we guarantee free-will? Well, we can't, at least not invoking any causality arguments. All events and process involved in the generation of the measurement choices and the preparation of the physical system do have a common cause (the Big Bang if you wish). But please notice that any experiment (in physics, biology or any other science) is subjected to the same constraints, unless we can assume free-will to hold true, any scientific endeavour is meaningless.

This does not mean, however, that we should not try to make the condition $p(x,y,\lambda) = p(x)p(y)p(\lambda)$ as plausible as possible. There have been many attempts in this direction. For instance, using cosmic photons as the inputs deciding which observable to measure [6]. This way, any correlations between our measurement choices and the physical systems being prepared during the Bell experiment have to have been generated hundred of years ago, thus before the apparatus preparing the physical had even been produced. Another proposal has been the use of human randomness as the inputs. More precisely, in a recent experimental demonstration [7], hundreds of thousands of people around the globe have decided (via an online videogame) which measurements to be performed (in realtime) in several Bell tests runned in labs in four continents. Such ideas certainly seem very conspiratory, but that is what it takes to try to explain quantum correlations with our everyday intuition!

[5] Michael JW Hall. Local deterministic model of singlet state correlations based on relaxing measurement independence. *Physical review letters*, 105(25):250404, 2010

[6] Johannes Handsteiner, Andrew S Friedman, Dominik Rauch, Jason Gallicchio, Bo Liu, Hannes Hosp, Johannes Kofler, David Bricher, Matthias Fink, Calvin Leung, et al. Cosmic bell test: measurement settings from milky way stars. *Physical review letters*, 118 (6):060401, 2017

[7] BIG Bell Test Collaboration et al. Challenging local realism with human choices. *Nature*, 557(7704):212, 2018

4.3.2 Locality Loophole

Another explicit assumption in our derivation of the LHV model (4.4), is the locality assumption stating that $p(a,b|\lambda,x,y) = p(a|\lambda,x)p(b|\lambda,x)$. That is, the measurement outcome in a given lab (Alice or Bob) can only depend on the local measurement choice (variables x and y) and the state of the physical system (represented by λ). For that, we are invoking local causality: since the sub-systems being measured are far apart (more precisely, the measurement outcomes a and b are space-like separated events), no action taken in one lab should affect in any way the statistics observe in the other. However, to invoke special relativity, our experimental implementation should indeed use far away labs (remember, light is really fast). And here comes the problem. We can easily, generate entangled systems these days (ions traps, superconducting devices, photonic experiments) but bringing these systems apart is tough. For instance, an ion trap (the device where the entangled system is located) is typically on the scale of micrometers, order of magnitudes smaller than the required distances. Clearly, to cover the large distances required to impose the locality condition, photons are the obvious candidates [8]. As a matter of fact, entangled photon pairs can nowadays be

[8] Gregor Weihs, Thomas Jennewein, Christoph Simon, Harald Weinfurter, and Anton Zeilinger. Violation of bell's inequality under strict einstein locality conditions. *Physical Review Letters*, 81(23):5039, 1998

distributed at distances surpassing thousand kilometers. Photons, however, run into yet another problem to be circumvented in a Bell experiment, the so called detection efficiency loophole.

4.3.3 Detection efficiency Loophole

Suppose you have setup your Bell experiment with photons considering two enough far away labs. You then start running your experiment and after sufficiently many data is collected, compute the expectation values entering in the CHSH inequality and find a larger than 2. Understanding the meaning of Bell's theorem you are indeed astonished, you have just proved that nature is non-local. At lunch, you are happily telling your colleagues your discovery when a somber fellow theorist asks you in grave tone: what are the detection efficiencies of your photon detectors? Unaware of what is to come, you say, well, around 10% of the photons being generated are actually detected. That is the typical number of the most common kind of detectors, avalanche photo-detectors, that upon the arrival of a photon generate an strong enough signal to be read. Most of the photons hitting the detector, however, do not generate such current, that is, they are not detected.

The problem with that, is that you are in fact post-selecting the data: from all the events (the entangled pairs being generated) only a small fraction of them is being considered when you compute your CHSH value. Let's call $p_D(a,b|x,y)$, the probability distribution you obtain by considering the detected events and $p_U(a,b|x,y)$ the undetected ones. The actual probability distribution of your experiment is $p(a,b|x,y) = \eta\, p_D(a,b|x,y) + (1-\eta)p_D(a,b|x,y)$, however, you do not have empirical access to $p(a,b|x,y)$ but rather to $p_D(a,b|x,y)$. To understand, why this is problematic consider the best scenario where p_D gives rise to a CHSH value of $2\sqrt{2}$. On the opposite direction, suppose a worst case scenario where all undetected events lead to a p_U such that the CHSH value is 0 (that is, p_U is local). The actual violation of the CHSH inequality you should be computing is that given by p and that in this case is then given $\eta 2\sqrt{2}$ that is larger than the classical bound of 2 only if $\eta \geq 1/\sqrt{2} \approx 0.7071$. Since your actual efficiency is way below the 70%, your experiment is not conclusive. In order to say that you have indeed violated a Bell inequality, you should make a further assumption, called fair-sampling assumption [9]. It basically states that the detected and undetected events are of the same type and not something like the example above (where p_D is non-local and p_U is local).

[9] Janke Larsson. Loopholes in bell inequality tests of local realism. *Journal of Physics A: Mathematical and Theoretical*, 47(42): 424003, 2014

When it comes to Bell's theorem, however, we are paranoids. Remember, even our free-will has been questioned in order to salvage Nature against non-locality. To really be sure, we should be using good enough detectors. Detector achieving almost 100% efficiency are available for the measurement of ions (remember, however, that in this case we cannot ensure the locality condition). We are really in a tight spot: using photons we can achieve the

locality condition but detection efficiencies are low; using ions we have no issues with detection efficiencies but no way we can enforce locality. Because, of that it was only in 2015 that the first loophole free Bell experiments have been performed. In one of these experiments, we had two distant diamonds and inside them electrons that emit a photon in such a way that their spin get entangled with such a photon. These photons are then directed to a central lab (in between Alice and Bob) and then measured in an entangled basis, that via an entanglement swapping protocol guarantees that the electronic spins of the two distant diamonds (that have never interacted before) are now entangled. The detection efficiency of the electronic spin is very high and thus we close both the detection and locality loopholes. The other loophole free experiments, used entangled light and superconducting photon detectors achieving detection efficiencies beyond 90% and thus also closing the loopholes. In all these experiments, Bell inequalities have been violated, conclusively showing (unless you are ready to give up free-will) that Nature is non-local.

4.4 Quantum cryptography

In 1994, the field of quantum computation emerged from the anonymity [10]. Peter Shor, working at the Bell Labs, put out his now famous algorithm, showing how quantum computers could perform integer factorization in polynomial time. This was big news for two main reasons. First, the best known classical algorithm (known as number field sieve) has an exponential complexity. Second, exactly because of this exponential complexity, integer factorization lies at the core of many cryptographic protocols, including the widely used RSA (the security of which relies on the difficulty to factorize a large number into its prime factors). However, if one could implement Shor's algorithm, this kind of crytography was doomed. For instance, the factorization of a 2048 digit number (typically used in RSA) – that on a classical computer could take longer than the age of the universe– on a quantum computer could be performed on a matter of minutes. This was the first example of how quantum computers could offer an exponential advantage on a practical and very timely information problem.

[10] Peter W Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *Proceedings 35th annual symposium on foundations of computer science*, pages 124–134. Ieee, 1994

In short, Shor has shown that quantum mechanics put in danger our communication security. Ten years ago, however, Bennett and Brassard showed that at the same time quantum theory also offer novel forms of cryptography [11].

[11] Charles H Bennett and Gilles Brassard. Quantum cryptography: public key distribution and coin tossing. *Theor. Comput. Sci.*, 560(12):7–11, 2014

4.4.1 The BB84 protocol

Let's say that Alice wants to communicate with Bob using a qubit. A simple way of doing so would be to prepare the eigenstates of a given observable, say $\sigma_z = |0\rangle\langle 0| - |1\rangle\langle 1|$, the state $|0\rangle$ if she wants to send the bit 0 and the state $|1\rangle$ if she wants to send the bit 1. Bob can recover this information, by

measuring the qubit sent by Alice by measuring the same observable σ_z. Clearly, if the state received is $|0\rangle$ we have certainty that the measurement outcome will be $+1$; if the state is $|1\rangle$ we have certainty that the measurement outcome will be -1. The same will be true for any other observable. For instance, Alice could encode her 0s and 1s in the eigenstates of $\sigma_x = |+\rangle\langle+| - |-\rangle\langle-|$, preparing $|+\rangle$ if she wants to send bit 0 and $|-\rangle$ if she wants to send bit 1.

Now, let's say that Alice decides randomly which basis she will use to encode her message (with probability $1/2$ she uses the $|0,1\rangle$ basis or the $|+,-\rangle$ basis). Bob does not know which basis she has used, so half the time he will make a wrong choice. If Alice encodes the bit 0 in the state $|0\rangle$ but Bob measured σ_x, with probability $1/2$ he will obtain the outcome -1 and thus be led to the conclusion that the bit send was 1. To avoid these errors, Alice and Bob have to announce publicly which basis they have used at each round. Introducing this element of randomness seems stupid, as half of the communication is just being thrown away. However, as we will see next, it guarantees that no eavesdropper can have access to their communication without being detected.

Suppose that Eve, the eavesdropper, is intercepting the qubits being sent by Alice. The aim of Eve is to measure the qubits, extract information and resend these qubits to Bob as if nothing had happened. As Bob, Eve does not know which basis Alice is using to encode her bits. As Bob, Eve has to randomly choose a basis. If the basis is correct, Eve will have access to the information. If, however, the basis is different from the one chosen by Alice, Eve not only will not have access to the information but as well will destruct it. For example, if Alice encode the bit 0 in the state $|0\rangle$ but Eve measured σ_x, with probability $1/2$ Eve will obtain the outcome -1, projecting the state to be sent to Bob to $|-\rangle$. Let's say that at that run, Bob choose to measure in the σ_z basis. If the qubit had had not been intercepted by Eve (in that case the state would be $|0\rangle$), he would obtain the outcome $+1$ with certainty, thus recovering the correct information. However, as the state is now in the state $|-\rangle$, he has probability $1/2$ of obtaining the wrong result -1.

The cryptography protocols proceed as follows. Alice and Bob announce their choice of basis, they take a fraction of those ones where their choice was the same: Alice also announces the bit she wanted to encode and Bob announces his measurement outcome. If no Eve was in between, they should have announced the same bit. However, if Eve was in between, half the times she will make the wrong answer of the basis and in these cases, also half the time, Bob will obtain as a measurement outcome a bit flip of the message sent by Alice. Overall, if Eve is in between, with probability $1/4$ Alice and Bob will announce a different bit and thus led to the conclusion that Eve is listening to them.

The only way Eve could pass undetected would be for her to find a way to distinguish between the basis $|0,1\rangle$ and $|+,-\rangle$. As we will see later on,

that turns out to be impossible, as it violates one of the consequences of quantum theory, the so-called, no-cloning theorem. We are thus led to the apparent conclusion, that unless Eve can break the laws of quantum theory, we have just achieved full cryptographic security. In 2010, however, a paper appeared in the arXiv [12], claiming to have cracked the uncrackable. Makarov and collaborators, have hacked a commercial crypto system based on the BB84 protocol. Have they just proved quantum mechanics wrong? Of course not. What they brought to our attention was the fact that the BB84 protocol assumes its perfect realization. That is, that Alice is preparing and Bob is measuring on the Z,X basis. Any deviations from that open to way to attacks and ways to crack the security of the protocol.

[12] Lars Lydersen, Carlos Wiechers, Christoffer Wittmann, Dominique Elser, Johannes Skaar, and Vadim Makarov. Hacking commercial quantum cryptography systems by tailored bright illumination. *Nature photonics*, 4(10):686, 2010

What was needed, was a new type of information protocol, where simply based on the statistics we observe in a experiment (without any assumptions about the internal working of the preparing and measuring devices) we can achieve success in a given task. The answer to that is in the violation of a Bell inequality.

4.4.2 Quantum criptography 2.0

Assuming that free-will holds, the violation of a Bell inequality shows us that we have no choice: if we assume locality, the only way to explain the violation of a Bell inequality is give up the reality of physical properties. That is, our measurement is giving us outcome $+1$ (spin pointing up, for instance), however, this does not mean that the system had spin-up previously to the measurement taking place. A bit that is not really defined until we decide to measure it, seems like a nice way to secure information. Let's do this argument a bit more precise [13].

[13] Artur K Ekert. Quantum cryptography based on bellâĂŹs theorem. *Physical review letters*, 67(6):661, 1991

As we will see in the next chapter, the maximum violation of the CHSH inequality is given by $\text{CHSH} = 2\sqrt{2}$ and the only way to achieve this value is if we measure on a maximally entangled two-qubit state $1/\sqrt{2}(|00\rangle + |11\rangle)$ (or locally unitarily equivalent to it). This is a particular case of what is known as self-testing [14]: by the measurement statistics alone (without assuming anything about which measurement are being performed) we can infer something about the physical system. But maximally entangled states respect a strong constraint known as entanglement monogamy [15]. The more a given system A is entangled with B, the less it can be entangled with any other system C. In particular, if A and B are maximally entangled, then they should be completely uncorrelated from the rest of the universe. That is, if we have maximum violation of the CHSH inequality, this implies that the physical system shared by Alice and Bob is completely uncorrelated with any other device (in particular those owned by Eve). At the same time, we know that if Alice and Bob measure the same observable they will get the same outcome, that is, they generated a string of correlated bits that are known only to them. That is, they achieve a secret bit string that can be used to

[14] Dominic Mayers and Andrew Yao. Self testing quantum apparatus. *arXiv preprint quant-ph/0307205*, 2003

[15] Valerie Coffman, Joydip Kundu, and William K Wootters. Distributed entanglement. *Physical Review A*, 61(5):052306, 2000

cryptographic purposes.

Quite generally, we can model the Alice, Bob and Eve state of affairs via a tripartite state ρ_{ABE}. Similarly to Alice and Bob, Eve is also measuring her shared of this state, with the aim of obtaining an outcome e equal to the outcome a of Alice. The tripartite distribution is given by

$$p(a,b,e|x,y) = \mathrm{Tr}\left[\left(M_a^x \otimes M_b^y \otimes M_e\right)\rho_{ABE}\right] \tag{4.8}$$

such that $p(a,b|x,y) = \sum_e p(a,b,e|x,y)$. Since $p(a,b|x,y)$ maximally violates the CHSH inequality $\rho_{AB} = |\Phi\rangle\langle\Phi|$ with $|\Phi\rangle = 1/\sqrt{2}(|00\rangle + |11\rangle)$ and the entanglement monogamy then implies that $\rho_{ABE} = \rho_{AB} \otimes \rho_E$. That is, the outcome e is completely independent of the outcomes a and b, in other terms, $p(a,b,e) = p(a,b)p(e)$. The objective of Eve is to maximize $p(e=a)$, that in this case is simply $p(e=a) = 1/2$, that is, a random guess.

4.5 PBR theorem

Let's know focus our attention to another postulate of quantum mechanics, that one saying that a physical system is represented by a state vector (more generally a density operator) in a complex vector space. Via Born's rule we know how to connect what we observe in the lab with this abstract description. But the question remains: what does the wave-function stand for? Is it just a mathematical tool used to calculate probabilities? In that case, the quantum state represents only knowledge or information about some aspect of reality. Or can it be that a pure quantum state corresponds directly to reality? As we will see in what follows, under some assumptions, the Pusey-Barrett-Rudolph (PBR) theorem [16], proves that indeed the wave-function cannot be interpreted statistically.

[16] Matthew F Pusey, Jonathan Barrett, and Terry Rudolph. On the reality of the quantum state. *Nature Physics*, 8(6):475, 2012

The big question is how should one understand the wave-function. There are basically two broad views. In the Psi-epistemic view, the wave vector indeed only carries knowledge. Within this camp we can further refine two possibilities:

1) $|\Psi\rangle$ is epistemic but there is an underlying ontic state. That is, we can understand quantum mechanics as the statistical theory of these ontic states (in analogy with statistical mechanics and Newtonian physics).

2) $|\Psi\rangle$ is epistemic but there is no underlying reality. That is, the wave function represents a state of knowledge but is complete simply because there is nothing else there to be known. This is the view broadly represented by the Copenhagen interpretation.

The second broad view is one where

3) $|\Psi\rangle$ is ontic. It represents a fundamental piece of the underlying reality but there might be some additional ontic degrees of freedom that are not described by quantum theory. Because of Bell's theorem, we know that such extended theories should be non-local. The most famous interpretation in this camp is the de Broglie-Bohmian interpretation.

The PBR theorem excluded option **1)** as a possibility.

Before we turn to the proof of the theorem let's first connect these epistemic/ontic with classical physics. In classical physics, we typically talk about ontic state (states of reality). For instance, all we need to know is the position and momentum at a given instant of time t $(x(t_0), p(t_0))$; knowing the relevant forces (or the Hamiltonian) we can predict deterministically the state $(x(t), p(t))$ at any other later time. In turn, in statistical physics we often talk about epistemic states (states of knowledge): we replace $(x(t), p(t))$ by a probability distribution $\mu(x, p)$ that represents our epistemic state. That is, this probability distribution represents an experimenter's uncertainty about the physical state of the particle but it does not represent reality directly.

In general, we can also speak about other physical properties, functions of the physical state, for instance, the energy $E(x, p)$. If all we know is the energy E of the system, there will be a number of possible physical state (x, p) compatible with it, each described by a given probability distribution $\mu_E(x, p)$. If instead we have energy E' we will have another distribution $\mu_{E'}(x, p)$. However, if energy is indeed a physical property (uniquely defined by the physical state) then it follows that $\mu_E(x, p)$ and $\mu_{E'}(x, p)$ should have disjoint supports. If, on the contrary, both distributions would overlap, this means that systems with the same (x, p) would be compatible with different values of energy and thus energy would not be correspond to a physical property.

Now consider a quantum system. The hypothesis is that the quantum state is a state of knowledge, representing uncertainty about the real physical state of the system. Hence assume some theory or model, perhaps undiscovered, which associates a physical state λ to the system. If a measurement is performed, the probabilities for different outcomes are determined by λ. If a quantum system is prepared in a particular way, then quantum theory associates a quantum state (assume for simplicity that it is a pure state) $|\Psi\rangle$. But the physical state λ need not be fixed uniquely by the preparation. Rather, the preparation results in a physical state λ according to some probability distribution $\mu_\Psi(\lambda)$. Clearly, the variable λ is the analogous of (x, p) and $|\Psi\rangle$ the analagous of a physical property (energy, for example).

If, for any pair of distinct quantum states $|\Psi_0\rangle$ and $|\Psi_1\rangle$, the corresponding distributions $\mu_0(\lambda)$ and $\mu_1(\lambda)$ do not overlap, this means that the quantum state $|\Psi\rangle$ can be inferred uniquely from the physical state of the system and hence satisfies the above definition of a physical property. However, if $\mu_0(\lambda)$ and $\mu_1(\lambda)$ overlap for at least one pair of quantum states, then $|\Psi\rangle$ can be regarded as simply information/knowledge. The PBR theorem shows that for distinct quantum states $|\Psi_0\rangle$ and $|\Psi_1\rangle$, if the distributions $\mu_0(\lambda)$ and $\mu_1(\lambda)$ overlap then there is a contradiction with the predictions of quantum theory. In other terms, the wave vector indeed should be seen as a physical property.

Before we sketch the general proof, let's look at an example. Consider that $|\Psi_0\rangle = |0\rangle$ and $|\Psi_1\rangle = |+\rangle$ and thus with a overlap $\langle 0|+\rangle = 1/\sqrt{2}$. If $\mu_0(\lambda)$

and $\mu_1(\lambda)$ overlap then there exists a $q > 0$ such that either of the state results in a λ from the overlap region with probability at least q. Now consider we have two independent preparation devices (each preparing $|\Psi_0\rangle$ or $|\Psi_1\rangle$) that are measured jointly. So, with probability $q^2 > 0$ the physical states λ_1 (first preparation device) and λ_2 (second preparation device) are from the overlap region. The four possible states being prepared (by both devices) are then: $|00\rangle$, $|0+\rangle$, $|+0\rangle$ or $|++\rangle$.

Consider now that these states are measured in the following measurement basis:

$$
\begin{aligned}
|\varepsilon_1\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \\
|\varepsilon_2\rangle &= \frac{1}{\sqrt{2}}(|0-\rangle + |1+\rangle), \\
|\varepsilon_3\rangle &= \frac{1}{\sqrt{2}}(|+1\rangle + |-0\rangle), \\
|\varepsilon_4\rangle &= \frac{1}{\sqrt{2}}(|+-\rangle + |-+\rangle).
\end{aligned}
\tag{4.9}
$$

Using Born's rule it is not difficult to see that

$$
\begin{aligned}
&\langle 00|\varepsilon_1\rangle = 0, \\
&\langle 0+|\varepsilon_2\rangle = 0, \\
&\langle +0|\varepsilon_3\rangle = 0, \\
&\langle ++|\varepsilon_4\rangle = 0.
\end{aligned}
\tag{4.10}
$$

We have just arrived at a contradiction. If $\mu_0(\lambda)$ and $\mu_1(\lambda)$ indeed overlap, then at least with probability $q^2 > 0$ we should be uncertain of which of the states has been prepared. However, quantum theory predicts that for certain measurement outcomes we can be certain that some preparations have not been performed, for instance, we we obtain $|\varepsilon_1\rangle$ we are sure the prepared state was not $|00\rangle$. The proof extends to any two different quantum states (we only have to find a suitable measurement such that one of the outcomes excludes one of the possible preparations to arrive at the same contradiction). This means, that our original assumption that $\mu_0(\lambda)$ and $\mu_1(\lambda)$ overlap must be wrong. The quantum state $|\Psi\rangle$ can be uniquely identified from its distribution $\mu_\psi(\lambda)$. Notice, that to arrive at this conclusion we are using two assumptions. First, that a quantum system has a real physical state (the state λ). That is the reason why the PBR theorem says nothing about the Copenhagen interpretation. Second, systems prepared independently have independent physical states.

5 Causality and Quantum Mechanics

5.1 Superluminal communication? The no-cloning theorem says no way!

The terminology quantum non-locality is often misleading as it seems to imply that quantum mechanics would allow for superluminal communication. Clearly, that is not the case (even though many science media outlets often say the contrary). Even if not possible, considerations about superluminal communication exploiting quantum entanglement is at the origin of a famous no-go theorem in quantum information: the no-cloning theorem [1]. In the following we consider the original superluminal protocol designed at Ref [2] employing a copying device that soon afterwards was shown to be incompatible with the quantum mechanical rules [3].

Let's first devise an apparent superluminal protocol. Consider the following setup: Alice and Bob share a maximally entangled state $|\Psi^+\rangle = (1/\sqrt{2})(|00\rangle + |11\rangle)$ and Alice has two possible measurement setups, where she is going to measure either σ_z or σ_x on her share of the state. Let us decompose the entangled state according to the measurement basis of Alice. We have two possibilities that should be indistinguishable:

$$|\Psi^+\rangle = (1/\sqrt{2})(|00\rangle + |11\rangle) \tag{5.1}$$
$$= (1/\sqrt{2})(|++\rangle + |--\rangle) \tag{5.2}$$

But know let us suppose that we have a special device, a cloning machine that can copy perfectly any quantum state. In particular, it can produce many copies of the measurement basis used by Alice, for example, $|0\rangle \to |0\rangle^{\otimes N}$ and $|+\rangle \to |+\rangle^{\otimes N}$.

If we apply this cloning machine to the two decompositions of $|\Psi^+\rangle$ we get

$$\to (1/\sqrt{2})(|0\rangle|0\rangle^{\otimes N} + |1\rangle|1\rangle^{\otimes N}), \tag{5.3}$$
$$\to (1/\sqrt{2})(|+\rangle|+\rangle^{\otimes N} + |-\rangle|-\rangle^{\otimes N}). \tag{5.4}$$

But these two states are not equivalent! In particular, if Bob always measure in the computational basis all the N clones of his state, he is going to observe a different statistics depending on whether Alice measured σ_z or σ_x.

[1] DGBJ Dieks. Communication by epr devices. *Physics Letters A*, 92(6):271–272, 1982; and William K Wootters and Wojciech H Zurek. A single quantum cannot be cloned. *Nature*, 299(5886):802, 1982

[2] Nick Herbert. FlashâĂŤa superluminal communicator based upon a new kind of quantum measurement. *Foundations of Physics*, 12(12):1171–1179, 1982

[3] DGBJ Dieks. Communication by epr devices. *Physics Letters A*, 92(6):271–272, 1982; and William K Wootters and Wojciech H Zurek. A single quantum cannot be cloned. *Nature*, 299(5886):802, 1982

For instance, if σ_z all the N qubits will either give $|0\rangle$ or $|1\rangle$. However, for a σ_x half of the measurement will return $|0\rangle$ and the other half $|1\rangle$. Apparently, using this cloning machine we can communicate faster than light! Can we, really?

Our conclusion is based on the assumption that we have a perfect clone machine, that in particular can clone orthogonal states ($|0\rangle$ and $|+\rangle$) perfectly. Does quantum mechanics allow for this device? The negative answer to this question is known as the no-cloning theorem [4], that we know prove.

[4] William K Wootters and Wojciech H Zurek. A single quantum cannot be cloned. *Nature*, 299(5886):802, 1982

Our general purpose machine (describe by a unitary evolution U) takes as input the state $|\psi\rangle$ to be cloned and using some ancillary system produces two (or more copies of it):

$$|\psi\rangle \otimes |s\rangle \to U(|\psi\rangle \otimes |s\rangle) = |\psi\rangle \otimes |\psi\rangle. \tag{5.5}$$

If we now consider the action of this hypothetical device on two different input states we have

$$|\psi\rangle \otimes |s\rangle \to U(|\psi\rangle \otimes |s\rangle) = |\psi\rangle \otimes |\psi\rangle, \tag{5.6}$$
$$|\phi\rangle \otimes |s\rangle \to U(|\phi\rangle \otimes |s\rangle) = |\phi\rangle \otimes |\phi\rangle. \tag{5.7}$$

If we take the inner product of both sides of the equations above we obtain that for any machine U we should respect

$$\langle\psi|\phi\rangle = (\langle\psi|\phi\rangle)^2, \tag{5.8}$$

a simple quadratic equation of the form $x = x^2$ and that has only two solutions $x = 0$ or $x = 1$. That is, such a hypothetical cloning machine U only works if the input states are the same ($\langle\psi|\phi\rangle = 1$) or are orthogonal ($\langle\psi|\phi\rangle = 0$). No quantum mechanical evolution allows for the perfect cloning of non-orthogonal states! Thus, no superluminal communication.

5.2 Tsirelson's bound

Sure, quantum mechanics does not allow for superluminal communication but given Bell's theorem, we know that some of our intuitive ideas about cause and effect have to give way in the description of quantum systems. Accepting Bell non-locality as a fact, we can instead of trying to simulate quantum correlations with classical physics to ask ourselves how much non-local can Nature be? To make the question more concrete: given some Bell inequality, what is the maximum violation of it allowed by quantum mechanics?

The first to ask this question, was Boris Tsirelson in the seminal paper [5]. Considering the paradigmatic CHSH inequality,

[5] Boris S Cirel'son. Quantum generalizations of bell's inequality. *Letters in Mathematical Physics*, 4(2):93–100, 1980

$$\langle \text{CHSH} \rangle = \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle \leq 2 \tag{5.9}$$

what is the maximum quantum violation allowed by it? One of the quantum mechanics postulates (Born's rule), states that any quantum probability distribution should be of the form

$$p(a,b|x,y) = \mathrm{Tr}\left[\left(M_a^x \otimes M_b^y\right)\rho\right], \tag{5.10}$$

where M_a^x and M_b^y are the measurement operators of Alice and Bob and ρ is their shared state. Remember also that $\langle A_x B_y \rangle = \sum a,b(-1)^{a+b} p(a,b|x,y)$ is the expectation value of the joint measurement outcomes of Alice and Bob (assumed to take the values $a,b = 0,1$).

The problem we have at hand is thus the following: maximize a linear function of $p(a,b|x,y)$ given that it should be of the form (5.10). As it turns out, (5.10) defines a convex set, however, the precise characterization of this set is yet a stone in our shoes (we are optimizing over all quantum states and measurements, including potentially infinite dimensional operators!). The best description we have so far, is a hierarchy of semi-definite approximations that only asymptotically converge to the quantum set [6]. In other terms, in general, the best we can hope for is to put upper bounds to the maximum quantum violation of a given Bell inequality. We can also straightforwardly obtain a lower bound to the maximum quantum violation, choosing specific measurements and a specific quantum state. When both these lower and upper bounds coincide, we have our precise quantum maximum.

[6] Miguel Navascués, Stefano Pironio, and Antonio Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7):073013, 2008

In the specific case of the CHSH inequality (5.9) we easily obtain an upper bound noticing the following. As we are not restricting the dimension of our quantum states, we can without loss of generality assume our measurements to be projective (this is a direct consequence of Neumark's theorem [7]). If the measurements are projective, then it follows that $A_x^2 = A_x$ and $B_y^2 = B_y$. Using that we can simply the square of the CHSH operator to the following expression

[7] Asher Peres. Neumark's theorem and quantum inseparability. *Foundations of Physics*, 20(12):1441–1453, 1990

$$\mathrm{CHSH}^2 = 4\mathbb{I} - [A_0,A_1]\,[B_0,B_1]\,. \tag{5.11}$$

In the case where the measurement operator commute of either Alice or Bob commute, we recover the usual classical bound $\langle \mathrm{CHSH} \rangle \leq 2$. In the quantum case we notice that norm of the commutator is bounded as $||\,[A_0,A_1]\,|| \leq 2||A_0||||A_1|| \leq 2$ to obtain that

$$\mathrm{CHSH}_{\mathrm{Q}}^2 \leq 8, \tag{5.12}$$

implying that

$$\langle \mathrm{CHSH} \rangle_{\mathrm{Q}} \leq 2\sqrt{2}, \tag{5.13}$$

that is precisely the value (known as Tsirelson's bound) we have obtained by considering the maximum violation with two-qubit entangled states and projective measurements.

5.3 Popescu-Rohrlich's boxes

Considering the simplest possible Bell scenario (the CHSH scenario), our conclusions so far can be summarized as

$$\langle \text{CHSH} \rangle \overset{\text{L}}{\leq} 2 \overset{\text{Q}}{\leq} 2\sqrt{2}, \tag{5.14}$$

where the letter over the inequality specifies the local and quantum sets (L and Q, respectively). The local description follows from our everyday intuitive concepts of cause and effect. The quantum description from Born's rule (one of the postulates of quantum mechanics).

Let's now add a third possible description. Forget about quantum mechanics and local realism for the moment. Suppose that after taking a course on special relativity you are asked to describe what are the possible correlations (probability distributions) in the Bell scenario. Remember, Alice and Bob, are far apart, space-like separated from each other. That is, whatever Alice is doing on her lab cannot have any influence on what Bob is observing in his own distant lab (and vice-versa). In general, the measurement outcomes of Alice and Bob are going to be correlated, but these correlations are only due to the source at their common past. Mathematically, special relativity implies the so-called no-signalling constraints:

$$\begin{aligned} p(a|x) &= \sum_b p(a,b|x,y) = \sum_b p(a,b|x,y'), \\ p(b|y) &= \sum_a p(a,b|x,y) = \sum_a p(a,b|x',y). \end{aligned} \tag{5.15}$$

The no-signalling conditions basically state that the marginal distribution for Alice and Bob are well-defined. In other terms, what Alice observes locally (the measurement outcomes a) cannot depend in any way of Bob's choice y (and vice-versa). Let's call the set of correlations compatible with the no-signalling conditions NS. Clearly,

$$\text{L} \subset \text{Q} \overset{??}{\subset} \text{NS}, \tag{5.16}$$

that is, the local and quantum descriptions respect the no-signalling constraints. Moreover, we know the first inclusion to be proper (there are quantum correlations beyond what is achievable in a classical description). What about the second inclusion relation? Does the quantum and no-signalling sets coincide?

In fact, they don't. That is, there are correlations (more generally, probabilistic theories) compatible with special relativity but still beyond what is achievable with quantum mechanics. In other terms, special relativity is not enough to single out quantum mechanics from the many possible probabilistic theories. To show that, it is enough to find a NS correlation surpassing the Tsirelson's bound, since it is respected by quantum correlations. A paradigmatic example is the so called Popescu-Rohrlich(PR)-box [8], defined as:

[8] Sandu Popescu and Daniel Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, 1994

$$p(a,b|x,y) = \frac{1}{2}\delta_{a\oplus b,xy}. \tag{5.17}$$

That is, the PR-box is such that the outcomes are equal (either 0 or 1 with probability half) if any of the measurements choices are different from 1 ($x \neq 1$ or $y \neq 1$) and the outcomes are different if both the measurements are equal to 1 ($x = y = 1$). It is easy to convince yourself that this correlation is no-signalling (respect (5.15)) and achieves CHSH $= 4$ (the algebraic maximum of the CHSH operator). Summing up our knowledge so far:

$$\mathrm{L} \subset \mathrm{Q} \subset \mathrm{NS}, \tag{5.18}$$

that is, the local is a proper subset of the quantum that by itself is also a proper subset of NS set. It is not difficult to see that the no-signalling constraints (5.15) define a convex set, again a polytope as in the case of classical (local) correlations.

5.4 Information Causality

Let's make a quick and superficial comparison between quantum mechanics and the special relativity theory. All the strange consequences of relativity (clocks that tick differently, the relativity of simultaneity, etc) are direct consequences of two simple and physically well motivated postulates. Quantum mechanics in its turn leads to counter-intuitive concepts such as wave-particle duality, uncertainty principle, entanglement. As opposed to relativity, however, these are consequences of highly abstract and not physically motivated (apart from the fact that they work) postulates. The natural question is then: is there any more intuitive way we can try to understand quantum mechanics? As we have seen above, special relativity alone is not enough to single out quantum correlations. Is there any other principle we can introduce in order to recover Tsirelson's bound? Notice that here we are not interested in any dynamics, just the possible results of measurements.

There have been many attempts of information theoretical principles introduced to explain why nature is described by quantum theory rather than some post-quantum one (with a higher degree of non-locality). Among these, the so called information causality (IC) principle [9] is among the most famous ones. Before we can state it, we first have to introduce the famous measure of information introduce by Shannon.

[9] Marcin Pawłowski, Tomasz Paterek, Dagomir Kaszlikowski, Valerio Scarani, Andreas Winter, and Marek Żukowski. Information causality as a physical principle. *Nature*, 461(7267):1101, 2009

5.4.1 A crash course on information theory

The Shannon entropy (see [10] for further details) of a random variable X defined as

$$H(X) = -\sum_x p(x)\log p(x), \tag{5.19}$$

where the log is taken to be in the basis 2 (with $0\log 0 = 0$). To illustrate, consider a variable X that can assume only two possible values $x = 0$ and

[10] Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002; and Raymond W Yeung. *Information theory and network coding*. Springer Science & Business Media, 2008

$x = 1$. Consider now that X is a deterministic variable, it always assume the value $x = 0$, that is, $p(x = 0) = 1$ and $p(x = 1) = 0$. It is easy to see that the Shannon entropy in this case is $H(X) = 0$. Consider now that both possibilities are equally probable, that is, $p(x = 0) = p(x = 1) = 1/2$. In this case, we have $H(X) = 1$. In the first case, we have a deterministic process, associated with a null entropy. In the other, we have a process with maximum uncertainty associated with maximum entropy. This illustrates the fact that the Shannon entropy is associated with the uncertainty of a given variable X before we know its value x. Alternatively, we can understand the Shannon entropy as the amount of information we gain after discovering the value x of X. For instance, in the deterministic process we already know what will be the value of X, we gain no new information by reading it. In contrast, in the second case, we have maximum uncertainty, by reading the value x we indeed gain 1 bit of information.

There are a few straightforward extensions of the Shannon entropy. For instance for two random variables X and Y we can define the Shannon entropy

$$H(X,Y) = -\sum_{x,y} p(x,y) \log p(x,y). \tag{5.20}$$

We can also define the conditional entropy

$$H(X|Y) = -\sum_{x,y} p(x,y) \log p(x|y), \tag{5.21}$$

a measure of how much uncertainty we have about X given that we know Y. Similarly we can also define a measure correlation, the mutual information defined as

$$I(X:Y) = -\sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}. \tag{5.22}$$

Finally, we can also define a measure of conditional correlation, the conditional mutual information, defined as

$$I(X:Y|Z) = -\sum_{x,y,z} p(x,y,z) \log \frac{p(x,y|z)}{p(x|z)p(y|z)}. \tag{5.23}$$

It is not difficult to obtain a few relations between these measures:

$$\begin{aligned} &H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) \qquad (5.24)\\ &I(X:Y) = H(X) + H(Y) - H(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)\\ &I(X:Y|Z) = H(X,Z) + H(Y,Z) - H(X,Y,Z) - H(Z) \end{aligned}$$

A basic result in information theory is the fact that entropies should respect the so called basic inequalities, basically saying that all these information measures introduced above are positive. Inequalities of the type

$$H(X,Y) \geq H(Y) \rightarrow H(X|Y) \geq 0, \tag{5.25}$$

are known as monotonicity inequalities and express the fact that (for classical variables), the uncertainty about two variables should be at least as large as the uncertainty about one of them. The second class of constraints is the positivity of the conditional mutual information

$$I(X:Y|Z) \geq 0 \rightarrow H(X,Z) + H(Y,Z) \geq H(X,Y,Z) + H(Z), \tag{5.26}$$

also known as strong subadditivity (SSA).

In the quantum case, we replace probability distributions by density operators then defining the quantum version of entropy, the von Neumann entropy defined as

$$S(\rho) = -\mathrm{Tr}\left(-\rho \log \rho\right). \tag{5.27}$$

If we use the spectral decomposition $\rho = \sum_i p_i |i\rangle\langle i|$ we see that the von Neumann entropy is simply given by the Shannon entropy of the eigenvalues p_i of the state ρ. One can also define joint entropies, conditional entropies and mutual information in the quantum case. The important difference here is the fact that although the von Neumann entropy respects the SSA constraint it can violate the monotonicity one. This is a consequence of entanglement, implying that the uncertainty of the whole can be smaller than the uncertainty of its parts.

5.4.2 Information Causality as a information theoretical principle

To understand information causality [11] let's consider the following task. Alice would like to send 2 bits of information (described by the random variables X_0 and X_1, for which we will assume $H(X_0,X_1) = H(X_0) + H(X_1)$, that is, these two variables are independent) to Bob. However, Alice has a channel with limited capacity

[11] Marcin Pawłowski, Tomasz Paterek, Dagomir Kaszlikowski, Valerio Scarani, Andreas Winter, and Marek Żukowski. Information causality as a physical principle. *Nature*, 461(7267):1101, 2009

$$H(M) < H(X_0) + H(X_1) \tag{5.28}$$

That is, Alice can encode the bits to be sent into a message M, but the bits X_0 and X_1 contain more information than it is possible to encode in the message (for instance, the message could be a single bit). Upon receiving the message, Bob applies a decoding protocol to generate a random variable G containing as much information as possible about X_0 and X_1. What they want is to maximize $I(X_0,X_1 : G)$, the mutual information between Bob's guess G and the input bits of Alice. To the aim, they can also use some pre-established correlations, classical randomness, quantum states or even post-quantum correlations. It is possible to show that for all these kind of shared correlations it follows that

$$I(X_0,X_1 : G) \leq H(M). \tag{5.29}$$

As expected, we cannot use a classical message to transmit more information than that contained in the message, irrespectively of the which kind of correlations the parties are sharing.

Let's consider a small twist to this protocol. Instead of maximizing its information about X_0 and X_1 jointly, the aim of Bob now is to guess only one of them, but this choice of which bit to guess at each round is made randomly (that is, both Alice and Bob do not now before hand which bit, X_0 or X_1 Bob should try to guess). In this case our figure of merit is given by $I(X_0 : G_0) + I(X_0 : G_1)$, where G_0 is the guess of Bob given that he decided to recover the information contained in the bit X_0 (similarly for G_1). It seems, reasonable to expect that

$$I(X_0 : G_0) + I(X_0 : G_1) \leq H(M), \tag{5.30}$$

that is, the amount of information available to Bob about the input bits cannot be larger that that contained in the message sent by Alice. Indeed, as we will see below classical and quantum correlations respect this constraint. However, there are non-signalling correlations that can violate it. In fact, considering a slightly more complicated version of this protocol, it can be shown that any correlation violating the CHSH inequality beyong the quantum maximum of $2\sqrt{2}$ will also violate the IC inequality. Nicely, even though these post-quantum theories are compatible with special relativity, they violate a basic premise at the interface between causality and information theory.

To exemplify, we consider a PR-box shared between Alice and Bob. The protocol proceed as follows. Alice uses as input to her side of the PR-box $x_0 \oplus x_1$. Alice obtains an outcome a and uses that to produce the message to be sent to Bob as $m = x_0 \oplus a$ Bob uses as input $y = 0, 1$, choosing which bit X_y he wants to retrieve in a given round of the experiment. The measurement outcome of Bob is b and he uses as a decoding (his guess) the function $g = m \oplus b$.

Using the properties of the PR-box, we see that if $y = 0$ then $a \oplus b = 0$. Using that $a = m \oplus x_0$ and $b = g \oplus m$ we see that $x_0 \oplus g = 0$, that is, $g = x_0$ and thus Bob can perfectly recover the bit X_0. If $y = 1$ then $a \oplus b = x_0 \oplus x_1$. Proceeding as before we obtain, $x_0 \oplus g = x_0 \oplus x_1$ and then $g = x_1$, that is, Bob can perfectly recover the bit X_1. We have that $I(X_0 : G_0) = I(X_1 : G_1) = 1$ and $H(M) = 1$, thus violating the IC inequality. In spite of the fact that Alice is only sending one bit of information, Bob can decide which of the two bits he can decode.

5.4.3 Information Causality inequality proof

In the IC scenario we have a few variables of interest, some are classical, some are quantum. On the classical side we have the bits of Alice X_0 and X_1, the message M sent by Alice, the input Y of Bob (deciding which bit he will try to recover) and the guess G made by Bob. On the quantum side we have the quantum state ρ_{AB} shared between Alice and Bob and its marginals ρ_A and ρ_B. Importantly, we have the independence relation between the input bits of Alice and the shared state, in particular $I(X_0, X_1 : \rho_B) = 0$, that is,

$S(X_0,X_1,\rho_B) = S(X_0,X_1) + S(\rho_B) = H(X_0,X_1) + S(\rho_B)$ (notice that we have the entropy of a mixed object, part classical and part quantum).

Let's start rewriting the left side of the IC inequality as

$$\begin{aligned} I(X_0 : G_0) + I(X_1 : G_1) \leq & \quad I(X_0 : M,\rho_B) + I(X_1 : M,\rho_B) \\ = & \quad H(X_0) + H(X_1) + 2S(M,\rho_B) - S(X_0,M,\rho_B) - S(X_1,M,\rho_B) \end{aligned} \tag{5.31}$$

Using the strong subadditivity property

$$S(X_0,M,\rho_B) + S(X_1,M,\rho_B) \leq S(M,\rho_B) + S(X_0,X_1,M,\rho_B) \tag{5.32}$$

we can rewrite (5.31) as

$$I(X_0 : G_0) + I(X_1 : G_1) \leq H(X_0) + H(X_1) + S(M,\rho_B) - S(X_0X_1,M,\rho_B) \tag{5.33}$$

Using the monotonicity property $S(X_0X_1,M,\rho_B) \geq S(X_0X_1,\rho_B)$ we can rewrite

$$I(X_0 : G_0) + I(X_1 : G_1) \leq H(X_0) + H(X_1) + S(M,\rho_B) - S(X_0X_1,\rho_B) \tag{5.34}$$

Using the fact that $S(X_0,X_1,\rho_B) = H(X_0X_1) + S(\rho_B)$ we arrive at

$$\begin{aligned} I(X_0 : G_0) + I(X_1 : G_1) \leq & \quad H(X_0) + H(X_1) - H(X_0X_1) + S(M,\rho_B) - S(\rho_B) \\ \leq & \quad I(X_0 : X_1) + H(M) - I(M : \rho_B) \\ \leq & \quad H(M), \end{aligned} \tag{5.35}$$

where in the last step we have used the fact that we assume $I(X_0 : X_1) = 0$ (the variables are considered independent) and also used that $-I(M : \rho_B) \leq 0$. We have then proved the IC inequality.

6 Appendix: Probability Theory Basics

Quantum mechanics (QM) is a probabilistic model of Nature that assigns to the state of a physical system a vector in a complex inner-product vector space. Given a system described by a d-dimensional vector $\Psi \in \mathbb{C}^d$, the probability of finding the system in the state $\Phi \in \mathbb{C}^d$ is given by the modulus square of the inner-product between these two vectors, i.e., $|\langle \Phi, \Psi \rangle|^2$.

It's clear from this very concise, and rather abstract, description of Quantum Mechanics, that two branches of mathematics will be strongly employed during this course. They are: *i)* probability theory, and *ii)* linear algebra. In this appendix, and the next two, we'll see a quick'n'dirty review of the main elements of these topics that will be of importance for us here.

A sample space is a set containing all possible outcomes of an "experiment". The outcomes are mutually exclusive. Given a sample space Ω (here assumed countable for simplicity), the function $\Pr : \Omega \mapsto [0,1]$ is said to be a *probability distribution* if

$$\sum_{e \in \Omega} \Pr(e) = 1. \tag{6.1}$$

The probability of an element $e \in \Omega$ gives the likelihood of e to happen.

An *event* E is any subset of the sample space Ω. In this way, the probability of a given event to happen is:

$$\Pr(E) = \sum_{e \in E} \Pr(e). \tag{6.2}$$

A (numerical) *random variable* X is a function from Ω to the real numbers. For such random variables we can define their expectation value $\langle \cdot \rangle$ as:

$$\langle X \rangle = \sum_{e \in \Omega} X(e) \Pr(e). \tag{6.3}$$

Example 1: Consider the throw of a coin, such that the probability of getting heads is p_h, and to get tails is $p_t = 1 - p_h$. The sample space is thus $\Omega = \{\texttt{head}, \texttt{tail}\}$. Now define the random variable X as $X(\texttt{head}) = 1$ and $X(\texttt{tail}) = -1$. The expectation value of X is therefore $\langle X \rangle = 1.p_h + (-1).p_t = 2p_h - 1$. As expected, if the coin is unbiased, i.e., $p_h = p_t = 1/2$, then $\langle X \rangle = 0$.

Another way of "seeing" random variables, is that they define a new sample space $\mathscr{X} := X(\Omega) = \{X(e) | e \in \Omega\}$. In this way, the probability of an

event $x \in \mathscr{X}$ is given by

$$\Pr(x) = \sum_{e|X(e)=x} \Pr(e). \tag{6.4}$$

Random variables are thus *labelling* functions for the original sample space — note that X does not even need to be defined over the whole sample space Ω. Observe that within this way of seeing things, the expected value of a random variable is simply given by $\sum x \Pr(x)$, with the sum taken over all $x \in \mathscr{X}$.

A *joint* probability distribution of multiple random variables is defined over the Cartesian product of the multiple sample spaces $\Pr : \mathscr{X} \times \mathscr{Y} \times \ldots \times \mathscr{Z} \mapsto [0,1]$, such that

$$\sum_{x\in\mathscr{X}} \sum_{y\in\mathscr{Y}} \cdots \sum_{z\in\mathscr{Z}} \Pr(x,y,\ldots,z) = 1. \tag{6.5}$$

For simplicity, from now on we look at distributions over the Cartesian product of just two sample spaces. Given a probability distribution $\Pr_{XY} : \mathscr{X} \times \mathscr{Y} \mapsto [0,1]$, the *marginal* distribution $\Pr_X : \mathscr{X} \mapsto [0,1]$, is given by:

$$\Pr_X(x) = \sum_{y\in\mathscr{Y}} \Pr_{XY}(x,y). \tag{6.6}$$

A pair of events $(x,y) \in \mathscr{X} \times \mathscr{Y}$ is said *uncorrelated* or *independent* if

$$\Pr_{XY}(x,y) = \Pr_X(x) \Pr_Y(y). \tag{6.7}$$

Otherwise, the events are *correlated*. The random variables X and Y are said uncorrelated, if every event pair in $\mathscr{X} \times \mathscr{Y}$ is uncorrelated.

Given a joint distribution, one modifies its state of knowledge about the probability of an event in $\mathscr{X}$ to happen given that an event in $\mathscr{Y}$ already did happen. The updated probability distribution is called *conditional* probability distribution, and the update is given by *Bayes rule*:

$$\Pr_X(x|y) = \frac{\Pr_{XY}(x,y)}{\Pr_Y(y)}. \tag{6.8}$$

The left hand side of the equation above is read as "the probability of x given y". Clearly, if x and y are independent events, then $\Pr_X(x|y) = \Pr_X(x)$, i.e., one learns nothing about x given y.

6.0.1 The (weak) law of large numbers

As QM is a probabilistic model, any experiment need to be repeated many times in order to estimate the value of the measurement being made. It is thus important to determine how close (in probability) the estimator is from the expected value.

We start by deriving the so called *Markov inequality*. Assume for the moment that a random variable X only leads to non-negative values, i.e., every

element in $\mathscr{X}$ is non-negative. Symbolically one writes $X \geq 0$. Moreover, let $\mu = \langle X \rangle$. Then, for any $\eta > 0$ it follows that

$$\begin{aligned} \Pr(X > \eta) &= \sum_{x>\eta} \Pr(x) \\ &< \sum_{x>\eta} \frac{x}{\eta} \Pr(x) \\ &\leq \frac{1}{\eta} \sum_{x \in \mathscr{X}} x \Pr(x) \\ &= \frac{\mu}{\eta}. \end{aligned} \tag{6.9}$$

Now forget about the positivity restriction on X, but define the random variable $(X-\mu)^2$. Surely this new random variable is non-negative, and its mean value is $(\langle X^2 \rangle - \mu^2) := \mathrm{var}(X)$, i.e., the variance of X that we call σ^2. Therefore, using Markov's inequality above one obtains:

$$\Pr\left((X-\mu)^2 > \eta^2\right) < \frac{\sigma^2}{\eta^2}. \tag{6.10}$$

Which is equivalent to:

$$\Pr\left(|X-\mu| > \eta\right) < \frac{\sigma^2}{\eta^2}, \tag{6.11}$$

known as the Chebyshev's inequality.

Finally, take N independent identically distributed (i.i.d.) random variables X_i such that $\langle X_i \rangle = \mu$ and $\mathrm{var}(X_i) = \sigma^2$, with $i \in [N]$. Define $\overline{X}_N = (1/N)\sum_i X_i$. The mean of $\overline{X}_N$ is also μ, but its variance is σ^2/N (check that!). Then by Chebyshev's inequality we arrive at:

$$\Pr\left(|\overline{X}_N - \mu| > \eta\right) < \frac{\sigma^2}{N\eta^2}. \tag{6.12}$$

This means that by repeating the experiment N times, the probability of an estimation to be far from the expected value by an amount bigger than η decreases with N. In other words, the estimation concentrates around the expected value with increasing number of runs N of the experiment. Formally, the statement of the weak law of large numbers reads as:

Theorem 5 (The weak law of large numbers)**.** *For every positive ε and δ there exists N such that for the i.i.d. random variables X_1, X_2, ..., X_N, each with finite mean μ and variance σ^2, it holds that*

$$\Pr\left(\left|\frac{1}{N}\sum_{i=1}^{N} X_i - \mu\right| > \varepsilon\right) < \delta. \tag{6.13}$$

Although Eq.(6.12) guarantees that in the limit of $N \to \infty$ the estimation converges (in probability) towards the expected value, the rate of convergence is much faster than predicted by this inequality. Hoeffding's inequality provides a much more stringent bound even with less restrictive assumptions.

Theorem 6 (Hoeffding's inequality)**.** *Let $X_1, X_2, \ldots, X_N$ be independent random variables. Assume that each X_i is almost surely bounded in a non-empty interval $[a_i, b_i]$.* [1] *Then*

$$\Pr\left(\left|\frac{1}{N}\sum_{i=1}^{N} X_i - \mu\right| > \varepsilon\right) < 2\exp\left(-\frac{2N^2\varepsilon^2}{\sum_i (b_i - a_i)^2}\right). \tag{6.14}$$

The proof of this inequality is not difficult, but a little beyond the scope of these notes. The important point to notice, though, is that Hoeffding's inequality predicts an *exponential* concentration around the mean value with increasing N. This is definitely important for experiments, since the number of runs can be dramatically decreased without jeopardizing too much the precision of the measurement.

[1] That means that $\Pr(X_i \in [a_i, b_i]) = 1$.

7 Appendix: Linear Algebra Basics

In quantum mechanics the state of physical systems are described by vectors in complex Hilbert spaces, and transformations on the system are described by linear operators acting on those vectors. It's therefore mandatory to have very clear in mind some basic linear algebra concepts. Here we collect some basic facts that will prove handy on the way.

7.1 Vector Spaces

Definition 7.1.1 (Vector space)**.** *A vector space* V *over the field* $\mathfrak{F}$*, is a set of elements, called vectors, on which two binary operations are defined: i) addition: that for* Ψ *and* Φ *in* V *returns* $\Psi+\Phi\in\mathsf{V}$*; ii) multiplication by a scalar: that for* $\Psi\in\mathsf{V}$ *and* $\alpha\in\mathfrak{F}$ *returns* $\alpha\Psi$*. Furthermore, for all* $\Psi,\Phi,\Upsilon\in\mathsf{V}$ *and* $\alpha,\beta\in\mathfrak{F}$*, the following properties of the two operations hold:*

Addition commutativity	$\Psi+\Phi=\Phi+\Psi$
Associativity	$(\Psi+\Phi)+\Upsilon=\Psi+(\Phi+\Upsilon)$ *and* $(\alpha\beta)\Psi=\alpha(\beta\Psi)$
Zero vector	$\exists 0\in\mathsf{V}$ *such that* $0+\Psi=\Psi$
Additive inverse	$\forall\Psi\in\mathsf{V},\exists(-\Psi)\in\mathsf{V}$ *such that* $\Psi+(-\Psi)=0$
Distributive	$(\alpha+\beta)\Psi=\alpha\Psi+\beta\Psi$ *and* $\alpha(\Psi+\Phi)=\alpha\Psi+\alpha\Phi$
Multiplication identity	$\exists 1\in\mathfrak{F}$ *such that* $1\Psi=\Psi$.

Example 1: For any $d\in\mathbb{N}$ the symbol $\mathbb{R}^d$ denotes the Euclidean d-dimensional vector space over the real numbers. The elements of $\mathbb{R}^d$ are the ordered lists $(c_1,c_2,\ldots,c_d)^\mathsf{T}$, with $c_i\in\mathbb{R}$ for all $i\in[d]$. The binary operations of addition and multiplication by a scalar are the usual ones.

A (vector) *subspace* U of V is a subset $\mathsf{U}\subset\mathsf{V}$ such that U is in itself a vector space. Note that the imposition that U must be a vector space implies that $0\in\mathsf{U}$.

Given a subset $\mathsf{T}\subset\mathsf{V}$, then the $\mathrm{Span}(\mathsf{T})$ is the set containing all the linear combinations of the elements of T. It should be clear that $\mathrm{Span}(\mathsf{T})$ is a subspace of V.

Example 2: Given a non-null vector $\Psi\in\mathsf{V}$, then the set $\mathsf{U}=\{\alpha\Psi|\alpha\in\mathbb{C}\}$ is a subspace of V. Moreover, by definition, the $\mathrm{Span}(\{\Psi\})=\mathsf{U}$.

The vectors in a set U are said to be *linearly independent*, or L.I. for brevity, if no element $u\in\mathsf{U}$ can be written as a linear combination of the other

vectors in U. The set is said *linearly dependent*, or L.D., otherwise. It follows directly from this definition, that a set U is L.I. if and only if $\forall i \in [|U|]$ it holds that $u_i \neq 0$ and the only solution for $\sum_i \alpha_i u_i = 0$, with $\alpha_i \in \mathfrak{F}$, is to set $\alpha_i = 0$ for all i.

Definition 7.1.2 (Basis)**.** *A basis of a vector space* V *is a subset* $\mathsf{B} \subset \mathsf{V}$ *such that* B *is L.I. and* $\mathrm{Span}(\mathsf{B}) = \mathsf{V}$.

Therefore, if B is a basis for V, then there exists a unique way to write each element of V as a linear combination of the vectors of B. If $\mathsf{B} = \{b_1, \ldots, b_d\}$ is a basis for V and $v = \alpha_1 b_1 + \ldots + \alpha_d b_d$, then the coefficients $\alpha_i \in \mathfrak{F}$ are called the *coordinates* of v in the basis B.

It is not difficult to show, although admittedly lengthy, that if a vector space V has a basis with d elements, then any other basis of V is also of size d. The size of a basis B of V, and therefore the size of any basis, is called the *dimension* of V.

Example 3: The set of d L.I. vectors $\{(1,0,\ldots,0)^\mathsf{T}, (0,1,\ldots,0)^\mathsf{T}, \ldots, (0,0,\ldots,1)^\mathsf{T}\}$ forms a basis for the complex vector space $\mathbb{C}^d$. This basis is usually called *canonical basis* among mathematicians, and *computational basis* among computer scientists.

Definition 7.1.3 (Complex Hilbert space)**.** *A complex Hilbert space* $\mathscr{H}$ *is a complex vector space equipped with a sesquilinear form* [1] $\langle \cdot, \cdot \rangle : \mathscr{H} \times \mathscr{H} \mapsto \mathbb{C}$*, called inner-product, that satisfies the following conditions:*

$\forall \Psi, \Phi \in \mathscr{H}$*, it holds that* $\langle \Psi, \Phi \rangle = \langle \Phi, \Psi \rangle^*$

$\forall \Psi, \Phi, \Upsilon \in \mathscr{H}$*, and* $\alpha, \beta \in \mathbb{C}$*, it holds that* $\langle \Psi, \alpha\Phi + \beta\Upsilon \rangle = \alpha\langle \Psi, \Phi \rangle + \beta\langle \Psi, \Upsilon \rangle$

$\forall \Psi \in \mathscr{H}$ *with* $\Psi \neq 0$*, then* $\langle \Psi, \Psi \rangle > 0$.

[1] A sesquilinear form is a binary function that is linear in one argument and antilinear in the other.

The first two properties imply that $\langle \alpha\Psi, \Phi \rangle = \alpha^* \langle \Psi, \Phi \rangle$, and therefore the inner-product is antilinear in the first argument

Example 4: The canonical inner-product in $\mathbb{C}^d$ is defined, for $\Psi = (\psi_1, \psi_2, \ldots, \psi_d)^\mathsf{T}$ and $\Phi = (\phi_1, \phi_2, \ldots, \phi_d)^\mathsf{T}$, as $\langle \Psi, \Phi \rangle = \psi_1^* \phi_1 + \psi_2^* \phi_2 + \cdots + \psi_d^* \phi_d$.

The scalar product induces a *norm*, which in turn defines a *distance*. The norm of a vector $\Psi \in \mathbb{C}^d$ is given by the positive quantity $\|\Psi\| = \sqrt{\langle \Psi, \Psi \rangle}$. The induced distance between two vectors is then $D(\Psi, \Phi) = \|\Psi - \Phi\|$. For a generic function $Dist : X \times X \mapsto \mathbb{R}$, with X some set, to be called a *distance*, for all $r, u, v \in X$ it must abide by the following properties:

Non-negativity:	$Dist(u,v) \geq 0$
Identity of indiscernibles:	$Dist(u,v) = 0$ if and only if $u = v$
Symmetry:	$Dist(u,v) = Dist(v,u)$
Triangle-inequality:	$Dist(u,v) \leq Dist(u,r) + Dist(r,v)$

It is left as exercise to show that the distance $D : \mathbb{C}^d \times \mathbb{C}^d \mapsto \mathbb{R}$ induced by the scalar product is indeed a distance function.

The geometrical picture of the inner product comes about via the very important and useful *Cauchy-Schwarz* inequality:

$$|\langle \Psi, \Phi \rangle| \leq \|\Psi\| \|\Phi\|. \tag{7.1}$$

This inequality is very general, in the sense that it holds true for any type of vector space over any field. Its proof is also left as an exercise. In general, the (smallest) “angle" between two vectors is then defined via the expression

$$\text{angle}(\Psi,\Phi) := \arccos \frac{|\langle\Psi,\Phi\rangle|}{\|\Psi\|\|\Phi\|}, \tag{7.2}$$

which is in the interval $[0,\pi/2]$. This definition matches and generalizes the case of vectors in $\mathbb{R}^d$, and moreover respects the intuition that in the case $\Phi = \alpha\Psi$, i.e., parallel vectors, the “angle" between the two vectors is zero.

Going beyond

The above is the very minimum to get us going. Linear operators and Dirac notation are subject of the next lecture. If you are not very comfortable with the concepts reviewed above, or want to know more about them, I strongly suggest that you take some time doing related exercises in any introductory book on probability theory and linear algebra. For instance you could look at:

- *Quantum processes systems, and information*, by Benjamin Schumacher and Michael Westmoreland. Cambridge University Press. (Appendix A).
- *Álgebra Linear – Coleção Matemática Universitária*, by Elon Lages Lima. IMPA.

We mentioned that QM assigns as the state of a physical system a vector Ψ in a Hilbert space $\mathscr{H}$. The next item in the agenda is to describe how we can act on the system. Within QM this is done by linear operators, and finish by introducing the so called Dirac notation.

7.2 Linear Operators

Let U and V be vector spaces over a field $\mathscr{F}$. A linear transformation $\mathsf{A} : \mathsf{U} \mapsto \mathsf{V}$ is a mapping that associates to each vector $u \in \mathsf{U}$ a vector $\mathsf{A}(u) \equiv \mathsf{A}.u \equiv \mathsf{A}u \in \mathsf{V}$, such that for all $u, u' \in \mathsf{U}$ and $\alpha \in \mathscr{F}$ it holds

$$\begin{aligned} \mathsf{A}(u+u') &= \mathsf{A}u + \mathsf{A}u' \\ \mathsf{A}(\alpha u) &= \alpha \mathsf{A}u. \end{aligned} \tag{7.3}$$

Note that for any linear operator it holds that $\mathsf{A}.0 = 0$. Indeed $A.0 = A.(0+0) = A.0 + A.0$ and therefore $A.0 = 0$.

The sum of linear operators $\mathsf{A}, \mathsf{B} : \mathsf{U} \mapsto \mathsf{V}$, and the product of $\mathsf{A} : \mathsf{U} \mapsto \mathsf{V}$ by a scalar $\alpha \in \mathscr{F}$ are the transformations $\mathsf{A}+\mathsf{B} : \mathsf{U} \mapsto \mathsf{V}$ and $\alpha\mathsf{A} : \mathsf{U} \mapsto \mathsf{V}$, such that

$$\begin{aligned} (\mathsf{A}+\mathsf{B})u &= \mathsf{A}u + \mathsf{B}u \\ (\alpha\mathsf{A})u &= \alpha(\mathsf{A}u). \end{aligned} \tag{7.4}$$

Let $\mathscr{L}(\mathsf{U};\mathsf{V})$ be the set of all linear operators between U and V. Given the above properties it is easy to realize that $\mathscr{L}(\mathsf{U};\mathsf{V})$ is a vector space (see exercises).

One nice thing about linear operators is that their action is fully determined over all the space if one knows how they act on a basis of the domain space.

Theorem 7. *Let* U *and* V *be vector spaces over the field* $\mathscr{F}$*, and* $\mathscr{B}(\mathsf{U}) \equiv \mathscr{B} = \{u_i\}_{i=1}^{dim(\mathsf{U})}$ *be a basis for* U*. If for all* $u_i \in \mathscr{B}$ *we assign a vector* $v_i \in \mathsf{V}$*, then there exists a unique linear transformation* $\mathsf{A} : \mathsf{U} \mapsto \mathsf{V}$*, such that* $\mathsf{A}u_i = v_i$*.*

Proof. First we prove the existence of the required linear operator.

Given any vector $u \in \mathsf{U}$ we can write it as a unique linear combination of the basis vectors:

$$u = \sum_{i=1}^{\dim(\mathsf{U})} \alpha_i u_i \tag{7.5}$$

with all $\alpha_i \in \mathscr{F}$. Now define a transformation $\mathsf{A} : \mathsf{U} \mapsto \mathsf{V}$ setting

$$\mathsf{A}u \equiv \mathsf{A}(\sum_i \alpha_i u_i) = \sum_i \alpha_i v_i. \tag{7.6}$$

Note that we are *not* using linearity, that *is* what we want to prove!

Take another vector $u' \in \mathsf{U}$, with $u' = \sum_i \beta_i u_i$ such that $\mathsf{A}(\sum_i \beta_i u_i) = \sum_i \beta_i v_i$. Then $u + u' = \sum_i (\alpha_i + \beta_i) u_i$. By the definition of A then follows:

$$\begin{aligned}
\mathsf{A}(u+u') &= \mathsf{A}[\sum_i (\alpha_i + \beta_i) u_i] \\
&= \sum_i (\alpha_i + \beta_i) v_i \\
&= \sum_i \alpha_i v_i + \sum_i \beta_i v_i \\
&= \mathsf{A}u + \mathsf{A}u'.
\end{aligned}$$

By the same token:

$$\mathsf{A}(\gamma u) = \mathsf{A}(\gamma \sum_i \alpha_i u_i) = \mathsf{A}(\sum_i \gamma \alpha_i u_i) = \sum_i \gamma \alpha_i v_i = \gamma \sum_i \alpha_i v_i = \gamma \mathsf{A}u. \tag{7.7}$$

We then constructed a linear operator for which $\mathsf{A}u_i = v_i$, as requested.

To show the uniqueness of A, suppose that there exists another linear operator $\mathsf{B} : \mathsf{U} \mapsto \mathsf{V}$ such that $\mathsf{B}u_i = v_i$. Then for all $u \in \mathsf{U}$ we have;

$$\mathsf{B}u = \mathsf{B}(\sum_i \alpha_i u_i) = \sum_i \alpha_i \mathsf{B}u_i = \sum_i \alpha_i \mathsf{v}_i = \mathsf{A}u; \tag{7.8}$$

and therefore $\mathsf{B} = \mathsf{A}$. □

7.2.1 Matrix representation of linear operators

In what follows, we'll see that the above theorem allows us to represent linear operators as matrices. This is specially useful for finite vector spaces, of course.

Let $\mathsf{A} : \mathsf{U} \mapsto \mathsf{V}$ be a linear operator, $\{u_i\}$ a basis for U, and $\{v_i\}$ a basis for V. From the previous theorem we know that for defining A we need only to specify how it acts on each element u_i of a basis. Therefore, setting

$$\mathsf{A}u_i = \sum_j a_{ji} v_j \tag{7.9}$$

for each u_i fully determines A. Moreover, the matrix of coefficients $[a_{ji}]$ is a matrix representation of A.

Example 5: Consider the transformation $\mathsf{X} : \mathbb{C}^2 \mapsto \mathbb{C}^2$, that maps $u_0 \mapsto u_1$, and $u_1 \mapsto u_0$, with $\{u_0, u_1\}$ a basis for $\mathbb{C}^2$. Then

$$\begin{aligned} Xu_0 &= a_{00}u_0 + a_{10}u_1 = 0u_0 + 1u_1 \\ Xu_1 &= a_{01}u_0 + a_{11}u_1 = 1u_0 + 0u_1 \end{aligned}$$

and thus

$$[X] = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{7.10}$$

Where the brackets $[\cdot]$ indicate the matrix representation of the linear operator. We will forget this notation very soon, though.

Note that the matrix representation of a linear operator depends on the choice of bases for the vector spaces.

Example 6: Consider the same linear transformation X of the example above. Now, however, we want to find its matrix representation in the basis $\{u_+ = u_0 + u_1, u_- = u_0 - u_1\}$. As before

$$\begin{aligned} Xu_+ &= X(u_0 + u_1) = u_0 + u_1 = u_+ = a_{++}u_+ + a_{-+}u_- = 1u_+ + 0u_- \\ Xu_- &= X(u_0 - u_1) = u_1 - u_0 = -u_- = a_{+-}u_+ + a_{--}u_- = 0u_+ - 1u_- \end{aligned}$$

and thus

$$[X] = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{7.11}$$

7.2.2 Dual space and adjoint linear transformations

The space $\mathscr{L}(\mathsf{U};\mathbb{C})$, that is, the space containing the linear operators $\mathsf{A} : \mathsf{U} \mapsto \mathbb{C}$, is known as the dual space of U, and it is usually denoted as $\mathsf{U}^\dagger$.

For vector spaces of finite dimension equipped with a scalar product, i.e., finite Hilbert spaces, we can define for each vector a linear operator in the dual space as follows. Let $u \in \mathsf{U}$, we define $u^\dagger : \mathsf{U} \mapsto \mathbb{C}$ as

$$u^\dagger(w) \equiv u^\dagger w = \langle u, w \rangle, \tag{7.12}$$

for all $w \in \mathsf{U}$. It's simple to show that $u^\dagger$ is a linear operator (see exercises).

Example 7: Consider the dual space $\mathscr{L}(\mathbb{C}^d, \mathbb{C})$ with the usual scalar product, and $\{e_1, \ldots, e_d\}$ an orthonormal basis for $\mathbb{C}^d$. Then for each vector $v = \sum_i \alpha_i e_i$ we can define $v^\dagger : \mathbb{C}^d \mapsto \mathbb{C}$ by

$$v^\dagger w = \langle v, w \rangle. \tag{7.13}$$

In this way, writing $w = \sum_i \beta_i e_i$ we have:

$$\begin{aligned} v^\dagger w &= \langle \sum_i \alpha_i e_i, \sum_j \beta_j e_j \rangle \\ &= \sum_i \sum_j \langle \alpha_i e_i, \beta_j e_j \rangle \\ &= \sum_i \sum_j \alpha_i^* \beta_j \langle e_i, e_j \rangle \\ &= \sum_i \alpha_i^* \beta_i. \end{aligned}$$

Therefore, if v is represented by a complex vector in $\mathscr{M}_{d\times 1}$, then $v^\dagger$ is represented by the conjugated transposed vector in $\mathscr{M}_{1\times d}$.

Returning to the matrix representation of operators, consider again the linear transformation $\mathsf{A} : \mathsf{U} \mapsto \mathsf{V}$, and now take $\{u_i\}$ and $\{v_i\}$ as orthonormal bases for U and V, respectively. As before, the operator is fully determined by setting:

$$\mathsf{A}.u_i = \sum_j a_{ji} v_j, \tag{7.14}$$

for all u_i in the basis for U. Since we are now taking orthonormal bases, we can evaluate:

$$\begin{aligned} v_k^\dagger.\mathsf{A}.u_i &= v_k^\dagger (\sum_j a_{ji} v_j) \\ &= \sum_j \langle v_k, a_{ji} v_j \rangle \\ &= \sum_j a_{ji} \langle v_k, v_j \rangle \\ &= a_{ki}. \end{aligned}$$

It is then clear that the linear operator can be written as a linear combination of dyadic products, that is, $\mathsf{A} = \sum_{ij} a_{ij} v_i u_j^\dagger$.

Also using the scalar product of Hilbert spaces, given a linear transformation $\mathsf{A} : \mathsf{U} \mapsto \mathsf{V}$ we can define the transformation $\mathsf{A}^\dagger : \mathsf{V} \mapsto \mathsf{U}$ in such way that $\forall u \in \mathsf{U}$ and $\forall v \in \mathsf{V}$:

$$\langle u, \mathsf{A}^\dagger v \rangle = \langle \mathsf{A}u, v \rangle. \tag{7.15}$$

The operator $\mathsf{A}^\dagger$ is also linear (see exercises), and is called the *adjoint* of A.

By the definition of the adjoint operator it is simple to see that the following properties hold:

$$\begin{aligned} (\mathsf{A}+\mathsf{B})^\dagger &= \mathsf{A}^\dagger + \mathsf{B}^\dagger \\ (\alpha \mathsf{A})^\dagger &= \alpha^* \mathsf{A} \\ (\mathsf{A}\mathsf{B})^\dagger &= \mathsf{B}^\dagger \mathsf{A}^\dagger \\ (\mathsf{A}^\dagger)^\dagger &= A. \end{aligned}$$

For finite dimensions the matrix representations of A and of $\mathsf{A}^\dagger$ are closely related.

Theorem 8. *Let $\{u_1,\ldots,u_m\}$ and $\{v_1,\ldots,v_n\}$ be orthonormal bases for U and V, respectively. If $[a_{ij}] \in \mathscr{M}_{m\times n}$ is a matrix representation of the linear transformation $\mathsf{A} : \mathsf{U} \mapsto \mathsf{V}$, then $[a^*_{ji}] \in \mathscr{M}_{n\times m}$ is the representation of the adjoint operator $\mathsf{A}^\dagger : \mathsf{V} \mapsto \mathsf{U}$.*

Proof. By definition, for all $j \in \{1,\ldots,n\}$ and $i \in \{1,\ldots,m\}$ we have:

$$\mathsf{A}u_j = \sum_{i=1}^{m} a_{ij} v_i$$
$$\mathsf{A}^\dagger v_i = \sum_{j=1}^{m} b_{ji} u_j.$$

We want to determine the coefficients b_{ij} in terms of the a_{ij}.

Since we chose orthonormal bases, then:

$$\begin{aligned}\langle u_k, \mathsf{A}^\dagger v_i\rangle &= \langle u_k, \sum_{j=1}^{n} b_{ji} u_j\rangle \\ &= \sum_{j=1}^{n} b_{ji}\langle u_k, u_j\rangle \\ &= b_{ki}.\end{aligned}$$

On the other hand

$$\begin{aligned}\langle u_k, \mathsf{A}^\dagger v_i\rangle &= \langle \mathsf{A}u_k, v_i\rangle \\ &= \langle \sum_{j=1}^{m} a_{jk} v_j, v_i\rangle \\ &= \sum_{j=1}^{m} a^*_{jk}\langle v_j, v_i\rangle \\ &= a^*_{ik}.\end{aligned}$$

Therefore $b_{ki} = a^*_{ik}$, that is, the matrix representation of the adjoint $\mathsf{A}^\dagger$ is given by the complex-transposition of the matrix representing A.

□

7.3 Dirac notation

Up to now we have used standard linear algebra notation. This notation may become awkward when describing systems defined by many quantum numbers. That is the case, for instance, for the Hydrogen atom, that each state is defined by the total momentum p, and internal quantum numbers n,l,m and perhaps two more for the spins of the proton and electron, leading to u_{p,n,m,l,s_e,s_p}. Not very neat.

To avoid this problem, Dirac introduced the "bra-ket" notation. The state vector for the Hydrogen example is then written as $|p,n,m,l,s_e,s_p\rangle$, and its adjoint as $\langle p,n,m,l,s_e,s_p|$.

The scalar product between two vectors hitherto denoted by $\langle u, v\rangle$, then becomes $\langle u|v\rangle$. The following table gives the translation between the two notations.

	LinAlg ($\mathbb{C}$)	Dirac
vector (column, ket)	v	$\|v\rangle$
co-vector (row, bra)	$v^\dagger$	$\langle v\|$
scalar product	$\langle u, v\rangle$	$\langle u\|v\rangle$
dyadic product	$vu^\dagger$	$\|v\rangle\langle u\|$
linear operation	$\mathsf{A}v$	$\mathsf{A}\|v\rangle$
adjoint operator	$\langle u, \mathsf{A}v\rangle = \langle A^\dagger u, v\rangle$	$\langle u\|(A\|v\rangle) = (\langle u\|A)\|v\rangle$

Further reading

- *Álgebra Linear – Coleção Matemática Universitária*, by Elon Lages Lima. IMPA.
- *Quantum Theory: Concepts and Methods*, by Asher Peres. Kluwer.

8 Appendix: Composition of Hilbert Spaces

We mentioned that QM assigns to the state of a physical system a vector Ψ in a Hilbert space $\mathscr{H}$. But what if the system is composed by many parties, and we want to describe not only the global system but also its components? Below we'll describe different ways to compose Hilbert spaces and theirs properties.

Cartesian product. The simplest way to combine the vectors of two vector spaces, say U and V, is by just juxtaposing the different vectors in a tuple. This composition is simply the Cartesian product between the two vector spaces, i.e., $\mathsf{U} \times \mathsf{V} := \{(u,v) | u \in \mathsf{U} \text{ and } v \in \mathsf{V}\}$. For this composed space the following rules apply:

$\forall u_1, u_2 \in \mathsf{U}$, and $\forall v_1, v_2 \in \mathsf{V}$, holds $(u_1, v_1) + (u_2, v_2) = (u_1 + u_2, v_1 + v_2)$
$\forall u \in \mathsf{U}$, $\forall v \in \mathsf{V}$, and $\alpha \in \mathfrak{F}$, holds $\alpha(u,v) = (\alpha u, \alpha v)$.

It's simple to convince oneself (meaning you should prove it!) that $\mathsf{U} \times \mathsf{V}$ is a vector space with dimension $d_{\mathsf{U}\times\mathsf{V}} = d_{\mathsf{U}} + d_{\mathsf{V}}$. The canonical inner-product between two vectors (u_1, v_1) and (u_2, v_2), both in $\mathsf{U} \times \mathsf{V}$, is given by $\langle (u_1, v_1), (u_2, v_2) \rangle = \langle u_1, u_2 \rangle + \langle v_1, v_2 \rangle$.

> **Example 1**: The obvious example here is the $\mathbb{R}^2 := \mathbb{R} \times \mathbb{R} = \{(x,y) | x, y \in \mathbb{R}\}$. The real plane is the vector space formed by the Cartesian product of two linearly independent real lines.

Direct sum. Now consider the subspaces V_1 and V_2 of a vector space U. The subspace generated by the vectors in $\mathsf{V}_1 \cup \mathsf{V}_2$ is given by the sums $v_1 + v_2$ with $v_1 \in \mathsf{V}_1$ and $v_2 \in \mathsf{V}_2$. This subspace is represented by $V_1 + V_2 := \{v_1 + v_2 | v_1 \in \mathsf{V}_1 \text{ and } v_2 \in \mathsf{V}_2\}$. In the case when the intersection between V_1 and V_2 is given only by the 0 vector (it must be present), one writes $V_1 \oplus V_2$, instead of $V_1 + V_2$, and the subspace is called the *direct sum* between V_1 and V_2. Note that, differently from the Cartesian product, the direct sum can only be employed between subspaces of the same vector space.

As for the Cartesian product, the dimension of $V_1 \oplus V_2$ is $d_{V_1} + d_{V_2}$, and similar rules also apply. The important property of the direct sum is that

for any $w \in V_1 \oplus V_2$ there always exists a unique decomposition in terms of vectors $v_1 \in \mathsf{V}_1$ and $v_2 \in \mathsf{V}_2$. This means that the direct sum splits a vector space into orthogonal subspaces. This can be realized by evaluating the scalar product between $v_1 + v_2$ and $w_1 + w_2$ for $v_1, w_1 \in \mathsf{V}_1$ and $v_2, w_2 \in \mathsf{V}_2$:

$$\begin{aligned}\langle v_1 + v_2, w_1 + w_2\rangle &= \langle v_1, w_1\rangle + \langle v_1, w_2\rangle + \langle v_2, w_1\rangle + \langle v_2, w_2\rangle \\ &= \langle v_1, w_1\rangle + \langle v_2, w_2\rangle.\end{aligned}$$

Where we used that $\langle v_1, w_2\rangle = \langle v_2, w_1\rangle = 0$, as the inner-product between vectors from disjoint subspaces (or by the zero vector) is zero.

Example 2: Let V_1 and V_2 be subspaces of $\mathbb{R}^4$ generated by $\{(1,0,0,0)^\mathsf{T}, (0,0,1,0)^\mathsf{T}\}$ and $\{(0,1,0,0)^\mathsf{T}, (0,0,0,1)^\mathsf{T}\}$, respectively. The vectors in V_1 are therefore of the form $(\alpha_1, 0, \beta_1, 0)^\mathsf{T}$ and the ones in V_2 are of the form $(0, \alpha_2, 0, \beta_2)^\mathsf{T}$, with $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{R}$. Clearly $\mathsf{V}_1 \cap \mathsf{V}_2 = \{(0,0,0,0)^\mathsf{T}\}$, and $\mathsf{V}_1 \oplus \mathsf{V}_2 = \mathbb{R}^4$.

Tensor product. A general and formal definition of the tensor product between vector spaces is beyond the scope of this course. However, given the importance of this type of composition between spaces, we will restrict here to the Kronecker product between vector spaces and, following the literature, call it tensor product.

Given matrices $A \in \mathscr{M}_{m,n}$ and $B \in \mathscr{M}_{p,q}$, the tensor product between them, denoted by $A \otimes B$, is given by:

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}. \tag{8.1}$$

Or more explicitly:

$$A \otimes B = \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & \cdots & a_{11}b_{1q} & \cdots & \cdots & a_{1n}b_{11} & a_{1n}b_{12} & \cdots & a_{1n}b_{1q} \\ a_{11}b_{21} & a_{11}b_{22} & \cdots & a_{11}b_{2q} & \cdots & \cdots & a_{1n}b_{21} & a_{1n}b_{22} & \cdots & a_{1n}b_{2q} \\ \vdots & \vdots & \ddots & \vdots & & & \vdots & \vdots & \ddots & \vdots \\ a_{11}b_{p1} & a_{11}b_{p2} & \cdots & a_{11}b_{pq} & \cdots & \cdots & a_{1n}b_{p1} & a_{1n}b_{p2} & \cdots & a_{1n}b_{pq} \\ \vdots & \vdots & & \vdots & \ddots & & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & & \ddots & \vdots & \vdots & & \vdots \\ a_{m1}b_{11} & a_{m1}b_{12} & \cdots & a_{m1}b_{1q} & \cdots & \cdots & a_{mn}b_{11} & a_{mn}b_{12} & \cdots & a_{mn}b_{1q} \\ a_{m1}b_{21} & a_{m1}b_{22} & \cdots & a_{m1}b_{2q} & \cdots & \cdots & a_{mn}b_{21} & a_{mn}b_{22} & \cdots & a_{mn}b_{2q} \\ \vdots & \vdots & \ddots & \vdots & & & \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{p1} & a_{m1}b_{p2} & \cdots & a_{m1}b_{pq} & \cdots & \cdots & a_{mn}b_{p1} & a_{mn}b_{p2} & \cdots & a_{mn}b_{pq} \end{bmatrix}. \tag{8.2}$$

The operation $\otimes$ is such that it fulfils the following properties (for $A_1, A_2 \in \mathscr{M}_{m,n}$, $B_1, B_2 \in \mathscr{M}_{p,q}$, $C \in \mathscr{M}_{r,s}$, $D \in \mathscr{M}_{n,w}$, $F \in \mathscr{M}_{q,k}$ and $\alpha \in \mathfrak{F}$):

Bilinearity $A_1 \otimes (B_1 + B_2) = A_1 \otimes B_1 + A_1 \otimes B_2$
$(A_1 + A_2) \otimes B_1 = A_1 \otimes B_1 + A_2 \otimes B_1$

Associativity $(\alpha A_1) \otimes B_1 = A_1 \otimes (\alpha B_1) = \alpha(A_1 \otimes B_1)$
$(A_1 \otimes B_1) \otimes C = A_1 \otimes (B_1 \otimes C)$

Mixed-product $(A_1 \otimes B_1)(D \otimes F) = A_1 D \otimes B_1 F$

Given the above (operational) definition of the tensor product and its properties, it is now simple to see that in general, the tensor product between two vector spaces U and V, denoted by $\mathsf{U} \otimes \mathsf{V}$, has dimension $d_u \cdot d_v$. Moreover, the scalar product $\langle u_1 \otimes v_1, u_2 \otimes v_2 \rangle = \langle u_1, u_2 \rangle \langle v_1, v_2 \rangle$. Note, however, that, differently from the direct sum, it is *not* always the case that given $w \in \mathsf{U} \otimes \mathsf{V}$ there exists $u \in \mathsf{U}$ and $v \in \mathsf{V}$ such that $w \stackrel{!}{=} u \otimes v$. As we will see, this is at the core of the phenomenon of entanglement!

Clearly all the descriptions above can be easily extended to more vector spaces (think inductively).

Bibliography

John S Bell. On the einstein podolsky rosen paradox. *Physics Physique Fizika*, 1(3):195, 1964.

Charles H Bennett and Gilles Brassard. Quantum cryptography: public key distribution and coin tossing. *Theor. Comput. Sci.*, 560(12):7–11, 2014.

Boris S Cirel'son. Quantum generalizations of bell's inequality. *Letters in Mathematical Physics*, 4(2):93–100, 1980.

John F Clauser, Michael A Horne, Abner Shimony, and Richard A Holt. Proposed experiment to test local hidden-variable theories. *Physical review letters*, 23(15):880, 1969.

Valerie Coffman, Joydip Kundu, and William K Wootters. Distributed entanglement. *Physical Review A*, 61(5):052306, 2000.

BIG Bell Test Collaboration et al. Challenging local realism with human choices. *Nature*, 557(7704):212, 2018.

DGBJ Dieks. Communication by epr devices. *Physics Letters A*, 92(6): 271–272, 1982.

Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical review*, 47(10):777, 1935.

Artur K Ekert. Quantum cryptography based on bellâĂŹs theorem. *Physical review letters*, 67(6):661, 1991.

Marissa Giustina, Marijn AM Versteegh, Soeren Wengerowsky, Johannes Handsteiner, Armin Hochrainer, Kevin Phelan, Fabian Steinlechner, Johannes Kofler, Carlos Larsson, et al. Significant-loophole-free test of bellâĂŹs theorem with entangled photons. *Physical review letters*, 115(25): 250401, 2015.

Michael JW Hall. Local deterministic model of singlet state correlations based on relaxing measurement independence. *Physical review letters*, 105 (25):250404, 2010.

Johannes Handsteiner, Andrew S Friedman, Dominik Rauch, Jason Gallicchio, Bo Liu, Hannes Hosp, Johannes Kofler, David Bricher, Matthias Fink, Calvin Leung, et al. Cosmic bell test: measurement settings from milky way stars. *Physical review letters*, 118(6):060401, 2017.

Bas Hensen, Hannes Bernien, Anaïs E Dréau, Andreas Reiserer, Norbert Kalb, Machiel S Blok, Just Ruitenberg, Raymond FL Vermeulen, Raymond N Schouten, Carlos Abellán, et al. Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres. *Nature*, 526(7575): 682, 2015.

Nick Herbert. FlashâĂŤa superluminal communicator based upon a new kind of quantum measurement. *Foundations of Physics*, 12(12):1171–1179, 1982.

Janke Larsson. Loopholes in bell inequality tests of local realism. *Journal of Physics A: Mathematical and Theoretical*, 47(42):424003, 2014.

Lars Lydersen, Carlos Wiechers, Christoffer Wittmann, Dominique Elser, Johannes Skaar, and Vadim Makarov. Hacking commercial quantum cryptography systems by tailored bright illumination. *Nature photonics*, 4 (10):686, 2010.

Dominic Mayers and Andrew Yao. Self testing quantum apparatus. *arXiv preprint quant-ph/0307205*, 2003.

Miguel Navascués, Stefano Pironio, and Antonio Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7):073013, 2008.

Michael A Nielsen and Isaac Chuang. Quantum computation and quantum information, 2002.

Marcin Pawłowski, Tomasz Paterek, Dagomir Kaszlikowski, Valerio Scarani, Andreas Winter, and Marek Żukowski. Information causality as a physical principle. *Nature*, 461(7267):1101, 2009.

Asher Peres. Neumark's theorem and quantum inseparability. *Foundations of Physics*, 20(12):1441–1453, 1990.

Sandu Popescu and Daniel Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, 1994.

Matthew F Pusey, Jonathan Barrett, and Terry Rudolph. On the reality of the quantum state. *Nature Physics*, 8(6):475, 2012.

Lynden K Shalm, Evan Meyer-Scott, Bradley G Christensen, Peter Bierhorst, Michael A Wayne, Martin J Stevens, Thomas Gerrits, Scott Glancy, Deny R Hamel, Michael S Allman, et al. Strong loophole-free test of local realism. *Physical review letters*, 115(25):250402, 2015.

Peter W Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *Proceedings 35th annual symposium on foundations of computer science*, pages 124–134. Ieee, 1994.

Gregor Weihs, Thomas Jennewein, Christoph Simon, Harald Weinfurter, and Anton Zeilinger. Violation of bell's inequality under strict einstein locality conditions. *Physical Review Letters*, 81(23):5039, 1998.

William K Wootters and Wojciech H Zurek. A single quantum cannot be cloned. *Nature*, 299(5886):802, 1982.

Raymond W Yeung. *Information theory and network coding*. Springer Science & Business Media, 2008.