Quantum Information - Primitive notions and quantum correlations

Scarani

October 22, 2009

Contents

Abstract

This Lecture covers basic material, that can be found in many sources. If I'd have to suggest some general bibliography, here are my preferences of the moment: for quantum physics in general, the book by Sakurai [Sakurai 1993]; for a selection of important topics that does not follow the usual treatises, the book by Peres [Peres 1995]; and for quantum information, the books of Nielsen and Chuang [Nielsen and Chuang 2000] or the short treatise by Le Bellac [Le Bellac 2006].

1. Lecture: Quantum Theory

1.1. Axioms, assumption and theorems

1.1.1. Classical versus quantum physics

It is convenient to start this series of lectures by reminding the meaning of some frequently used expressions:

Physical system: the degrees of freedom under study. For instance, the system "point-like particle" is defined by the degrees of freedom $(\vec{x}, \vec{p})$. A system is called *composite* if several degrees of freedom can be identified: for instance, "Earth and Moon".

State at a given time: the collection of all the properties of the system; a *pure state* is a state of maximal knowledge, while a *mixed state* is a state of partial knowledge.

An extremely important distinction has to be made now: the distinction between the *description* (or kinematics) of a physical system and its *evolution* (or dynamics). The distinction is crucial because the *specificity of quantum physics lies in the way physical systems are described, not in the way they evolve!*

Classical physics

In classical physics, the issue of the description of a physical system is not given a strong emphasis because it is, in a sense, trivial. Of course, even in classical physics, before studying "how it evolves", we have to specify "what it is", i.e. to identify the degrees of freedom under study. But once the system is identified, its properties combine according to the usual rules of set theory. Indeed, let P represent properties and s pure states. In classical physics, s is a point in the configuration space and P a subset of the same space. Classical logic translates into the usual rules of set theory:

- The fact that system in state s possesses property P translates as $s \in P$;
- For any two properties P_1 and P_2, one can construct the set $P_1\hat{P}_2$ of the states that possess both properties with certainty;
- If a property P_1 implies another property P_2, then $P_2 \subset P_1$,

and so on. In particular, by construction $s = \bigcap_k P_k$ where P_k are all the properties that the system possesses at that time. As is well-known, only part of this structure can be found in quantum physics: in particular, a pure state is specified by giving the list of all *compatible* physical properties that it possesses with certainty. Another interesting remark can be made about composite systems. Let us consider for instance "Earth and Moon": the two subsystems are obviously interacting, therefore the dynamics of the Earth will be influenced by the Moon. However, at each time one can consider the state of the Earth, i.e.' give its position and its momentum. If the initial state was assumed to be pure, the state of each sub-system at any given time will remain pure (i.e., the values sharply defined). This is not the case in quantum physics: when two sub-systems interact, generically they become "entangled" in a way that only the global state remains pure.

Quantum physics

In all these lectures, I shall focus on *non-relativistic quantum physics* and use the following definition of quantum physics:

(i) Physical systems (degrees of freedom) are described by a Hilbert space;

(ii) The dynamics of a closed system is reversible.

Admittedly, (i) is not a very physical definition: it is rather a description of the formalism. Several attempts have been made to found quantum physics on more "physical" or "axiomatic" grounds, with some success; but I consider none of them conclusive enough to be worth while adopting in a school. We shall come back to this issue in Lecture 5. Requirement (ii) is not specific to quantum physics: in all of modern physics, irreversibility is practical but not fundamental (the result of the system interacting with a large number of degrees of freedom, which one does not manage to keep track of).

1.1.2. Kinematics: Hilbert space

Note: the students of this school are supposed to be already familiar with the basic linear algebra used in quantum physics, as well as with the Dirac notation. Therefore, I skip the mathematical definitions and focus only on the correspondence between physical notions and mathematical objects.

Pure states, Born's rule

Basically the only statement that should be taken as an axiom is the description of states in quantum physics: the space of states is not a set, like the classical configuration space, but a *vector space with scalar product*, defined on the complex field, called *Hilbert space*[1] and written $\mathcal{H}$. In this space, *pure states are described by one-dimensional subspaces.* A one-dimensional subspace is identified by the corresponding projector P; or alternatively, one can choose any vector $|v\rangle$ in the subspace as representative, with the convention however that every vector $c|v\rangle$ differing by a constant represents the same state.

With this construction, as well known, after a measurement a system may be found to possess a property that it did not possess with certainty before the measurement. In other words, given a state $|v_1\rangle$, there is a non-zero probability that a measurement finds the system in a *different* state $|v_2\rangle$. The probability is given by

$$P(v_2|v_1) = \mathrm{Tr}(P_1P_2) = |\langle v_1\,|\,v_2\rangle|^2 \tag{1.1}$$

where we assume, as we shall always do from now onwards, that the vectors are normalized as $\langle v\,|\,v\rangle = 1$. This is called *Born's rule for probabilities.*

What we have discussed in this paragraph is the essence of quantum physics.

[1] Strictly speaking, the Hilbert space is the infinite-dimensional vector space used in quantum physics to describe the point- like particle. But it has been customary to extend the name to all the vector spaces of quantum theory, including the finite-dimensional ones on which we mostly focus here.

It is generally called *superposition principle*: states are treated as vectors, i.e., the vector sum of states defines another state that is not unrelated to its components[2]. All the usual rules of quantum kinematics follow from the vector space assumption. It is useful to sketch this derivation.

Ideal measurements

Consider a measurement, in which different positions of the pointer can be associated with d different states $|\phi_1\rangle, ..., |\phi_d\rangle$. Among the features of an ideal measurement, one tends to request that the device correctly identifies the state. In quantum physics, this cannot be enforced for all states because of the Born's rule; but at least, one can request the following: if the input system is in state $|\phi_j\rangle$, the measurement should produce the outcome associated to that state with certainty (for non-destructive measurements, this implies that the measurement outcome is reproducible). Therefore, given Born's rule, *an ideal measurement is defined by an orthonormal basis, i.e., a set of orthogonal vectors.*

A very compact way of defining an ideal measurement is provided by self-adjoint operators, because any self-adjoint operator A can be diagonalized on an orthonormal basis:

$$A = \sum_k a_k |\phi_k\rangle \langle\phi_k|$$

with a_k real numbers. In this case, one typically considers A to be the "physical quantity" that is measured and the a_k to be the "outcomes" of the measurement. In turn, one is able to compute derived quantities, in particular average values over repeated measurements. Indeed, let $|\psi\rangle$ be the state produced by the source, $a^{(i)}$ be the outcome registered in the i-th measurement and n_k the number of times the outcome a_k was registered: then one has

$$\langle A\rangle_\psi = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} a^{(i)} = \sum_{k=1}^{d} \left(\lim_{N\to\infty} \frac{n_k}{N}\right) a_k = \sum_{k=1}^{d} P(\phi_k|\psi) a_k = \langle\psi| A |\psi\rangle$$

[2] The classical space of states has tyne structure of a set, not a vector space, in spite of being identified with $\mathbb{R}^n$. Indeed, the vector sum is not defined in the configuration space: one can formally describe a point as the vector sum of two others, but there is no relation between the physical states (the "sum" of a car going East at 80km/h and a car going West at 80km/h is a parked car located at the mid-point). Yet another example of a set of numbers, on which operations are not defined, is the set of telephone numbers.

It is however important to keep in mind that, strictly speaking, *only the orthonormal basis defines an ideal measurement, while the labeling of the outcomes of a measurement is arbitrary*: even if real numbers are a very convenient choice in many cases, one can use complex numbers, vectors, colors or any other symbol. In other words, *what is directly measured in an ideal measurement are the probabilities of each outcome*; average values are derived quantities.

Mixed states

The simplest way of introducing mixed states is to think of a source that fluctuates, so that it produces the pure state $|\psi_1\rangle$ with probability p_1, $|\psi_2\rangle$ with probability p_2 and so on. If nothing is known about these fluctuations, the statistics of any ideal measurement will look like

$$P(\phi_j|\rho) = \sum_k p_k P(\phi_j|\psi_k) = \mathrm{Tr}(P_{\phi_j}\rho) \quad \text{with} \quad \rho = \sum_k p_k P_{\psi_k}$$

This is the usual definition of the *density matrix*. It is again important to stress that the notion of mixed state is not proper to quantum physics: in fact, in the formula above, only the $P(\phi_j|\psi_k)$ are typically quantum, but the probabilities p_k are classical (at least formally; their ultimate origin may be entanglement, see below, but this is another matter).

A remark on Gleason's theorem

At this stage, it is worth while mentioning Gleason?s theorem [Gleason 1957]. Suppose that properties P are defined by sub-spaces $\mathcal{E}_P$ of a Hilbert space $\mathcal{H}$; and suppose in addition that orthogonal sub-spaces are associated to distinguishable properties. Let ω be an assignment of probabilities, i.e., $\omega(\mathcal{E}_P) \in [0,1]$, $\omega(\mathcal{H}) = 1$ and $\omega(\mathcal{E}_P \oplus \mathcal{E}_{P'}) = \omega(\mathcal{E}_P) + \omega(\mathcal{E}_{P'})$ if $\mathcal{E}_P \perp \mathcal{E}_{P'}$. Then, if the dimension of the Hilbert space is $d \geq 3$, to each such ω one can associate a non-negative Hermitian operator ρ_ω such that $\omega(\mathcal{E}_P) = \mathrm{Tr}(\rho_\omega \Pi_P)$ with Π_P the projector on the subspace $\mathcal{E}_P$.

This complicated statement basically says that Born's rule can be *derived* if one decides to associate proper- ties to sub-spaces of a Hilbert space and to identify orthogonality with distinguishability. Curiously enough, if one allows for generalized measurements (see below), the theorem becomes valid for $d = 2$ and the proof is considerably simplified [Caves et al., 2004].

1.1.3. Dynamics of closed systems: reversibility

Why reversible evolution must be unitary

The requirement of reversible dynamics for closed systems implies that *the evolution operator must be unitary.* As a simple way of understanding this, let us take an analogy with classical computation, that can be seen as an evolution of an initial string of bits. A computation is obviously reversible if and only if it consists in a permutation: indeed, a permutation is reversible; any other operation, that would map two strings onto the same one, is clearly not reversible. A reversible evolution in a vector space is slightly more general: it is *a change of basis.* A change of basis is obviously reversible; the brute force proof of the converse is hard, but we can provide a proof "by consistency of the theory" that is quite instructive.

We are first going to argue that any operation that describes *reversible* evolution in time must preserve the modulus of the scalar product, $\chi = |\langle\psi_1|\psi_2\rangle|$, between any two vectors. The point is that, as a direct consequence of Born's rule for probabilities, χ has a crucial operational interpretation in the theory: two states can be perfectly discriminated if and only if $\chi = 0$, and are identical if and only if $\chi = 1$; this indicates (and it can be proved rigorously, see Section 3.1) that χ quantifies the "distinguishability" between the two states in the best possible measurement. At this stage, we have only spoken of preparation and measurement and invoked Born's rule; so we have made no assumption on time evolution. So now suppose that an evolution in time can change χ: if χ decreases, then the states become more distinguishable, contradicting the fact that χ quantifies the maximal distinguishability (in other words, in this case the best measurement would consist in waiting for some time before performing the measurement itself). So, for our measurement theory to be meaningful, χ an only increase. But if the evolution is reversible, by reverting it we would have a valid evolution in which χ decreases. The only remaining option is that χ does not change during reversible evolution. Unitarity follows from the fact that only unitary and anti-unitary operations preserve χ, but anti-unitary operations cannot be used for a symmetry with a continuous parameter as translation in time is.

The status of the Schrödinger equation

If the generator of the evolution is supposed to be independent of time, standard group theoretical arguments and correspondence with classical dynamics allow writing

$$U = e^{-iHt/\hbar}$$

where H is the Hamilton operator; from this expression, the corresponding differential equation

$$i\hbar \frac{d}{dt} |\psi\rangle = H |\psi\rangle \quad \text{(Schrödinger equation)}$$

is readily derived. For the case where the dynamics itself varies with time, there is no such derivation: rather, one assumes the same Schrödinger equation to hold with $H = H(t)$. Needless to say, the corresponding dynamics remains unitary.

1.2. Composite systems

In the classical case, as we have argued above, a pure state of a composite system is always a *product state*, i.e., a state of the form $s_A \times s_B$, where s_j is a pure state of system j; in the space of states, this translates by the Cartesian product of the sets of properties of each system.

In quantum theory, product states obviously exist as well, as they correspond to possible physical situations. But the space of states of the whole system cannot be simply $\mathcal{H}_A \times \mathcal{H}_B$, because this is not a vector space. The vector space that contains all product states and their linear combinations is the *tensor product*

$$\mathcal{H}_A \otimes \mathcal{H}_B$$

We start with a rapid survey of the algebraic structure of the tensor product - it really behaves like a product.

1.2.1. Tensor product algebra

The space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ is constructively defined out of its components by requiring that: for any basis $\{|\alpha_j\rangle\}_{j=1....d_A}$ of $\mathcal{H}_A$ and for any basis $\{|\beta_k\rangle\}_{k=1....d_B}$ of $\mathcal{H}_B$, the set of states $\{|\alpha_j\rangle \otimes |\beta_k\rangle\}_{j,k=...}$ forms a basis of $\mathcal{H}$. It follows immediately that the dimension of $\mathcal{H}$ is $d_A d_B$, as it should[3].

[3] Indeed, one can do the counting on product states: two product states of the composite system are perfectly distinguishable if either the two states of the first system are perfectly

The scalar product is defined on product states as

$$(\langle\psi|\otimes\langle\phi|)(|\psi'\rangle\otimes|\phi'\rangle = \langle\psi\,|\,\Psi'\rangle\,\langle\phi\,|\,\phi'\rangle$$

and extended to the whole space by linearity.

The space of linear operators on $\mathcal{H}$ corresponds with the tensor product of the spaces of linear operators on $\mathcal{H}_A$ and on $\mathcal{H}_B$. An operator of the product form, $A\otimes B$, acts on product states as

$$(A\otimes B)(|\psi\rangle\otimes|\phi\rangle) = A\,|\psi\rangle\otimes B\,|\phi\rangle$$

by linearity, this defines uniquely the action of the most general operator on the most general state.

1.2.2. Entanglement

Any linear combination of product states defines a possible pure state of the composite system. However, a state like

$$|\psi(\theta)\rangle = \cos\theta Ket\alpha_1\otimes|\beta_1\rangle + \sin\theta Ket\alpha_2\otimes|\beta_2\rangle \tag{1.2}$$

cannot be written as a product state, as it is easily checked by writing down the most general product state

$$\left(\sum_j a_j\,|\alpha_j\rangle\right)\otimes\left(\sum_k b_k\,|\beta_k\rangle\right)$$

and equating the coefficients. In fact, it?s easy to convince oneself that product states form a set of measure zero (according to any reasonable measure) in the whole set of states. A *pure state, that cannot be written as a product state, is called "entangled"*.

The most astonishing feature of entanglement is that we have a pure state of the composite system that does not arise from pure states of the components: in other words, the properties of the whole are sharply defined, while the properties of the sub-systems are not.

distinguishable, or the two states of the second system are perfectly distinguishable, or both.

Case study: the singlet state

We are going to study a specific example that plays an important role in what follows. Like most of the examples in this series of lectures, this one involves *qubits*, i.e., two-level systems. The algebra of two-level systems is supposed to be known from basic quantum physics; for convenience, it is reminded as an Appendix to this lecture (section 1.3).

Consider two qubits. The state

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|0\rangle \otimes |1\rangle - |1\rangle \otimes |0\rangle) \tag{1.3}$$

is called *singlet* because of its particular status in the theory of addition of angular momenta. The projector on this state reads

$$P_{\Psi^-} = \frac{1}{4}(\mathbb{1} \otimes \mathbb{1} - \sigma_x \otimes \sigma_x - \sigma_y \otimes \sigma_y - \sigma_z \otimes \sigma_z)) \tag{1.4}$$

It appears clearly from the projector[4] that this state is invariant by bilateral rotation:

$$u \otimes u\,|\Psi^-\rangle = |\Psi^-\rangle$$

(possibly up to a global phase). In particular, consider a measurement in which the first qubit is measured along direction $\vec{a}$ and the second along direction $\vec{b}$. The statistics of the outcomes $r_A, r_B \in \{-1, +1\}$ are given by

$$P_{\vec{a},\vec{b}}(r_A, r_B) = \left|\left(\langle r_A\vec{a}| \otimes \left\langle r_B\vec{b}\right|\right)|\Psi^-\rangle\right|^2$$

The calculation leads to

$$P_{\vec{a},\vec{b}}(++) = P_{\vec{a},\vec{b}}(--) = \frac{1}{4}(1 - \vec{a} \otimes \vec{b}) \tag{1.5}$$

$$P_{\vec{a},\vec{b}}(+-) = P_{\vec{a},\vec{b}}(-+) = \frac{1}{4}(1 + \vec{a} \otimes \vec{b}) \tag{1.6}$$

[4] For those who have never done it, it is useful to check that indeed the state has the same form in any basis. To do so, write a formal singlet in a different basis

$$\frac{1}{\sqrt{2}}(|+\hat{n}\rangle \otimes |-\hat{n}\rangle + |-\hat{n}\rangle \otimes |+\hat{n}\rangle)$$

open the expression up using (1.22) and verify that this state is exactly the same state as (1.3).

Much information can be extracted from those statistics, we shall come back to them in Lecture 4. For the moment, let us just stress the following features:

- On the one hand,

  $$P_{\vec{a},\vec{b}}(r_A = +) = P_{\vec{a},\vec{b}}(r_A = -) = \frac{1}{2}$$

  and

  $$P_{\vec{a},\vec{b}}(r_B = +) = P_{\vec{a},\vec{b}}(r_B = +) = \frac{1}{2}$$

  the outcomes of the measurements of each qubit appear completely random, independent of the directions of the measurements. This implies the operational meaning of the fact that the state of each qubit cannot be pure (and here it is actually maximally mixed, see below).

- On the other hand, there are very sharp correlations: in particular, whenever $\vec{a} = \vec{b}$, the two outcomes are rigorously opposite.

This can be seen as defining the relation "being opposite" for two arrows in itself(the same arrow), without specifying in which direction each arrow points. While this possibility is logically compelling, it is worth while reminding that this is impossible in our everyday life.

A warning: we stressed that the singlet has a special role to play in the addition of two spins 1/2: indeed, it defines a one-dimensional subspace of total spin 0, as opposite to the three states that form the spin-1 triplet. However, in quantum information theory this feature is *not* crucial. In other words, in quantum information the state $|\Psi^-\rangle$ is as good as any other state

$$u_1 \otimes u_2 \,|\Psi^-\rangle = \frac{1}{\sqrt{2}}\big(|{+}\hat{m}\rangle \otimes |{+}\hat{n}\rangle + |{-}\hat{m}\rangle \otimes |{-}\hat{n}\rangle\big)$$

that can obtained from it with local unitaries. In particular, it is customary to define the states

$$|\Psi^+\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle \otimes |1\rangle + |1\rangle \otimes |0\rangle\big) \tag{1.7}$$

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle\big) \tag{1.8}$$

$$|\Phi^-\rangle = \frac{1}{\sqrt{2}}(|0\rangle \otimes |0\rangle - |1\rangle \otimes |1\rangle) \tag{1.9}$$

that are orthogonal to $|\Psi^-\rangle$ and with which they form the so-called *Bell basis*.

Entanglement for mixed states

It is in principle not difficult to decide whether a pure state is entangled or not: just check if it is a product state; if not, then it is entangled. For mixed states, the definition is more subtle, because a mixed state may exhibit classical correlations. As an example, consider the mixture "half of the times I prepare $|0\rangle \otimes |0\rangle = |00\rangle$, and half of the times I prepare $|1\rangle \otimes |1\rangle = |11\rangle$", that is,

$$\rho = \frac{1}{2}|00\rangle\langle 00| + \frac{1}{2}|11\rangle\langle 11|$$

This state exhibits correlations: whenever both particles are measured in the basis defined by $|0\rangle$ and $|1\rangle$, the outcomes of both measurements are the same. However, by the very way the state was prepared, it is clear that no entanglement is involved.

We shall then call *separable* a mixed state for which a decomposition as a convex sum of product state exists:

$$\rho = \sum_k p_k(|\psi_k\rangle\langle\psi_k|)_A \otimes (|\phi_k\rangle\langle\phi_k|)_B \tag{1.10}$$

It is enough that a single mixture[5] of product states exists, for the state to be separable. A mixed state is called *entangled* if it is not separable, i.e. if some entangled state must be used in order to prepare it [Werner 1989]. Apart from some special cases, to date no general criterion is known to decide whether a given mixed state is separable or entangled.

1.2.3. Partial states, no-signaling and purification

We have just seen that, in the presence of entanglement, one cannot assign a separate pure state to each of the sub-systems. Still, suppose two entangled particles are sent apart from one another to two physicists, Alice and Bob.

[5] Recall that any mixed state that is not pure can be decomposed in an infinite number of ways as a mixture of pure states (although generally there is only one mixture of pure *mutually orthogonal states*, because of hermiticity).

Alice on her location holds particle A and can make measurements on it, without possibly even knowing that a guy named Bob holds a particle that is correlated with hers. Quantum physics should provide Alice with rules to compute the probabilities for the outcome she observes: there must be a "state" (positive, unit trace hermitian operator) that describes the information available on Alice's side. Such a state is called "partial state" or "local state". The local state cannot be pure if the state of the global system is entangled.

Suppose that the full state is ρ_{AB}. The partial state ρ_A is defined as follows: for any observable $\mathcal{A}$ on Alice's particle, it must hold

$$\begin{aligned} \mathrm{Tr}_A(\rho_A\mathcal{A}) &= \mathrm{Tr}_{AB}\left(\rho_{AB}(\mathcal{A}\otimes \mathbb{1}_B)\right) \\ &= \sum_{j=1}^{d_A}\sum_{k=1}^{d_B} \langle a_j, b_k|\,\rho_{AB}(\mathcal{A}\otimes \mathbb{1}_B)\,|a_j, b_k\rangle \\ &= \sum_{j=1}^{d_A} \langle a_j|\,(\mathrm{Tr}_B(\rho_{AB}))\,\mathcal{A}\,|a_j\rangle = \mathrm{Tr}_A\left(\mathrm{Tr}_B(\rho_{AB})\mathcal{A}\right) \end{aligned} \tag{1.11}$$

so by identification

$$\rho_A = \mathrm{Tr}_B(\rho_{AB}) = \sum_{k=1}^{d_B} \langle b_k|\,\rho_{AB}\,|b_k\rangle \tag{1.12}$$

This is the general definition of the partial state: the partial state on A is obtained by *partial trace* over the other system B. The result (1.12) is not ambiguous since the trace is a unitary invariant, that is, gives the same result for any choice of basis. This implies that *whatever Bob does, the partial state of Alice will remain unchanged.* This fact has an important consequence, namely that Bob cannot use entanglement to send any message to Alice. This is known as the principle of *no-signaling through entanglement*[6].

When it comes to computing partial states, one can always come back to the general definition (1.12), but there are often more direct ways. For instance,

[6] This principle is sometimes stated by saying that entanglement does not allow to signal "faster-than-light". Indeed, if entanglement would allow signaling, this signal would travel faster than light; and this is why we feel relieved when we notice that quantum physics does not allow such a signaling. However, entanglement does not allow to signal *tout court(with no qualification)*, be it faster or slower than light.

no-signaling itself can be used: to obtain (say) ρ_A, one may consider that Bob performs a given measurement and compute the mixture that would be correspondingly prepared on Alice's side. This mixture must be ρ_A itself, since the partial state is the same for any of Bob's actions.

The notion of "purification"

The notion of *purification* is, in a sense the reverse, of the notion of partial trace. It amounts to the following: any mixed state can be seen as the partial state of a pure state in a larger Hilbert space. To see that this is the case, note that any state ρ can be diagonalized, i.e., there exist a set of orthogonal vectors $\{|\varphi_k\rangle\}_{k=1\ldots r}$ (r being the rank of ρ) such that

$$\rho = \sum_{k=1}^{r} p_k \, |\varphi_k\rangle \, \langle\varphi_k|$$

Then one can always take r orthogonal vectors $\{|e_k\rangle\}_{k=1\ldots r}$ of an ancilla and construct the pure state

$$|\Psi\rangle = \sum_{k=1}^{r} \sqrt{p_k} \, |\varphi_k\rangle \otimes |e_k\rangle$$

of which ρ is the partial state for the first sub-system.

This construction shows that a purification always exists; but because the decomposition of a mixed state onto pure states is not unique, the purification is not unique as well. In particular, if $|\Psi\rangle$ is a purification of ρ and U is a unitary on the ancilla, then $\mathbb{1} \otimes U \, |\Psi\rangle$ is also a purification of ρ, because it just amounts at choosing another set $\{|e'_k\rangle\}_{k=1\ldots r}$ of orthogonal vectors. Remarkably though, this is the only freedom: it can indeed be proved that all purification are equivalent up to local operations on the ancilla. In this sense, the purification can be said to be unique.

1.2.4. Measurement and evolution revisited

To conclude this first Lecture, we have to mention one of the most important extensions of the notion of tensor product, namely the definitions of *generalized measurements and evolution.* I shall not go into details of the formalism; whenever these notions appear later in these lectures, it will always be in a rather natural way.

Evolution: CP maps, decoherence

We have stressed above that the evolution of a closed system is reversible, hence unitary. In general, however, it is impossible to have a perfectly isolated system: even if some degrees of freedom ("the system") are carefully selected and prepared in a given state, the evolution may imply some interaction with other degrees of freedom ("the environment"). In principle, it is obvious what has to be done to describe such a situation: consider both the system and the environment, study the unitary evolution, then trace the environment out and study the resulting state of the system. By doing this study for any given time t, one defines an effective map T_t such that $\rho_S(t) = T_t[\rho_S(0)]$. It turns out that the maps defined this way coincide with the family of *trace-preserving, completely positive (CP) maps.* Let us comment on each of these important properties:

- A map is positive if it transforms positive operators into positive operators; in our case, this means that the density matrix will never develop any negative eigenvalue;
- The trace-preserving property is self-explained: $\text{Tr}[\rho_S(t)] = 1$ for all t. Together with the previous, it implies that a density matrix remains a density matrix - an obvious necessity.
- Not all positive maps, however, define possible evolutions of a subsystem. For instance, time-reversal is a positive map, but if one applies time-reversal only to a sub-system, it seems clear that one may get into trouble. A map T is called "completely positive" if $T \otimes \mathbb{1}$ is also a positive map for any possible enlarged system. Interestingly, non-CP maps play an important role in quantum information: since the non-positivity can appear only if the system is entangled with the environment, these maps act as entanglement witnesses (see series of lectures on entanglement theory).

Trace-preserving CP maps are the most general evolution that a quantum system can undergo. Whenever the map on the system is not unitary, the state of the system becomes mixed by entanglement with the environment: this is called *decoherence.* An obvious property of these maps is that states can only become "less distinguishable". A last remark: it would of course be desirable to derive a differential equation for the state of the system, whose solution implements the CP-map, without having to study the environment fully. In full generality, this task has been elusive; a general form has been derived by Lindblad for the case where the environment has no memory [Lindblad 1976]. For a detailed introduction to the theory of open quantum systems, we refer to the book by Breuer and Petruccione [Breuer and Petruccione 2002].

Generalized measurements

For measurements, a similar discussion can be made. The most general measurement on a quantum system consists in appending other degrees of freedom, then performing a measurement on the enlarged system. These additional degrees of freedom are called *ancillae* (latin for "servant maids") rather than environment, but play exactly the same role: the final measurement may project on states, in which the system and the ancillae are entangled. Note that the ancillae start in a state that is independent of the state of the system: they are part of the measuring apparatus, and the whole measurement can truly be said to be a measurement on the system alone.

The effective result on the system is captured by a family of positive operators: for all possible generalized measurement with D outcomes, there exist a family $\{A_k\}_{k=1,\dots,D}$ of positive operators, such that $\sum_k A_k^\dagger A_k = \mathbb{1}$. If the system has been prepared in the state ρ, the probability of obtaining outcome k is given by $p_k = \mathrm{Tr}(A_k \rho A_k^\dagger)$ and the state is subsequently prepared as $\rho_k = \frac{1}{p_k} A_k \rho A_k^\dagger$. Note that D must be at most the dimension of the total Hilbert space "system+ancillae", but can of course be much larger than the dimension of the Hilbert space of the system alone.

For some reason, the name that has been retained for such generalized measurements is *positive-operator-valued measurements*, or *POVMs*. Some specific examples will be presented in the coming lectures.

1.3. Appendix: one-qubit algebra

Since most of the examples in these lectures will be done on spin 1/2 systems (*qubits*), I remind here for convenience some basic elements of the Hilbert space $\mathcal{H} = \mathbb{C}^2$ and of the linear operators on that space. We start by recalling the Pauli matrices:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad , \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.13}$$

The computational basis $\{|0\rangle, |1\rangle\}$ is universally assumed to be the eigenbasis of σ_z so that:

$$\sigma_z |0\rangle = |0\rangle \quad , \quad \sigma_z |1\rangle = -|1\rangle \tag{1.14}$$

$$\sigma_x |0\rangle = |1\rangle \quad , \quad \sigma_x |1\rangle = -|0\rangle \tag{1.15}$$

$$\sigma_y |0\rangle = i|1\rangle \quad , \quad \sigma_y |1\rangle = -i|0\rangle \tag{1.16}$$

It is useful sometimes to write the Pauli matrices as

$$\sigma_x = |0\rangle\langle 1| + |1\rangle\langle 0| \ , \ \sigma_y = -i|0\rangle\langle 1| + i|1\rangle\langle 0| \ , \ \sigma_x = |0\rangle\langle 0| - |1\rangle\langle 1| \tag{1.17}$$

One has $\mathrm{Tr}(\sigma_k) = 0$ and $\sigma_k^2 = \mathbb{1}$ for $k = x, y, z$; moreover

$$\sigma_x\sigma_y = -\sigma_y\sigma_x = i\sigma_z \quad + \text{ cyclic permutations.} \tag{1.18}$$

The generic pure state of a qubit is written $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$. The associated projector is therefore

$$|\psi\rangle\langle\psi| = |\alpha|^2 |0\rangle\langle 0| + |\beta|^2 |1\rangle\langle 1|_\alpha \beta^* |0\rangle\langle 1| + \alpha^*\beta |1\rangle\langle 0| \tag{1.19}$$

$$= \frac{1}{2}\left(\mathbb{1} + (|\alpha|^2 - |\beta|^2)\sigma_z + 2\mathrm{Re}(\alpha\beta^*)\sigma_x + 2\mathrm{Im}(\alpha\beta^*)\sigma_y\right) \tag{1.20}$$

It follows immediately from $\sigma_k^2 = \mathbb{1}$ that

$$|\psi\rangle\langle\psi| = \frac{1}{2}\left(\mathbb{1} + \sum_k \langle\sigma_k\rangle_\psi \sigma_k\right)$$

One writes

$$|\psi\rangle\langle\psi| = \frac{1}{2}\left(\mathbb{1} + \hat{n}\cdot\vec{\sigma}\right) \ , \ \hat{n} = \begin{pmatrix} \langle\sigma_x\rangle_\psi \\ \langle\sigma_y\rangle_\psi \\ \langle\sigma_z\rangle_\psi \end{pmatrix} = \begin{pmatrix} 2\mathrm{Re}(\alpha\beta^*) \\ 2\mathrm{Im}(\alpha\beta^*) \\ |\alpha|^2 - |\beta|^2 \end{pmatrix} \tag{1.21}$$

The vector $\hat{n}$ is called Bloch vector, it corresponds to the expectation value of the "magnetic moment" $\vec{\sigma}$ in the given state. For pure states (the case we are considering here), its norm is one. It is actually well-known that such vectors cover the unit sphere (called the *Bloch sphere*, or the Poincaré sphere if the two-level system is the polarization of light). In fact, there is a one-to-one correspondence between unit vectors and pure states of a two-level system given by the following parametrization in spherical coordinates:

$$|\psi\rangle \equiv |{+\hat{n}}\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle \leftrightarrow \hat{n} \equiv \hat{n}(\theta,\varphi) = \begin{pmatrix}\sin\theta\cos\varphi\\ \sin\theta\sin\varphi\\ \cos\theta\end{pmatrix} \tag{1.22}$$

In turn, $|{+\hat{n}}\rangle$ is the eigenstate of $\hat{n}\cdot\vec{\sigma}$ for the eigenvalue +1; or alternatively,

$$\hat{n}\cdot\vec{\sigma} \equiv \sigma_n = |{+\hat{n}}\rangle\langle{+\hat{n}}| - |{-\hat{n}}\rangle\langle{-\hat{n}}| \tag{1.23}$$

One has $\mathrm{Tr}(\sigma_n) = 0$ and $\sigma_n^2 = \mathbb{1}$. For any basis $\{|{+\hat{n}}\rangle, |{-\hat{n}}\rangle\}$, one has the closure (completeness) relation

$$|{+\hat{n}}\rangle\langle{+\hat{n}}| + |{-\hat{n}}\rangle\langle{-\hat{n}}| = \mathbb{1}$$

The eigenstates of σ_x and σ_y are frequently used. They read, up to a global phase:

$$|{\pm x}\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle) \quad , \quad |{\pm y}\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm i\,|1\rangle) \tag{1.24}$$

Let's move to the study of *mixed states*. Given that any projector can be written as (1.21), obviously any mixed state takes exactly the same form. In fact, consider just the mixture of two pure states:

$$\rho = p_1|\psi_1\rangle\langle\psi_1| + p_2|\psi_2\rangle\langle\psi_2| = \frac{1}{2}\left(\mathbb{1} + (p_1\hat{n}_1 + p_2\hat{n}_2)\cdot\vec{\sigma}\right)$$

It is easy to verify that the resulting Bloch vector $\vec{m} = p_1\hat{n}_1 + p_2\hat{n}_2$ lies *inside* the unit sphere. In fact, the points in the volume of the Bloch sphere are in a one-to-one correspondence with all possible states of a single qubit, be they pure (in which case their Bloch vector lies on the surface) or mixed.

We can summarize all that should be known on the states of a single qubit as follows: generically, a state of a single qubit reads

$$\rho = \frac{1}{2}\left(\mathbb{1} + \vec{m}\cdot\vec{\sigma}\right) = \frac{1}{2}\left(\mathbb{1} + |\vec{m}\sigma_m\right) \quad \text{with} \quad \vec{m} = \begin{pmatrix}\mathrm{Tr}(\sigma_x\rho)\\ \mathrm{Tr}(\sigma_y\rho)\\ \mathrm{Tr}(\sigma_z\rho)\end{pmatrix} \tag{1.25}$$

The norm of the Bloch vector is $|\vec{m}| \leq 1$, with equality if and only if the state is pure. Since ρ is hermitian, there is one and only one decomposition as the sum of two orthogonal projectors. Clearly, the eigenstates of this decomposition are the eigenstates $|+\hat{m}\rangle$ and $|-\hat{m}\rangle$ of $\vec{m}\cdot\vec{\sigma}$, where $\hat{m} = \vec{m}/|\vec{m}|$. The orthogonal decomposition reads

$$\rho = \left(\frac{1+|\vec{m}|}{2}\right)|+\hat{m}\rangle\langle+\hat{m}| + \left(\frac{1-|\vec{m}|}{2}\right)|-\hat{m}\rangle\langle-\hat{m}| \qquad (1.26)$$

Recall that any density matrix ρ that is not a projector admits an infinity of decompositions as sum of projectors. All these decompositions are equivalent, because the density matrix carries all the information on the state, that is, on the actual properties of the system.

Finally, a useful result: the probability of finding $|+\hat{n}\rangle$ given the state $\rho = \frac{1}{2}(\mathbb{1} + \vec{m}\cdot\vec{\sigma})$ is

$$\text{Prob}(+\hat{n}| + \vec{m}) = \frac{1}{4}\text{Tr}[(\mathbb{1} + \hat{n}\cdot\vec{\sigma})(\mathbb{1} + \vec{m}\cdot\vec{\sigma})] = \frac{1+\hat{n}\cdot\vec{m}}{2} \qquad (1.27)$$

1.4. Tutorials

1.4.1. Problems

Exercise 1.1

Which of the following states are entangled?

1. $|\Psi_1\rangle = \cos\theta\,|0\rangle|0\rangle + \sin\theta\,|1\rangle|1\rangle$
2. $|\Psi_2\rangle = \cos\theta\,|0\rangle|0\rangle + \sin\theta\,|1\rangle|0\rangle$
3. $|\Psi_3\rangle = \frac{1}{2}(|0\rangle|0\rangle + |0\rangle|1\rangle - |1\rangle|0\rangle - |1\rangle|1\rangle)$
4. $|\Psi_4\rangle = \frac{1}{2}(|0\rangle|0\rangle + |0\rangle|1\rangle + |1\rangle|0\rangle - |1\rangle|1\rangle)$

For those that are not entangled, give the decomposition as a product state. For simplicity of notation, we write $|\psi_1\rangle|\psi_2\rangle$ instead of $|\psi_1\rangle\otimes|\psi_2\rangle$.

Exercise 1.2

Compute the partial states ρ_A and ρ_B for the following states of two qubits:

1. The pure state $|\Psi\rangle = \sqrt{\frac{2}{3}}|0\rangle|+\rangle + \sqrt{\frac{1}{3}}|+\rangle|-\rangle$; where $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$

2. The mixed state $W(\lambda) = \lambda |\Psi^-\rangle \langle\Psi^-| + (1-\lambda)\frac{\mathbb{1}}{4}$, call Werner state [R.F. Werner, Phys. Rev. A 40, 4277 (1989)].

Verify in both cases that ρ_A and ρ_B are mixed by computing the norm of their Bloch vectors.

Exercise 1.3

We consider the following decoherent channel [Scarani et al. 2002]. A qubit, initially prepared in a state ρ, undergoes sequential "collisions" with qubits coming from a reservoir. All the qubits of the reservoir are supposed to be in state $\xi = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|$. Each collision implements the evolution

$$U : \begin{cases} |0\rangle|0\rangle \Rightarrow |0\rangle|0\rangle \\ |0\rangle|1\rangle \Rightarrow \cos\phi|0\rangle|1\rangle + i\sin\phi|1\rangle|0\rangle \\ |1\rangle|0\rangle \Rightarrow \cos\phi|1\rangle|0\rangle + i\sin\phi|0\rangle|1\rangle \\ |1\rangle|1\rangle \Rightarrow |1\rangle|1\rangle \end{cases} \tag{1.28}$$

with $\sin\phi \neq 0$. We assume that each qubit of the reservoir interacts only once with the system qubit. Therefore, the state of the system after collision with $n+1$ qubits of the bath is defined recursively as

$$\rho^{(n+1)} \equiv T_\xi^{n+1}[\rho] = \mathrm{Tr}_B\left(U\rho^{(n)} \otimes \xi U^\dagger\right) \tag{1.29}$$

1. Let

$$\rho^{(n)} = d^{(n)}|0\rangle\langle 0| + (1-d^{(n)})|1\rangle\langle 1| + k^{(n)}|0\rangle\langle 1| + k^{(n)*}|1\rangle\langle 0|$$

   Prove that the CP-map (1.29) induces the recursive relations

$$d^{(n+1)} = c^2 d^{(n)} + s^2 p \quad , \quad k^{(n+1)} = ck^{(n)} \tag{1.30}$$

   with $c = \cos\phi$ and $s = \sin\phi$

2. By iteration, provide $d^{(n+1)}$ and $k^{(n+1)}$ as a function of the parameters of the initial state $d^{(0)}$ and $k^{(0)}$. Conclude that $T_\xi^n[\rho] \to \xi$ when $n \to \infty$, whatever the initial state ρ (pure or mixed).

3. We have just studied an example of "thermalization": a system, put in contact with a large reservoir, ultimately assumes the same state as the particles in the reservoir. Naively, one would have described this process as $\rho \otimes \xi^{\otimes N} \to \xi^{\otimes N+1}$ for all ρ. Why is such a process not allowed by quantum physics?

4. The condition $\sin\phi \neq 0$ is necessary to have a non-trivial evolution during each collision; however, to have a meaningful model of thermalization one has to enforce $\cos\phi >> |\sin\phi|$. What is the meaning of this condition? Hint: as a counter-example, consider the extreme case $\sin\phi = 1$; what is then U? What does the process look like in this case?

1.4.2. Solutions

Exercise 1.1

$|\Psi_1\rangle$ is entangled. $|\Psi_2\rangle = (\cos\theta\,|0\rangle + \sin\theta\,|1\rangle)\,|0\rangle$ is not entangled. $|\Psi_3\rangle = \frac{1}{2}(|0\rangle - |1\rangle)(|0\rangle + |1\rangle) = |-\rangle\,|+\rangle$ is not entangled. $|\Psi_4\rangle$ is entangled. this can be verified by direct calculation, or also by noticing that $|\Psi_4\rangle = \frac{1}{\sqrt{2}}(|0\rangle\,|+\rangle + |1\rangle\,|-\rangle)$; by just relabeling the basis of the second system, one sees that this state has the same form as $|\Psi_1\rangle$ with $\cos\phi = \sin\phi = \frac{1}{\sqrt{2}}$.

Exercise 1.2

1. For the pure state under study,

$$\rho_A = \frac{2}{3}\,|0\rangle\,\langle 0| + \frac{1}{3}\,|+\rangle\,\langle +| = \frac{1}{2}\left(\mathbb{1} + \frac{2}{3}\sigma_z + \frac{1}{3}\sigma_x\right)$$

In order to compute ρ_B, here is a possibility (maybe not the fastest one): first, rewrite the state as

$$|\Psi\rangle = |0\rangle\left(\sqrt{\frac{2}{3}}\,|+\rangle + \sqrt{\frac{1}{6}}\,|-\rangle\right) + \sqrt{\frac{1}{6}}\,|1\rangle\,|-\rangle$$

Then

$$\begin{aligned}\rho_B &= \left(\sqrt{\frac{2}{3}}\,|+\rangle + \sqrt{\frac{1}{6}}\,|-\rangle\right)\left(\sqrt{\frac{2}{3}}\,\langle +| + \sqrt{\frac{1}{6}}Bra-\right) + \frac{1}{6}\,|-\rangle\,\langle -| \\ &= \frac{2}{3}\,|+\rangle\,\langle +| + \frac{1}{3}\,|-\rangle\,\langle -| + \frac{1}{3}\,|+\rangle\,\langle -| + \frac{1}{3}\,|-\rangle\,\langle +| = \frac{2}{3}\,|0\rangle\,\langle 0| + \frac{1}{3}\,|+\rangle\,\langle +| = \rho_A\end{aligned}$$

since $|0\rangle = \frac{1}{\sqrt{2}}(|+\rangle + |-\rangle)$. Note how, in this last calculation, the normalization is taken care of automatically.

The Bloch vector of $\rho_A = \rho_B$ has norm $|\vec{m}| = \frac{\sqrt{5}}{3} < 1$, therefore the states are mixed.

2. For the Werner state: $\rho_A = \rho_B = \frac{\mathbb{1}}{2}$. These states are maximally mixed, indeed $|\vec{m}| = 0$.

Exercise 1.3

1. The first point is a matter of patience in writing down explicitly $U\rho^{(n)} \otimes \xi U^\dagger$, then noticing that

$$\mathrm{Tr}(|0\rangle\langle 0|) = \mathrm{Tr}(|1\rangle\langle 1|) = \mathrm{Tr}(|0\rangle\langle 1|) = \mathrm{Tr}(|1\rangle\langle 0|) = 0$$

2. For the off-diagonal term, the recursion is obviously

$$k^{(n+1)} = c^{n+1} k^{(0)}$$

For the diagonal term, one has

$$\begin{aligned} d^{(n+1)} &= c^2\big[c^2 d^{(n-1)} + s^2 p\big] + s^2 p = c^4 d^{(n-1)} + s^2(1+c^2)p = \ldots\ldots \\ &= c^{2(n+1)} d^{(0)} + s^2 \sum_{k=0}^{n} c^{2k} p = c^{2(n+1)} d^{(0)} + \big[1 - c^{2(n+1)}\big]p \end{aligned}$$

because

$$\sum_{k=0}^{n} c^{2k} = \frac{1 - c^{2(n+1)}}{1 - c^2} = \frac{1 - c^{2(n+1)}}{s^2}$$

Therefore

$$d^{(n+1)} \to p \quad \text{and} \quad k^{(n+1)} \to 0 \quad \text{for } n \to \infty$$

3. The evolution $\rho \otimes \xi^{\otimes N} \to \xi^{\otimes N+1}$ is not unitary, since two initially different states would end up being the same.

4. For $\sin\phi = 1$, U is the swap operation. In this case, the "thermalization" would consist in dumping the initial system in the reservoir and replacing it with one of the qubits of the reservoir. Such a process would introduce a very large fluctuation in the reservoir. By setting $\cos\phi \approx 1$, on the contrary, one has $\mathrm{Tr}(\rho_j A) \approx \mathrm{Tr}(\xi A)$ for any quibit j, for any single-particle physical quantity A. In other words, the system *appears* to be completely thermalized and one has to measure some multi-particle physical quantities to see some differences. This view is perfectly consistent with the idea that irreversibility is only apparent.

2. Lecture: Primitives of quantum information (I)

2.1. A tentative list of primitives

The main tasks of quantum information, at the present stage of its development, are quantum computing and quantum cryptography. These tasks are complex: they rely on simpler notions, most of which are of interest in themselves. These notions that subtend the whole field are those that I call “primitives”. Here is my tentative list of primitives, listed in chronological order of their appearance in the development of quantum physics:

1. Violation of Bell’s inequalities
2. Quantum cloning
3. State discrimination
4. Quantum coding
5. Teleportation
6. Error correction
7. Entanglement distillation

The common feature of all these primitives is that they have been studied in great detail. This does not mean that there are no open issues left; however, with a few remarkable exceptions, those are generally difficult points of rather technical nature. This is why you may not hear many talks dedicated to these topics in research conferences - but the notions are there and will appear over and over again, as something anyone should know. This is why it is important to review those basic notions in a school like this one.

In my series of lectures, I shall deal with quantum cloning, teleportation and entanglement distillation (this Lecture), state discrimination, quantum coding (Lecture 3) and the violation of Bell’s inequalities in greater detail (Lectures 4-6). Error correction is presented in this school by other lecturers.

2.2. Quantum cloning

The first primitive that should be considered is quantum cloning. The famous no-go theorem was formulated in 1982-83 [Wootters and Żurek 1982, Dieks 1982, Milonni and Hardies 1982, Mandel 1983]; much later, in 1996, came the idea of studying optimal cloning [Bužek and Hillery 1996]. Since then, the subject has been the object of rather thorough investigations; two very comprehensive review articles are available [Scarani et al. 2005, Cerf and Fiurášek 2006].

2.2.1. The no-go theorem

It is well-known that one cannot measure the state $|\psi\rangle$ of a single quantum system: the result of any single measurement of an observable A is one of its eigenstates, unrelated to the input state $|\psi\rangle$. To reconstruct $|\psi\rangle$ (or more generally ρ) one has to measure the average values of several observables; this implies a statistic over a large number of identically prepared systems (see Lecture 3).

One can imagine to circumvent the problem in the following way: take the system in the unknown state $|\psi\rangle$ and let it interact with N other systems previously prepared in a blank reference state $|R\rangle$, in order to obtain $N+1$ copies of the initial state:

$$|\psi\rangle \otimes |R\rangle \otimes |R\rangle \cdots \otimes |R\rangle \overset{?}{\rightarrow} |\psi\rangle \otimes |\psi\rangle \otimes |\psi\rangle \cdots \otimes |\psi\rangle \qquad (2.1)$$

Such a procedure would allow one to determine the quantum state of a single system, without even measuring it because one could measure the N new copies and let the original untouched. The no-cloning theorem of quantum information formalizes the suspicion that such a procedure is impossible:

No-cloning theorem: no quantum operation exists that can duplicate perfectly an unknown quantum state.

The theorem can be proved by considering the 1 → 2 cloning. Suppose first that perfect cloning is possible without any ancilla: this means that there exist a unitary operation such that

$$|\text{in}(\psi)\rangle \equiv |\psi\rangle \otimes |R\rangle \overset{?}{\rightarrow} |\psi\rangle \otimes |\psi\rangle \equiv |\text{out}(\psi)\rangle \qquad (2.2)$$

But such an operation cannot be unitary, because it does not preserve the scalar product:

$$\langle \text{in}(\psi) \,|\, \text{in}(\psi) \rangle = \langle \psi \,|\, \phi \rangle \neq \langle \text{out}(\psi) \,|\, \text{out}(\psi) \rangle = \langle \psi \,|\, \phi \rangle^2$$

Now we have to prove that perfect cloning is impossible also for CP maps, the most general evolution. So let's add an ancilla (the "machine") and suppose that

$$|\psi\rangle \otimes |R\rangle \otimes |M\rangle \overset{?}{\rightarrow} |\psi\rangle \otimes |\psi\rangle \otimes |M(\psi)\rangle \tag{2.3}$$

is unitary. The same type of proof as before can be done as in Sect. 9-4 of Peres' book [Peres 1995]; here we give a different one, closer to the Wootters-Zurek proof [Wootters and Żurek 1982]. We suppose that (2.3) holds for two orthogonal states, labelled $|0\rangle$ and $|1\rangle$:

$$|0\rangle \otimes |R\rangle \otimes |M\rangle \rightarrow |0\rangle \otimes |0\rangle \otimes |M(0)\rangle$$
$$|1\rangle \otimes |R\rangle \otimes |M\rangle \rightarrow |1\rangle \otimes |1\rangle \otimes |M(1)\rangle$$

Because of linearity (we omit tensor products) then:

$$(|0\rangle + |1\rangle)\,|R\rangle\,|M\rangle \rightarrow |00\rangle\,|M(0)\rangle + |11\rangle\,|M(1)\rangle$$

that cannot be equal to

$$(|0\rangle + |1\rangle)(|0\rangle + |1\rangle)\,|M(0+1)\rangle = (|00\rangle + |10\rangle + |01\rangle + |11\rangle)\,|M(0+1)\rangle$$

So (2.3) may hold for states of a basis, but cannot hold for all states. Since a unitary evolution with an ancilla is the most general evolution allowed for quantum systems, the proof of the theorem is concluded.

2.2.2. How no-cloning was actually discovered

Sometimes one learns more from mistakes than from perfect thought. This is, in my opinion, the case with the paper that triggered the discovery of no-cloning [Herbert 1982]. The author, Herbert, reasoned as follows. Consider two particles in the singlet state: one particle goes to Alice, the other to Bob. If Alice measures σ_z, she prepares effectively on Bob's side either $|+z\rangle$ or $|-z\rangle$, with equal probability: therefore Bob's local state is

$$\rho_z = \frac{1}{2}\,|+z\rangle\,\langle +z| + \frac{1}{2}\,|-z\rangle\,\langle -z| = \frac{1}{2}\mathbb{1}$$

If Alice measures σ_x, she prepares effectively on Bob's side either $|{+x}\rangle$ or $|{-x}\rangle$, with equal probability: Bob's local state is now

$$\rho_x = \frac{1}{2}\,|{+x}\rangle\,\langle{+x}| + \frac{1}{2}\,|{-x}\rangle\,\langle{-x}| = \frac{1}{2}\mathbb{1}$$

equal to ρ_z, as it should because of no-signaling.

But suppose now that Bob can make a perfect copy of his qubit. Now, if Alice measures σ_z, Bob ends up with either $|{+z}\rangle\,|{+z}\rangle$ or $|{-z}\rangle\,|{-z}\rangle$, with equal probability: therefore Bob's local state is

$$\rho_z = \frac{1}{2}\,|{+z}+z\rangle\,\langle{+z}+z| + \frac{1}{2}\,|{-z}-z\rangle\,\langle{-z}-z|$$

A similar reasoning leads to the conclusion that, if Alice measures σ_x, Bob's local state is

$$\rho_x = \frac{1}{2}\,|{+x}+x\rangle\,\langle{+x}+x| + \frac{1}{2}\,|{-x}-x\rangle\,\langle{-x}-x|$$

But now, $\rho_x \neq \rho_z$! Herbert suggested that he had discovered a method to send signals faster than light - had he been a bit more careful, he would have discovered the no-cloning theorem.

Astonishingly enough, Herbert's paper was published. Both referees are known: the late Asher Peres explained he knew the paper was wrong, but guessed it was going to trigger interesting developments [Peres 2002]; Gian-Carlo Ghirardi pointed out the mistake in his report, which may therefore be the first "proof" of the no-cloning theorem. Later, Gisin re-considered Herbert's scheme and studied how the two- copy state must be modified in order for ρ_z and ρ_x to be equal after imperfect duplication; he obtained an upper bound for the fidelity of the copies, that can actually be reached [Gisin 1998].

A final point is worth noting. The whole reasoning of Herbert implicitly assumes that something does change instantaneously on a particle upon measuring an entangled particle at a distance. This view is not shared by most physicists. The whole debate might have gone astray on discussions about "collapse", just as the vague reply of Bohr to the Einstein-Podolski-Rosen paper prevented people from defining local variables in a precise way. It is fortunate that people seemed to have learned the lesson, and the problem was immediately cast in an *operational way.* This is the “spirit” that later led to the rise of quantum information science.

2.2.3. The notion of Quantum Cloning Machines

Perfect cloning of an unknown quantum state is impossible; conversely, cloning of orthogonal states belonging to a known basis is trivially possible: simply measure in the basis, and produce as many copies as you like of the state you obtained. What is also possible, is to *swap* the state from one system to the other: $|\psi\rangle |R\rangle \rightarrow |R\rangle |\psi\rangle$ is unitary; one has then created a perfect image of the input state on the second system, at the price of destroying the initial one.

The notion of *Quantum Cloning Machines* (QCM) is a wide notion encompassing all possible intermediate cases. One needs a figure of merit. Here we focus on the *single-copy fidelity*, called “fidelity” for short. For each copied system j, this is defined as $F_j = \langle\psi| \rho_j |\psi\rangle$ for the initial state $|\psi\rangle$.

Here are some intuitive statements:

- If perfect cloning of an unknown quantum state is impossible, *imperfect* cloning should be possible. In particular, there should be an operation that allows to copy equally well, with a fidelity $F <$, any unknown state. Any such operation will imply a “degradation” of the state of the original system.

- The fidelity of the “original” and of the “copy” after the cloning need not be the same; the better the copy, the more the original is perturbed.

- One can also consider cloning $N \rightarrow M = N + k$; if $N \rightarrow \infty$, the fidelity of the final M copies can be arbitrarily close to 1.

- Also, one may be willing to copy only a subset of all states.

Along with the variety of possible approaches, some terminology has been created:

Universal QCM: copies equally well all the states; non-universal QCM, called state-dependent, have been studied only in some cases, mostly related to the attacks of the spy on some cryptography protocols.

Symmetric QCM: the original(s) and the copie(s) have the same fidelity. *Optional QCM*: or a given fidelity of the original(s) after interaction, the fidelities of the copie(s) is the maximal one.

2.2.4. Case study: universal symmetric QCM $1 \to 2$ for qubits

We study in detail the first example of a quantum cloning machine, the universal symmetric QCM $1 \to 2$ for qubits found by Bužek and Hillery [Bužek and Hillery 1996]. It is however instructive to start by analyzing first some trivial cloning strategies.

Trivial cloning

Consider the following strategy: *Let the prepared qubit fly unperturbed, and produce a new qubit in a randomly chosen state, say $|0\rangle$. Don't keep track of which qubit is which.*

Let us compute the single-copy fidelity. We detect one particle: the original one with probability $\frac{1}{2}$, the new one with the same probability. Thus the average single-copy fidelity is

$$\begin{aligned} F_{triv} &= \frac{1}{2} \times 1 + \frac{1}{2} \times \left(\frac{1}{4\pi} \int_0^{2\pi} d\varphi \int_{-1}^{1} d(\cos\theta) \langle\psi| P_0 |\psi\rangle \right) \\ &= \frac{1}{2} + \frac{1}{2} \left(\frac{1}{2} \int_{-1}^{1} d(\cos\theta) \frac{1+\cos\theta}{2} \right) = \frac{3}{4} \end{aligned} \tag{2.4}$$

It is interesting that a fidelity of 75% can be reached by such an uninteresting strategy. In particular, this implies that one must show $F > \frac{3}{4}$ in order to demonstrate non-trivial cloning.

Note that one can consider another trivial cloning strategy, namely: *measure the state in an arbitrary basis and produce two copies of the outcome*. A similar calculation to the one above shows that this strategy leads to $F = \frac{2}{3}$ for the average single-copy fidelity and is therefore worse than the previous one.

The Bužek-Hillery (B-H) QCM for qubits

The Bužek-Hillery (B-H) QCM is a universal symmetric QCM for $1 \to 2$

qubits, that was soon afterwards proved to be the optimal one. We give no derivation, but rather start from the definition and verify all the properties *a posteriori*.

The B-H cloner uses *three qubits*: the original (A), the copy (B) and an ancilla (C). For convention, B and C are initially set in the state $|0\rangle$. Here is the action in the computational basis of A:

$$\begin{aligned} |0\rangle|0\rangle|0\rangle &\to \sqrt{\tfrac{2}{3}}|0\rangle|0\rangle|0\rangle + \sqrt{\tfrac{1}{6}}[|0\rangle|1\rangle + |1\rangle|0\rangle]|1\rangle \\ |1\rangle|0\rangle|0\rangle &\to \sqrt{\tfrac{2}{3}}|1\rangle|1\rangle|1\rangle + \sqrt{\tfrac{1}{6}}[|1\rangle|0\rangle + |0\rangle|1\rangle]|0\rangle \end{aligned} \tag{2.5}$$

These two relations induce the following action on the most general input state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$:

$$|\psi\rangle|0\rangle|0\rangle \to \sqrt{\frac{2}{3}}|\psi\rangle|\psi\rangle|\psi^*\rangle + \sqrt{\frac{1}{6}}[|\psi\rangle|\psi^\perp\rangle + |\psi^\perp\rangle|\psi\rangle]|\psi^{*\perp}\rangle \tag{2.6}$$

We have written

$$|\psi^\perp\rangle = \beta^*|0\rangle - \alpha^*|1\rangle \quad , \quad |\psi^*\rangle = \alpha^*|0\rangle + \beta^*|1\rangle$$

Combining the two definitions one finds that $|\psi^{*\perp}\rangle = |\psi^{\perp *}\rangle$. Eq. (2.6) is the starting point for the subsequent analysis.

The verification that (2.6) follows from (2.5) is made as follows: from (2.5), because of linearity,

$$|\psi\rangle|0\rangle|0\rangle \to \alpha\sqrt{\frac{2}{3}}|000\rangle + \alpha\sqrt{\frac{1}{6}}[|011\rangle + |101\rangle] + \beta\sqrt{\frac{2}{3}}|111\rangle + \beta\sqrt{\frac{1}{6}}[|100\rangle + |010\rangle]$$

Then one writes explicitly the r.h.s. of (2.6) and finds the same state.

B-H: State of A and B

From Eq. (2.6), one sees immediately that A and B can be exchanged, and in addition, that the transformation has the same coefficients for all input states $|\psi\rangle$. Thus, the B-H QCM is symmetric and universal. Explicitly, the partial states are

$$\rho_A = \rho_B = \frac{2}{3}|\psi\rangle\langle\psi| + \frac{1}{3}\frac{\mathbb{1}}{2} = \frac{5}{6}|\psi\rangle\langle\psi| + \frac{1}{6}|\psi^\perp\rangle\langle\psi^\perp| = \frac{1}{2}\left(\mathbb{1} + \frac{2}{3}\hat{m}\cdot\vec{\sigma}\right) \tag{2.7}$$

From the standpoint of A then, the B-H cloner "shrinks" the Bloch vector by a factor $\frac{2}{3}$ without changing its direction.

For both the original and the copy, the B-H cloner gives the fidelity

$$F_A = F_B = \langle\psi|\,\rho_A\,|\psi\rangle = \frac{5}{6} \tag{2.8}$$

This is the optimal fidelity for a symmetric universal $1 \rightarrow 2$ cloner of qubits, a statement that is not evident and was proved in later papers [Gisin 1998, Bruss et al. 1998, Gisin and Massar 1997].

B-H: State of C

Although it is a departure from the main theme, I find it interesting to spend some words about the state of the ancilla C after cloning. No condition has been imposed on this, but it turns out to have a quite interesting meaning. We have

$$\rho_C = \frac{2}{3}\,|\psi^*\rangle\,\langle\psi^*| + \frac{1}{3}\,|\psi^{*\perp}\rangle\,\langle\psi^{*\perp}| = \frac{1}{2}\left(\mathbb{1} + \frac{1}{3}\hat{m}_* \cdot \vec{\sigma}\right) \tag{2.9}$$

with $\hat{m}_* = (m_x, -m_y, m_z)$. This state is related to another operation which, like cloning, is impossible to achieve perfectly, namely the NOT operation that transforms $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$ into $|\psi^\perp\rangle = \beta^*\,|0\rangle - \alpha^*\,|1\rangle$. Because of the need for complex conjugation of the coefficients, the NOT transformation is anti-unitary and cannot be performed[7]. Just as for the cloning theorem, one can choose to achieve the NOT on some states while leaving other states unchanged; or one can find the operation that approximates at best the NOT on all states, called the universal NOT [Bužek et al. 1999]. This operation needs some ancilla, and reads

$$|\psi\rangle \rightarrow \rho_{\mathrm{NOT}} = \frac{2}{3}\,|\psi^\perp\rangle\,\langle\psi^\perp| + \frac{1}{3}\,|\psi\rangle\,\langle\psi|$$

Now, it is easy to verify that $\rho_{\mathrm{NOT}} = \sigma_y \rho_C \sigma_y$ (just see the definition of $\hat{m}_*$). Thus, the ancilla qubit of the B-H QCM carries (up to a rotation of π around

[7] Here is an intuitive version of this impossibility result: any unitary operation on a qubit acts as a rotation around an axis in the Bloch sphere, while the NOT is achieved as the point symmetry of the Bloch sphere through its center. Obviously, no rotation around an axis can implement a point symmetry. A rotation ofπ around the axis z achieves the NOT only for the states in the (x, y) plane, while leaving the eigenstates of σ_z invariant.

the y axis of the Bloch sphere) the universal NOT of the input state.

B-H as the coherent version of trivial cloning

There is an intriguing link between the B-H QCM and the trivial cloning presented above. Recall that in the trivial cloning one has to "forget" which qubit is which in order to pick up one of the two qubits at random. In other words, the process involves summing over the classical permutations. One might ask what happens if this classical permutation is replaced by the *quantum (coherent) permutation*. This operation is defined as the following CP-map:

$$T[\rho] = \frac{2}{3} S_2(\rho \otimes \mathbb{1}) S_2 \tag{2.10}$$

where S_2 is the projector on the symmetric space of two qubits, i.e. the 3-dimensional subspace spanned by $\{|00\rangle, |11\rangle, |\Psi^+\rangle\}$; the factor $\frac{2}{3}$ guarantees that the map is trace-preserving. Remarkably, for $\rho = |\psi\rangle\langle\psi|$, $T[\rho] = \rho_{AB}(\psi)$ as obtained by applying the B-H QCM to $|\psi\rangle$. The map T is not unitary, which justifies the need for the ancilla. This elegant construction was noted by Werner, who used it to find universal symmetric $N \to M$ cloning [Werner 1998].

2.3. Teleportation

The second primitive we are considering is *teleportation*. Discovered over a black-board discussion, the teleportation protocol [Bennett et al. 1993] is a fascinating physical phenomenon (which is more, with a catchy name). Of course, contrary to quantum cloning, teleportation in itself admits few and rather obvious generalizations; in other words, "teleportation" is not and has never been a sub-field of quantum information. However, it is a seed for many other ideas and plays an important role in entanglement theory.

2.3.1. The protocol

As usual in these lectures, we present the protocol with qubits. Consider three qubits: qubit A is prepared in an arbitrary state $|\psi\rangle$; qubits B and C are prepared in a maximally entangled state, say $|\Phi^+\rangle$.

By writing $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ and expanding the terms, one can readily verify

$$\begin{aligned} |\psi\rangle_A |\Phi^+\rangle_{BC} = \frac{1}{2} [& |\Phi^+\rangle_{BC} |\psi\rangle_C + |\Phi^-\rangle_{BC} (\sigma_z |\psi\rangle)_C \\ & + |\Psi^+\rangle_{BC} (\sigma_x |\psi\rangle)_C + |\Psi^-\rangle_{BC} (\tilde{\sigma}_y |\psi\rangle)_C] \end{aligned} \tag{2.11}$$

where $\tilde{\sigma}_y = -i\sigma_y$. This identity is the basis of the *teleportation protocol*:

1. Prepare the three qubits as described above; bring qubits A and B together.

2. Perform a *Bell-state measurement* on qubits A and B, and send the result of the measurement (2 bits) to the location of qubit C.

3. Upon reception of this information, apply the suitable unitary operation to C in order to recover $|\psi\rangle$.

That is all for the protocol. Still, some remarks are worth making:

- It is customary to emphasize that *information, not matter, is teleported*: qubit C had to exist in order to receive the state of qubit A. That being clarified, the name "teleportation" is well chosen: the information has been transferred from A to C without ever being available in the region when particle B has propagated.

- Another point that is usually stressed is the fact that, of course, this task respects *no-signaling*: indeed, the teleportation can be achieved only when the two bits of classical communication are sent, and these cannot travel faster than light. The phenomenon is nonetheless remarkable, because the two classical bits are definitely not sufficient to carry information about a state of a qubit (a vector in the Poincare sphere is defined by three continuous parameters).

- Even though some of the particles may be "propagating", the qubit degree of freedom that is going to be teleported does not evolve in the protocol (there is no hamiltonian anywhere). Teleportation is due to the *purely kinematical* identity (2.11).

- Since the protocol can teleport every pure state with perfect fidelity, it can *teleport any mixed state* as well.

2.3.2. Entanglement swapping

If the qubit to be teleported is itself part of an entangled pair, the identity (2.11) applied to B-CD leads to

$$\begin{aligned} |\Phi^+\rangle_{AB}\,|\Phi^+\rangle_{CD} = \frac{1}{2}\,[&|\Phi^+\rangle_{AD}\,|\Phi^+\rangle_{BC} + |\Phi^-\rangle_{AD}\,|\Phi^-\rangle_{BC} \\ &+ |\Phi^+\rangle_{AD}\,|\Phi^+\rangle_{BC} + |\Phi^-\rangle_{AD}\,|\Phi^-\rangle_{BC}] \end{aligned} \tag{2.12}$$

because

$$\mathbb{1}\otimes\sigma_z\,|\Phi^+\rangle = |\Phi^-\rangle\ ,\ \mathbb{1}\otimes\sigma_x\,|\Phi^+\rangle = |\Psi^+\rangle \ \text{ and } \ \mathbb{1}\otimes\tilde{\sigma}_y\,|\Phi^+\rangle = |\Psi^-\rangle$$

Therefore, by performing the Bell-state measurement on B and C and sending the result to D, one can prepare the particles A and D in the state $|\Phi^+\rangle$, even if they have never interacted. This protocol is called *entanglement swapping* [Yurke and Stoler, 1992, Żukowski et al. 1993].

Entanglement swapping, in itself, is nothing but a special case of teleportation, in which the particle to be teleported is itself part of an entangled pair. However, it shows that *direct interaction is not needed to create entanglement.*

2.3.3. Teleportation and entanglement swapping as primitives for other tasks

Two-qubit maximally entangled states as universal resources The possibility of teleportation has a fundamental consequence in entanglement theory, namely that bipartite maximally entangled states are universal resources to distribute entanglement. Indeed, suppose that N partners want to share a fully N-partite entangled state. It is enough for one of the partners, Paul, to act as a provider. Indeed, if Paul shares a bipartite maximally entangled state with each of the others, he can prepare locally the N-partite state, then teleport the state of each particle to the suitable person.

Now, maximally entangled states of any dimension can be created from enough many copies of two-qubit maximally entangled states (at least on paper), as the following example makes clear:

$$\begin{aligned}|\Phi^+\rangle_{AB}|\Phi^+\rangle_{A'B'} &= \frac{1}{2}\left[|00\rangle_{AA'}|00\rangle_{BB'} + |01\rangle_{AA'}|01\rangle_{BB'}\right.\\ &\quad \left. + |10\rangle_{AA'}|10\rangle_{BB'} + |11\rangle_{AA'}|11\rangle_{BB'}\right]\\ &= \frac{1}{2}\left[|0\rangle_{\mathbf{A}}|0\rangle_{\mathbf{B}} + |1\rangle_{\mathbf{A}}|1\rangle_{\mathbf{B}} + |2\rangle_{\mathbf{A}}|2\rangle_{\mathbf{B}} + |3\rangle_{\mathbf{A}}|3\rangle_{\mathbf{B}}\right] \qquad (2.13)\end{aligned}$$

In conclusion, as soon as the partners can share pairwise maximally entangled states of two-qubits with one provider, they can share any entangled state of arbitrary many parties and dimensions. Let me stress again that this is a theoretical statement about the behavior of entanglement as a resource, not necessarily a practical or even feasible scheme to realize an experiment.

Quantum repeaters

At the other end of the spectrum, the idea of quantum repeaters is triggered by a very practical problem. In quantum communication, one normally uses photons because they propagate well and interact little. In this case, decoherence is by far not the most problematic issue: *losses* are. In other words, quantum communication schemes reach their limits when the photons just don't arrive often enough for the signal to overcome the noise of the detectors and other local apparatuses. Since you cannot amplify your signal because of the no-cloning theorem, losses seem unbeatable. Quantum repeaters are a clever solution to the problem [Briegel et al. 1998, Duan et al. 2001]. For exhaustive information, a recent review article on the topic is available [Sangouard et al. 2009]

Suppose you want to share an entangled pair between locations A and B, at a distance ℓ. The transmission typically scales exponentially $t = 10^{-\alpha\ell}$. What happens if A and B can both send one photon to an intermediate location C, where someone performs the Bell-state measurement? This does not seem to help, because the transmission from A to C is $\sqrt{t}$, and so is the transmission from B to C, so the probability that both photons arrive in C is still t. However, if C has *quantum memories*, the picture changes significantly: now, C can establish a link with A and *independently* a link with B.

This effect is best understood in terms of the time needed to establish an entangled pair between A and B. If photons have to travel from A to B, this time is (in suitable units)

$$\tau_{AB}^{(1)} = \frac{1}{t}$$

In the same units,

$$\tau_{AC}^{(1)} = \tau_{BC}^{(1)} = \frac{1}{\sqrt{t}}$$

If the two links can be established independently, it takes C on average $\frac{1}{2}\frac{1}{\sqrt{t}}$ to establish the first of the links, because the first can be either; conditioned now on the fact that one link was established, C has to wait in average $\frac{1}{\sqrt{t}}$ to establish the second link. After this, entanglement swapping can ?t be performed, thus establishing the entangled pair between A and B in a time

$$\tau_{AB}^{(1)} = \frac{3}{2}\frac{1}{\sqrt{t}}$$

For more details on this calculation, see Appendix B in [Scarani et al. 2009].

2.4. Entanglement distillation

This section is scandalously short. The reason is that I thought other lecturers would introduce the notion of distillation, but it turned out they had planned their lectures differently and there was no time for it! When we realized our lack of coordination, we decided that I would at least mention rapidly the idea, for the sake of completeness. So here it is. All the meaningful notions, extensions and references can be found for instance in the comprehensive review paper devoted to entanglement theory written by the Horodecki family [Horodecki et al. 2009].

2.4.1. The notion of distillation

We have seen above that two-qubit maximally entangled states are a universal resource for distributing quantum states. Suppose now two parties Alice and Bob (it can of course be generalized) share many copies of some other, less entangled state ρ_{AB}: can they somehow "concentrate" or "distill" the entanglement they have, in order to end up with fewer copies of those very useful maximally entangled state? Of course, for the task to make sense, the distillation must be performed only using operations that themselves do not increase entanglement on a single-copy level: these are *local operations* *and classical communication* (LOCC). In other words, Alice and Bob are in different locations and they cannot send quantum systems to each other (otherwise trivially Alice would prepare a maximally entangled state and give or send one particle to Bob). So, is it possible to achieve

$$\rho_{AB}^{\otimes N} \xrightarrow{LOCC} \Phi^{\otimes m} \otimes (\text{garbage}) \tag{2.14}$$

with $\Phi = |\Phi\rangle\langle\Phi|$ a maximally entangled state of two qubits? The answer is: often yes but sometimes no! For instance, entanglement is distillable for all entangled pure states; also for all entangled states of two qubits, or of a qubit and a qutrit. Some entangled mixed state of higher dimensions and/or of more parties, however, are such that their entanglement is not distillable! In other words, you need entanglement to create them, but this entanglement cannot be recovered. Such states are called *bound-entangled.*

At the moment of writing, the question of assessing whether or not an entangled state is distillable is still open in general. Those who want to have an idea of how many notions I am skipping here - partial transpose (a positive, but non-completely-positive map), different measures of entanglement, etc. - may browse the review paper mentioned above. But, in the context of a school, I cannot resist spelling out one of the most fascinating examples of bound-entangled states.

2.4.2. A nice example of bound entangled state

The example of bound entangled state that we are going to consider [Bennett et al. 1999] is a three-qubit state. Suppose that, for any reason, one starts out writing a basis of the Hilbert space as

$$|\varphi_1\rangle = |0\rangle|1\rangle|+\rangle \ , |\varphi_2\rangle = |1\rangle|+\rangle|0\rangle \ , |\varphi_3\rangle = |+\rangle|0\rangle|1\rangle \ , |\varphi_4\rangle = |-\rangle|-\rangle|-\rangle$$

these states are obviously orthogonal, but four more states are needed to have a basis. Now, it turns out that the four remaining states cannot be product states: they must contain some entanglement! The four product states we started with form a so-called *unextendible product basis.*

The next ingredient is the fact that the remaining four states can all be taken to be entangled only between the first two qubits; i.e. one can find four orthogonal vectors

$$|\varphi_k(CA|B)\rangle = |\Psi_k\rangle_{CA}|\psi_k\rangle_B \ \text{ or } \ |\varphi_k(BC|A)\rangle = |\Psi_k\rangle_{BC}|\psi_k\rangle_A$$

Consider now the state

$$\rho = \frac{1}{4}\left(\mathbb{1} - \sum_{k=1}^{4} |\varphi_k\rangle\langle\varphi_k|\right) \tag{2.15}$$

This is the maximally mixed state defined on the subspace that is complementary to the unextendible product basis. As such, it is obviously entangled, because there cannot be any product state in its support. But *where* is the entanglement? Notice that

$$\begin{aligned}\rho &= \frac{1}{4}\sum_{k=5}^{8} |\varphi_k(AB|C)\rangle\langle\varphi_k(AB|C)| \\ &= \frac{1}{4}\sum_{k=5}^{8} |\varphi_k(CA|B)\rangle\langle\varphi_k(CA|B)| \\ &= \frac{1}{4}\sum_{k=5}^{8} |\varphi_k(BC|A)\rangle\langle\varphi_k(BC|A)| \end{aligned}\tag{2.16}$$

therefore, according to *each* possible bipartition, the state is separable. It is therefore impossible that separate parties can distill entanglement: ρ is bound-entangled.

Is bound entanglement "useful"? For long time, the answer was supposed to be negative. In 2005, a theoretical breakthrough showed that some bound entangled states contain secrecy, i.e. can be used for cryptography. In order to explore this topic, I suggest to read first [Scarani et al. 2009], II.B.2, then the original references given there. However, no explicit protocol to distribute such states has ever been devised, nor will probably ever be: other protocols are much simpler and efficient for the task.

2.5. Tutorials

2.5.1. Problems

Exercise 2.1

Amplification of light is of course compatible with the no-cloning theorem, because spontaneous emission pre- vents amplification to be perfect [Mandel 1983]. Actually, if the amplifier is independent of the polarization, universal symmetric cloning of that degree of freedom is implemented [Simon et al.

2000, Kempe et al. 2000]. In this problem, we explore the basics of this correspondence.

Consider a single spatial mode of the electromagnetic field and focus on the polarization states; we denote by $|n,m\rangle$ the state in which n photons are polarized H and m photons are polarized V. Suppose one photon in mode H is initially present in the amplifier, and that after amplification 2 photons have been produced.

1. Compute the single-copy fidelity of this cloning process. Hint: if you don't remember the physics of amplification, you can reach the result by comparing $a_H^\dagger |1,0\rangle$ with $a_V^\dagger |1,0\rangle$.
2. How would you describe the state of the system (field + amplifier medium) in this process? Hint: compare with the B-H QCM.

2.5.2. Solutionss

Exercise 2.1

1. The theory of spontaneous and stimulated emission implies that, starting with $|1,0\rangle$, the probability of creating $|2,0\rangle$ is twice as large as the probability of creating $|1,1\rangle$. The single-copy fidelity is defined as the probability of finding one of the photons in the initial state, whence obviously
$$F = \frac{2}{3} \times 1 + \frac{1}{3} \times \frac{1}{2} = \frac{5}{6}$$
This is identical to the fidelity for optimal universal symmetric cloning.
2. The analogy with cloning is actually exact: indeed, by conservation of angular momentum, the emission of an H photon and of a V photon cannot be due to the same process. Therefore, after amplification and post-selection of the emission of two photons, the state of the system "field + amplifying medium" reads
$$\sqrt{\frac{2}{3}}\,|2,0\rangle \otimes |e_H\rangle + \sqrt{\frac{1}{3}}\,|1,1\rangle \otimes |e_V\rangle$$
i.e., in first-quantized notation
$$\sqrt{\frac{2}{3}}\,|H\rangle\,|H\rangle \otimes |e_H\rangle + \sqrt{\frac{1}{3}}\,|\Psi^+\rangle \otimes |e_V\rangle$$

and this exactly the state produced by the B-H QCM.

3. Lecture: Primitives of quantum information (II)

3.1. State discrimination

Under the head of state discrimination, a large variety of tasks can be accommodated. For a school, rather than reviewing each of them exhaustively, I find it more useful to present concrete examples of each. A review article was written by Chelfes in the year 2000 [Chefles 2000]: it contains most of the basic ideas; some recent developments will be mentioned below.

3.1.1. Overview

As the name indicates, state discrimination refers to obtaining information about the quantum state produced by a source that is not fully characterized. We can broadly divide the possible tasks in two categories:

- *Single-shot tasks*: the goal is to obtain information on each signal emitted by the source. Without further information, as well-known, the task is almost hopeless: basically, after measuring one system, the only thing one can be sure of is that the state was not orthogonal to the one that has been detected. However, the task becomes much more appealing if some additional knowledge is present: for instance, if one is guaranteed that each system can be either in state ρ_1 or in state ρ_2, with the ρ_j two well-specified states. In such situations, one can consider *probabilistic state discrimination* and try to minimize the probability of a wrong guess; or even *unambiguous state discrimination*, a POVM that either identifies the state perfectly or informs that the discrimination was inconclusive.

- *Multi-copy tasks*: if the source is guaranteed to produce always the same state, the state can be exactly reconstructed asymptotically; this process is called *state reconstruction*, or *state estimation*, or *state tomography.* If, in addition, one knows that the state is either ρ_1 or ρ_2, one can ask how fast the probability of wrong guess decreases with the number of copies and obtain a quantum version of the *Chernoff bound.*

If the source is not guaranteed to produce always the same state, the task seems hopeless, and in full generality it is; but if the observed statistics are symmetric under permutation, a quantum version of the *de Finetti theorem* exists.

3.1.2. Single-shot tasks (I): Probabilistic discrimination

Probabilistic discrimination of two states

Let us consider the simplest case: two states ρ_1 and ρ_2 are given each with probability η_1 and $\eta_2 = 1 - \eta_1$. At each run, one performs a measurement, whose outcome is used to guess which state was given; we want to *minimize the probability P_{error} that the guess is wrong.*

Since ultimately we want two outcomes, without loss of generality the measurement can be described by two projectors Π_1 and $\Pi_2 = \mathbb{1} - \Pi_1$ where $\mathbb{1}$ is the identity over the subspace spanned by ρ_1 and ρ_2. Therefore, the probability of guessing ρ_j correctly, i.e., the probability of guessing j given ρ_j is given by $\mathrm{Tr}(\Pi_j\rho_j)$; whence the average probability of error for this measurement is

$$P_{error}(\Pi_1) = 1 - \sum_{j=1,2} \eta_j \mathrm{Tr}(\Pi_j\rho_j) = \eta_1 - \mathrm{Tr}(\Pi_1(\eta_1\rho_1 - \eta_2\rho_2)) \tag{3.1}$$

In order to minimize this, we have to find the projector Π_1 that maximizes the second term on the right-hand side. The result is [Helstrom 1976, Herzog and Bergou 2004]

$$P_{error} = \frac{1}{2}\left[1 - \mathrm{Tr}|\eta_1\rho_1 - \eta_2\rho_2|\right] \tag{3.2}$$

A constructive measurement strategy that would lead to this optimal result is the following: measure the Hermitian operator $M = \eta_1\rho_1 - \eta_2\rho_2$; if the outcome is a positive eigenvalue, guess ρ_1, if it's a negative eigenvalue, guess ρ_2.

Let us prove (3.2) for the special case of *equal a priori probabilities* $\eta_1 = \eta_2 = \frac{1}{2}$. In this case, Π_1 is the projector on the subspace of positive eigenvalues of $\rho_1 - \rho_2$; but since $\mathrm{Tr}(\rho_1 - \rho_2) = 0$, the sum of the positive eigenvalues and of the negative ones must be the same in absolute value. Therefore we find

$$\max_{\Pi_1} \mathrm{Tr}(\Pi_1(\rho_1 - \rho_2)) = \frac{1}{2}\mathrm{Tr}|\rho_1 - \rho_2|$$

and finally

$$P_{error} = \frac{1}{2}\left[1 - \frac{1}{2}\mathrm{Tr}|\rho_1 - \rho_2|\right] \quad (\eta_1 = \eta_2 = \frac{1}{2} \tag{3.3}$$

Intermezzo: trace distance

The mathematical object

$$D(\rho,\sigma) = \frac{1}{2}\mathrm{tr}|\rho - \sigma| \tag{3.4}$$

that appeared in the previous proof is called *trace-distance* between two states ρ and σ. This quantity appears often in quantum information, so it is worth while spending some time on it (for all proofs and more information, refer to chapter 9 of [Nielsen and Chuang 2000]). Some simple properties of the trace distance are: $D(\rho,\sigma) = 0$ if and only if $\rho = \sigma$; the maximal value $D(\rho,\sigma) = 1$ is reached for orthogonal states; $D(\rho,\sigma) = D(\sigma,\rho)$. It can moreover be proved that the triangle inequality

$$D(\rho,\tau) \leq D(\rho,\sigma) + D(\sigma,\tau)$$

holds; therefore it has the mathematical properties of a "metric", i.e., it defines a valid distance between states. As we have seen, two states can be distinguished with probability at most

$$\frac{1}{2}[1 + D]$$

But $\frac{1}{2}$ is sheer random guessing: rewriting

$$\frac{1}{2}[1 + D] = D \times 1 + (1 - D) \times \frac{1}{2}$$

we see that, *in any task*, the two states behave differently with probability at most D. Indeed, suppose there is a task for which the two behaviors are more distinguishable: we would use that task as measurement for discrimination, thus violating the bound (3.3). The usefulness of this remark becomes even more apparent when phrased in a slightly different context. Instead of having to discriminate two states, suppose one has a state ρ and wants to compare it to an "ideal" state ρ_{ideal}. Then $D(\rho,\rho_{ideal})$ is the *maximal probability of failure*, i.e., the maximal probability that the real state will produce a result different than the one the ideal state would have produced. This plays a central role, for instance, in the definition of security in quantum cryptography;

see II.C.2 in [Scarani et al. 2009] for a discussion and original references. We shall find this idea in a different context in Section 6.2.2.

PSD of more than two states

In the most general case, as expected, the optimal PSD strategy is not known. Sub-optimal strategies are trivially found: just invent a measurement strategy and compute the probabilities of failure. A particular strategy performs often quite well, so much so that it has been called *pretty good measurement* (PGM) [Hausladen and Wootters 1994]. It is defined as follows: let $\{\rho_k, p_k\}$ be the set of states to be distinguished, with the corresponding a priori probabilities; and define

$$M = \sum_k p_k \rho_k$$

Then the PGM is the POVM whose elements are defined by

$$E_k = p_k M^{-1/2} \rho_k M^{1/2}$$

The special case, in which all the states to be distinguished are pure ($\rho_k = |\psi_k\rangle\langle\psi_k|$) and equally probable, is known as *square-root measurement*; in this case, the elements of the POVM are the suitably normalized projectors on the states $|\chi_k\rangle \propto M^{-1/2}|\psi_k\rangle$ [Hausladen et al. 1996]. There is a significant amount of literature on these measurements, that in some case define the optimal measurement strategy. Its review goes beyond our scope.

Let us finally mention that an upper bound on the guessing probabilities can be found by studying the dual problem[8] [Koenig et al. 2008].

PSD and cloning

There is an interesting, somehow intuitive link between optimal cloning and

[8] The dual problem is related to the definition of min-entropies. Let $\{(\rho_k, \eta_k)\}_{k=1\dots n}$ be the states to be distinguished with the respective a priori probabilities. Then one forms the *classical-quantum* state

$$\rho_{AB} = \sum_k \eta_k |k\rangle\langle k| \otimes \rho_k$$

Choose now a state σ_B and compute the minimum λ such that $M = \lambda \mathbb{1}_A \otimes \sigma_B - \rho_{AB}$ s a non-negative operator: then $\lambda \geq 1 - P_{error}$, with the guarantee of equality for the minimum over all possible σ_B.

PSD, namely: for any ensemble of *pure* states $\{(|\psi_k\rangle, \eta_k)\}_{k=1\ldots n}$ to be distinguished, optimal PSD is equivalent to optimal symmetric $1 \to N$ cloning in the limit $N \to \infty$ [Bae and Aćʔn 2006]. Note that this fact does not help to find the explicit strategy, since optimal state-dependent cloners are not known in general either.

The argument goes as follows. For convenience, define F_C as the single-copy fidelity of the optimal $1 \to \infty$ cloner and F_M as the fidelity of the state reconstructed after the optimal PSD measurement. It is obvious that $F_M \leq F_C$: after measurement, we have a guess for the state, so we can just create as many copies as we want of that state, so this defines a possible cloner. The proof of the converse is more tricky. Basically, one applies the cloner C to a half of a maximally entangled state: $(\mathbb{1} \otimes C)|\Phi^+\rangle = \rho_{AB_1\ldots B_N}$. By assumption, all the ρ_{AB} are equal. In the limit $N \to \infty$ the information of B_j is "infinitely shareable" and a theorem then guarantees that ρ_{AB_j} is separable. But then, the restriction $\tilde{C}$ defined as $(\mathbb{1} \otimes \tilde{C})|\Phi^+\rangle = \rho_{AB_1}$ describes an entanglement-breaking channel, and it is known that any such channel is equivalent to performing a measurement and forwarding the collapsed state. In conclusion, there exist a measurement strategy that achieves the single-copy fidelity of the optimal cloner.

3.1.3. Single-shot tasks (II): Unambiguous state discrimination

We consider now a different situation: now we want discrimination to be unambiguous; the price to pay is that sometimes the procedure will output an inconclusive outcome.

Unambiguous discrimination of two pure states

The case where the two state have equal a priori probabilities was solved independently by Ivanovics, Dieks and Peres [Ivanovic 1987, Dieks 1988, Peres 1988]; Jaeger and Shimony later solved the problem for arbitrary probabilities [Jaeger and Shimony 1995]. See also [Peres 1995], sect. 9-5. For two *pure* states, the discrimination succeeds at most with probability $1 - |\langle\psi_1|\psi_2\rangle|$; of course, with probability $|\langle\psi_1|\psi_2\rangle|$ one obtains the inconclusive outcome.

Let us begin by describing a simple setup that achieves USD of two pure states [Huttner et al. 1996]; an even simpler one is given as a tutorial. The setup is shown in figure 3.1.

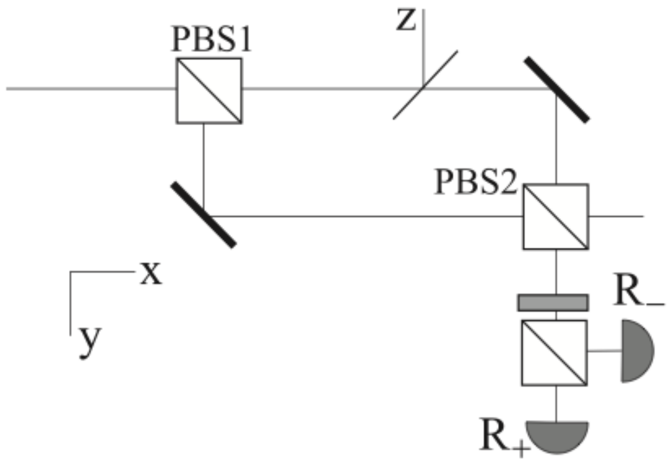

Figure 3.1: Optical setup for unambiguous discrimination of non-orthogonal polarization states.

The polarizing beam-splitters PBS1 and PBS2 are oriented such that $|H\rangle$ is transmitted and $|V\rangle$ is reflected. The goal is to distinguish between

$$|\psi_{\pm}\rangle = \cos\alpha\,|H\rangle \pm \sin\alpha\,|V\rangle \tag{3.5}$$

whose overlaps is $\chi = \cos 2\alpha$. Indeed, we have

$$\begin{aligned}(\cos\alpha\,|H\rangle \pm \sin\alpha\,|V\rangle)\,|x\rangle &\xrightarrow{PBS1} \cos\alpha\,|x\rangle \pm \sin\alpha\,|y\rangle \\ &\xrightarrow{BS\,x-z} \cos\alpha\sqrt{t}\,|H\rangle\,|x\rangle \pm \sin\alpha\,|V\rangle\,|y\rangle + \cos\alpha\sqrt{1-t}\,|H\rangle\,|z\rangle \\ &\xrightarrow{mirrors} i\cos\alpha\sqrt{t}\,|H\rangle\,|y\rangle \pm i\sin\alpha\,|V\rangle\,|x\rangle + \cos\alpha\sqrt{1-t}\,|H\rangle\,|z\rangle \\ &\xrightarrow{PBS2} i(\cos\alpha\sqrt{t}\,|H\rangle\,|y\rangle \pm i\sin\alpha\,|V\rangle)\,|y\rangle + \cos\alpha\sqrt{1-t}\,|H\rangle\,|z\rangle\end{aligned}$$

If $\cos\alpha\sqrt{t} = \sin\alpha$, that is for transmission of the BS $t = \tan^2\alpha$, then the two states that can appear in mode y are orthogonal and can be discriminated by a usual projective measurement (output modes $|R_{\pm}\rangle$ of a PBS). In summary, we have implemented the transformation

$$(\cos\alpha\,|H\rangle \pm \sin\alpha\,|V\rangle)\,|x\rangle \longrightarrow i\sqrt{2}\sin\alpha\,|\pm\rangle\,|R_{\pm}\rangle + \cos\alpha\sqrt{1-\tan\alpha}\,|H\rangle\,|z\rangle \tag{3.6}$$

where the probability of obtaining a conclusive result is the optimal one, since

$$2\sin^2\alpha = 1 - \cos 2\alpha = 1 - |\,\langle\psi_+\,|\,\psi_-\rangle\,|$$

Taking a more formal view, this setup realizes a POVM: there are three outcomes for a qubit. The two conclusive outcomes are described by

$$A_+ = \eta\,|+\rangle\,\langle|\psi_-^\perp\rangle|\ \ '\ \ A_- = \eta\,|-\rangle\,\langle|\psi_+^\perp\rangle| \tag{3.7}$$

with

$$|\psi_\pm^\perp\rangle = \sin\alpha\,|H\rangle \mp \cos\alpha\,|V\rangle$$

The factor η is determined by the constraint that the largest eigenvalue of $A_+^\dagger A_+ + A_-^\dagger A_-$ should be 1; direct inspection leads to

$$\eta = \frac{1}{\sqrt{2}\cos\alpha}$$

The probability of guessing correctly $|\psi_\omega\rangle$ is given by

$$\mathrm{Tr}(A_\omega\,|\psi_\omega\rangle\,\langle\psi_\omega|\,A_\omega^\dagger) = 2\sin^2\alpha = 1 - \cos 2\alpha$$

as it should.

Generalizations: more pure states, mixed states

For any number *pure* states, unambiguous state discrimination is possible if and only if the states are linearly independent. This implies in particular that no set consisting of $N > d$ states, where d is the dimension of the Hilbert space, can be discriminated unambiguously.

Though not optimal in general, the following strategy can always be implemented. Consider N linearly independent states $\{|\psi_k\rangle\}_{k=1,\dots,N}$ generating the N-dimensional subspace $\mathcal{E}$. Let $|\varphi_k\rangle$ be the vector in $\mathcal{E}$ that is orthogonal to all the vectors except $|\psi_k\rangle$: of course, if one performs a measurement and finds the result $|\varphi_k\rangle$, one can unambiguously identify $|\psi_k\rangle$ as being the measured state. This looks very much like a usual projective measurement defined by the projector $P_k = |\varphi_k\rangle\,\langle\varphi_k|$; however, the $|\varphi_k\rangle$'s are in general not orthogonal (if they are orthogonal, they coincide with $|\psi_k\rangle$ that are therefore already orthogonal), so $\sum_k P_k$ does not need to be proportional to $\mathbb{1}$ and its largest eigenvalue may be $\lambda > 1$. A possible way of obtaining a valid POVM consists in defining $A_k = P_k/\sqrt{\lambda}$ and associate the inconclusive outcome to $A_0^\dagger A_0 = \mathbb{1} - \sum_k P_k/\lambda$. When it comes to unambiguous discrimination of *mixed states*, the situation is far more complex. For instance, even two different mixed states cannot be discriminated unambiguously if their supports are identical (obviously, two pure states have identical support if and only if they are the same state). Some non-trivial examples have been worked out [Raynal and Lütkenhaus 2005].

3.1.4. Multi-copy tasks (I): tomography, Chernoff bound

Tomography of an a priori completely unknown state

The idea of tomography is inherent to the notion of state itself. A state is a description of our statistical knowledge. In classical statistics, a probability distribution can be reconstructed by collecting many independent samples that follow the distribution. Similarly, if one performs suitable measurements on N copies of an unknown quantum state ρ, one can obtain a faithful description of the state itself.

Some examples will suffice to clarify the principle. Any one-qubit state can be written as

$$\rho = \frac{1}{2}(\mathbb{1} + \vec{m} \cdot \vec{\sigma})$$

where $m_k = \langle \sigma_k \rangle$. Therefore, if one can estimate the average values $\langle \sigma_x \rangle$, $\langle \sigma_y \rangle$ and $\langle \sigma_z \rangle$, one can reconstruct the state. Similarly, for two qubits, one has to estimate the fifteen average values $\langle \sigma_k \otimes \mathbb{1} \rangle$ (three), $\langle \mathbb{1} \otimes \sigma_k \rangle$ (three) and $\langle \sigma_j \otimes \sigma_k \rangle$ (nine). It's exactly like reconstructing a classical distribution, but for the fact that one has to perform several different samplings (measurements) because of the existence of incompatible physical quantities.

While the principle is simple and necessary for the notion of state to have any meaning, the topic is not closed. For instance, one can look for "optimal" tomography according to different figure of merit (smallest size of the POVM, faster convergence with the number of samples N, etc.).

Among the rigorous results that have been obtained only recently, let us mention the computation of the *quantum Chernoff bound*. The task is somehow half-way between state estimation and tomography: given the promise that the state is either ρ_1 or ρ_2, one wants to estimate *how fast* the probability of distinguishing increases for increasing N. The result is most easily formulated in terms of the probability of error, which has been proved to decrease exponentially:

$$\frac{1}{2}\left[1-\frac{1}{2}\left|\eta_1\rho_1^{\otimes N}-\eta_2\rho_2^{\otimes N}\right|\right]\sim e^{-\xi N}\ \text{ with }\ \xi=-\log\left(\min_{0\leq s\leq 1}\mathrm{Tr}(\rho_0^4\rho_1^{1-s})\right)$$

[Audenaert et al., 2007].

3.1.5. Multi-copy tasks (II): de Finetti theorem and extensions

The tomography discussed above works only under the assumption that the source produces always the same state, i.e. that the state of N emitted particles is simply $\rho^{\otimes N}$. Operationally, this cannot be guaranteed: all that one can guarantee is that the experimental procedure is the same for each realization. This implies that the N-particle state ξ_N, whatever it is, is such that all possible statistics are invariant by permutation of the particles (for we have no information that allows to distinguish some realizations from others).

Now, the statement we want to make is the following: *in the limit of large N, any state ξ_N that is invariant by permutation is "close" to a product state* $\rho^{\otimes N}$ or to a classical mixture of such states[9]. It is not superfluous to notice that we use this result, implicitly, in every laboratory experiment, be it about classical or quantum physics.

There are several proofs of this statement, differing in important details (what "to be close" exactly means) and consequences. Most of them are called "de Finetti theorems", from the name of the Italian mathematician who first proved a similar statement in the context of classical probability theory; they consist on an estimate of the trace distance between the real state and the set of product states. For a very clear presentation, including a review of previous works, we refer the student to [Renner 2007]. The most recent

[9] The careful reader may immediately notice that the statement, as loosely stated, cannot be strictly true: for instance, the GHZ state $|000...0\rangle+|111...1\rangle$ is invariant under permutation but is far, under any measure, from a product state. Indeed, the exact statement says that: if ξ_N is invariant under permutation, there exist $k<N$ such that $\mathrm{Tr}_k\xi_N$ is close to a mixture of product states (see how this is the case for the GHZ state, already with $k=1$).

extension, that obtains a much better bound under relaxed assumptions, is based on the idea of post-selection [Christandl et al. 2009].

3.2. Quantum coding

By quantum coding I denote the generalization to quantum physics of the main results of classical information theory. A very thorough presentation of the basic material is already available in chapters 11 and 12 of Nielsen and Chuang's book [Nielsen and Chuang 2000]: unless other references are given, this section is a distillation of that source, to which the reader should also refer to obtain the references to the original works. The studies of security of quantum cryptography are of course an offspring of this field, so several notions will be presented in due detail in the specific series of lectures.

3.2.1. Shannon entropy and derived notions

It should be well-known that "entropy" is associated first with "uncertainty" in information theory. The elementary entropic quantity in quantum physics is the *von Neumann entropy*, which is nothing else than the Shannon entropy of the eigenvalues $\{\lambda_k\}$ of a given state ρ:

$$S(\rho) = -\mathrm{Tr}(\rho \log \rho) = -\sum_k \lambda_k \log \lambda_k \tag{3.8}$$

Many of the properties of Shannon entropy translate directly to von Neumann entropy. For instance, it?s a concave function: the entropy of a mixture is larger than the average of the entropies, formally

$$S\left(\sum_k p_k \rho_k\right) \geq \sum_k p_k S(\rho_k) \tag{3.9}$$

One can also define "relative entropy"

$$S(\rho\|\sigma) = -S(\rho) - \mathrm{Tr}(\rho \log \sigma)$$

with similar properties as the classical analog. Of course, the quantum joint entropy is just defined as $S(A,B) = S(\rho_{AB})$. Again, it behaves like the classical analog; in particular, the property called "strong subadditivity" holds:

$$S(A,B,C) + S(B) \leq S(A,B) + S(B,C)$$

Not exactly everything is a copy of classical information theory, though: a most remarkable exception is the possibility for *conditional entropy* to be negative. Classical conditional entropy for a probability distribution $P(a,b)$ is the uncertainty on the distribution on A knowing B: it's the entropy of the conditional probabilities, averaged over all possible conditions, i.e.,

$$H(A|B) = \sum_b P(b) \left[- \sum_a P(a|b) \log P(a|b) \right]$$

This definition cannot be generalized as such in quantum physics, because there is no obvious analog of $P(a|b)$; his definition cannot be generalized as such in quantum physics, because there is no obvious analog of P (a|b); however, it is simple to show that

$$H(A|B) = H(A,B) - H(B)$$

and this expression can be generalized:

$$S(A|B) = S(\rho_{AB}) - S(\rho_B) \tag{3.10}$$

Consider now, as an example, a maximally entangled state of two qubits $\rho_{AB} = |\Phi^+\rangle \langle\Phi^+|$: this state is pure, so $S(\rho_{AB}) - = 0$; however, ρ_B is maximally mixed, whence $S(\rho_B) = 1$; so in all $S(A|B) = -1$. This is a manifestation of one of the unexpected features of entanglement: the fact that one has sharp properties for the composite system that do not derive from sharp properties of the sub-systems.

Interestingly, conditional entropy has an operational interpretation as the amount of quantum information needed to transfer a state ρ_{AB} , initially shared between two parties, to one of the parties, while keeping the coherence with possible purifying systems that are not available to either party [Horodecki, Oppenheim and Winter 2005]. When the quantity is negative, it basically means that, after the transfer, the parties still keep some quantum correlations[10] that can be used to transfer additional quantum informations via teleportation.

[10] In the extreme case of pure states, this is pretty clear. Alice and Bob know which state they share. If the state is pure, there is no coherence with a purifying system to be preserved! Therefore, Bob can just generate the state in his own location: no resources are used, and the shared states are still available for teleportation.

3.2.2. From Holevo to Schumacher, and beyond

We start by presenting the two best-known early results in quantum coding: the Holevo bound and Schumacher's compression.

Holevo bound: classical information coded in quantum states

Suppose Alice prepares any of the states $\{\rho_x\}_{x=1,..,N}$, each with some probability p_x and sends them to Bob on a noiseless channel. Bob performs a measurement, generically a POVM, with outcomes $y \in \{1,...,m\}$. Since ultimately both Alice's and Bob's variables are classical, the amount of information that Bob has obtained through the measurement is given by the Shannon mutual information

$$I(X:Y) = H(X) + H(Y) - H(X,Y)$$

The *Holevo bound* is an upper bound on this amount of information:

$$I(X:Y) \leq \xi_{\{p_x,\rho_x\}} \equiv S\left(\sum_x p_x\rho_x\right) - \sum_x p_x S(\rho_x) \tag{3.11}$$

The left-hand side is a purely classical quantity, because X and Y are classical; the right-hand side is quantum, because classical information has been coded in quantum states. The bound can be saturated only if the states ρ_x are mutually commutative. We shall meet below an extension of this bound.

Maybe it's convenient to stress something at this point. The configuration envisaged in this paragraph looks at first like a mathematical exercise with little or no usefulness in practice. Indeed, the task is "Alice wants to send a non-secret message to Bob". To achieve this, Alice most probably would not use quantum states at all in the first place; or more precisely, she'd use orthogonal states (this is classical communication), so that Bob would have no trouble discriminating them. So, why should one bother about such an artificial task as sending messages with non-orthogonal states? The answer is the following: in itself, the situation is artificial indeed; but this situation may appear as natural in the context of another task. For instance, consider the effective channel linking Alice to Eve in quantum cryptography. Here, Alice does *not* want Eve to learn the message: it's Eve that sneaks in and get whatever she can. It is therefore normal that the channel Alice-Eve is not optimized for direct communication. It turns out that this channel is precisely of the Holevo type: Alice's bits are encoded in quantum states that Eve keeps (see also Tutorial, and the series of lectures on quantum cryptography in this school).

Schumacher compression of quantum information

Consider a source producing classical symbols x, independent and identically distributed (i.i.d.) according to a distribution $p(x)$. One wants to know how much a message can be compressed and later decompressed without introducing errors, i.e., $m(n)$ in

$$(x_1, ..., x_n) \xrightarrow{\mathcal{C}} (y_1, ..., y_m) \xrightarrow{\mathcal{D}} (x_1, ..., x_n)$$

A well-known theorem by Shannon says that, asymptotically, $m(n) = nH(X)$.

Schumacher's coding is the quantum generalization of this theorem. The source is now represented by the mixed state

$$\rho = \sum_x p(x) \left|x\right\rangle \left\langle x\right|$$

living in a d-dimensional Hilbert space $\mathcal{H}$ (i.e., $\log d$ qubits). The compression-decompression process is now

$$\rho^{\otimes n} \xrightarrow{\mathcal{C}} \sigma \xrightarrow{\mathcal{D}} \rho^{\otimes n}$$

The result is that σ must live in a Hilbert space of dimension $2^{nS(\rho)}$.

The richness of quantum information theory

With classical signals, one can basically send classical information. When one realizes the huge body of knowledge that this has generated [Cover and Thomas 2006], the complexity of opening the box to quantum physics becomes striking. Indeed, once quantum systems are brought into the game, the number of possible situations that one can envisage explodes:

- The *nature of the coding*: classical information theory has classical bits (*c-bits*). In quantum informa- tion theory, one has of course to add quantum bits (*q-bits* or *qubits*), but also more complex units like bits of entanglement (*e-bits*) because states of two qubits are not the same of two states of qubits.

- The *available resources*: in classical information theory, one has classical channels for communication and shared randomness for possible pre-established correlations. Now we have to add quantum channels (those that allow to send qubits), pre-established entanglement... For instance, one can study what can be done with a classical channel assisted with shared entangled pairs; and all possible combinations.

Many results we have reviewed, including Schumacher's, may give the impression that ultimately one will always find the analog of known results in classical information theory. However, this is *not* the case, and the reason is *entanglement*. Indeed, a crucial assumption in Schumacher?s result is that the source is i.i.d. We are going to gain more insight on this point by studying *channel capacities*, a field in which several ground-breaking results have been found only recently.

3.2.3. Channel capacities: a rapid overview

Definition of quantum channel capacity

Alice wants to send m x-bits to Bob (x can be c, q, e...). She encodes her information in a state of n qubits (possibly entangled) and sends these qubits over a quantum channel T to Bob, who decodes the information correctly with probability $\varepsilon(m,n)$. The capacity of this channel is given by

$$C_x(T) = \sup \lim_{n\to\infty} \frac{m}{n} \quad \text{such that} \quad \lim_{n\to infty} \varepsilon(m,n) = 0 \tag{3.12}$$

The supremum is taken over all possible choice of coding and decoding. In general therefore, for a given quantum channel T, several different capacities can be defined, depending on the nature of the information to be transmitted (x-bits), but also on possible additional resources, on the requested speed of convergence of the error rate etc.

Case study: classical capacity of a quantum channel

For definiteness, let us focus on the *classical capacity* of a quantum channel, written $\mathcal{C}(T)$. The scenario is the one considered by Holevo: Alice codes classical information X in quantum states, sends them to Bob who performs a POVM and extracts classical information Y. The only difference is that now the quantum systems are sent over a non-trivial channel T: i.e., if Alice sends ρ_k, Bob receives $T[\rho_x]$.

Under the assumption of an i.i.d. source, we can easily understand that

$$\mathcal{C}_{i.i.d.}(T) = \chi(T) = \max_{p_x,\rho_x} \chi_{\{p_x,T[\rho_x]\}} \tag{3.13}$$

But is it possible to achieve a larger capacity with a non i.i.d. source? In classical information, this is known not to help: one can always maximize the rate of a channel with i.i.d. sources. In this case, one says that *capacity is additive*. But quantum physics allows for entanglement: Alice may associate the sequence (x_1, x_2) to an entangled state ρ_{x_1,x_2}. Can this help? For long time, the answer was conjectured to be negative, though explicit proofs were available only for some particular channels. In September 2008, however, Hastings proved that the conjecture is in general wrong: there exist channels, whose full classical capacity cannot be reached by i.i.d. sources - in other words, for some channels, entanglement does help even if you are using the channel to share classical information!

Because of this, the general expression of the classical capacity of a quantum channel is

$$\mathcal{C}(T) = \lim_{n\to\infty} \frac{1}{n}\chi(T^{\otimes n}) \quad \text{with} \quad \chi(T^{\otimes n}) = \max_{p_{\mathbf{x}},\rho_{\mathbf{x}}} \chi_{\{p_{\mathbf{x}},T^{\otimes n}[\rho_{\mathbf{x}}]\}} \tag{3.14}$$

now the maximum must be taken over all possible choice of n-qubit states, each coding for the classical information $\mathbf{x} = (x_1, ..., x_n)$. There is manifestly no hope of computing such a maximum by brute force. This is the reason why even the “simple” classical capacity of quantum channels in not known for arbitrary channels.

Other capacities

Among the other possible capacities of a quantum channel, two are worth at least mentioning:

- The *classical private capacity* is associated to the task of sending classical information while keeping it secret from the environment. Its expression is [Devetak 2005]:

$$\mathcal{P}(T) = \lim_{n\to infty} \frac{1}{n} \max_{p_{\mathbf{x}},\rho_{\mathbf{x}}} \left[\chi_{\{p_{\mathbf{x}},T^{\otimes n}[\rho_{\mathbf{x}}]\}} - \chi_{\{p_{\mathbf{x}},\tilde{T}^{\otimes n}[\rho_{\mathbf{x}}]\}}\right] \tag{3.15}$$

where $\tilde{T}$ is the "complementary channel", i.e., the information that leaks into the environment - as such, the formula is somehow intuitive.

- The *quantum capacity* is the capacity of the channel when Alice wants to send quantum information. It is generally written $\mathcal{Q}$; its expression would require introducing additional notions that are beyond the scope of our survey.

In general, it holds

$$\mathcal{C}(T) \geq \mathcal{P}(T) \geq \mathcal{Q}(T) \tag{3.16}$$

he fact that the first inequality can be strict is almost obvious; less obvious is the fact that the second inequality can also be strict [Horodecki, Horodecki et al. 2005]. The proofs of non-additivity are all very recent: Smith and Yard had proved the non-additivity of the quantum capacity in Summer 2008 [Smith and Yard 2008]. Later, the conjecture that the private capacity may also be non-additive was formulated [Smith and Smolin 2009] and proved [Li et al., 2009]. At the moment of writing, the field is still very active.

3.3. Tutorials

3.3.1. Problems

Exercise 3.1

Prove that the trace distance between two pure states $|\psi_1\rangle$ and $|\psi_2\rangle$ is given by

$$D(\rho_1, \rho_2) = \sqrt{1 - |\langle\psi_1|\psi_2\rangle|^2} \tag{3.17}$$

Hint: Note that you can always find a basis in which

$$|\psi_1\rangle = c\,|0\rangle + s\,|1\rangle \quad \text{and} \quad |\psi_2\rangle = e^{i\varphi}(c\,|0\rangle - s\,|1\rangle)$$

where $c = \cos\theta$ and $s = \sin\theta$.

Exercise 3.2

A short laser pulse can be sent either at time t_1 or at time t_2. By detecting the time of arrival, one can obviously discriminate between these two cases. This rather trivial process is actually an example of *unambiguous state discrimination* of the two two-mode coherent states $|\psi_1\rangle = |0\rangle\,|\alpha\rangle$ and $|\psi_2\rangle = |\alpha\rangle\,|0\rangle$; it is used to create the raw key in the quantum cryptography protocol called COW [Stucki et al. 2005]. We recall the decomposition of the coherent state $|\alpha\rangle, \alpha \in \mathbb{C}$, on the number basis:

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle \tag{3.18}$$

1. What is the probability of success for optimal USD?
2. Prove that the POVM for optimal USD can simply be realized by detecting the time of arrival (with a perfect detector). Hint: What are the "inconclusive" events?
3. Discuss what happens if the detector is not perfect, in particular how the discussion is modified by (i) efficiency $\eta < 1$; (ii) dark counts.

Exercise 3.3

Let $\{|e_k\rangle\}_{k=1..4}$ be an orthonormal set of four vectors. We define

$$|\psi_1^{\pm}\rangle = \sqrt{1-\varepsilon}\,|e_1\rangle \pm \sqrt{\varepsilon}\,|e_2\rangle \quad \text{and} \quad |\psi_2^{\pm}\rangle = \sqrt{1-\varepsilon}\,|e_3\rangle \pm \sqrt{\varepsilon}\,|e_4\rangle$$

and we construct in turn the mixtures

$$\rho_0 = (1-\varepsilon)\,|\psi_1^+\rangle\langle\psi_1^+| + \varepsilon\,|\psi_2^+\rangle\langle\psi_2^+| \quad \text{and} \quad \rho_1 = (1-\varepsilon)\,|\psi_1^-\rangle\langle\psi_1^-| + \varepsilon\,|\psi_2^-\rangle\langle\psi_2^-|$$

1. Compute the Holevo bound $\chi(\rho_0, \rho_1)$, assuming $p_0 = p_1 = \frac{1}{2}$.
2. The states given above, of course, have a meaning: they describe Eve's states in the optimal eaves- dropping on the BB84 protocol of quantum cryptography, when an error rate ε is measured by Alice and Bob - see other series of lectures; and paragraph III.B.2 and Appendix A of [Scarani et al. 2009]. In this scenario, what does the Holevo bound mean? *Hint:* the index a or the matrices ρ_a in Alice's bit.

3.3.2. Solutions

Exercise 3.1

By writing

$$|\psi_1\rangle = c\,|0\rangle + s\,|1\rangle \quad \text{and} \quad |\psi_2\rangle = e^{i\varphi}(c\,|0\rangle - s\,|1\rangle)$$

we have

$$\langle \psi_1 \,|\, \psi_2 \rangle = e^{i\varphi}(c^2 - s^2) = e^{i\varphi} \cos 2\theta$$

and

$$\rho_1 - \rho_2 = 2cs\sigma_x$$

whence

$$D(\rho_1, \rho_2) = \frac{1}{2}(| + 2cs| + | - 2cs|) = 2cs = \sin 2\theta$$

The result follows immediately.

Exercise 3.2

1. The probability of optimal USD is

   $$p_{USD} = 1 - |\,\langle \psi_1 \,|\, \psi_2 \rangle\,|$$

   In this case, we have

   $$\langle \psi_1 \,|\, \psi_2 \rangle = |\,\langle 0 \,|\, \alpha \rangle\,|^2 = e^{-|\alpha|^2}$$

2. As soon as the detector fires, the two states can be distinguished; so the "inconclusive" events are the events in which the detector did not fire; if the detector has perfect efficiency, this can only happen because of the vacuum component of the state. In both $|\psi_1\rangle$ and $|\psi_2\rangle$, the amplitude of the vacuum component is $e^{-|\alpha|^2/2}$; therefore the probability that the detector fires is $1 - e^{-|\alpha|^2} = p_{USD}$.

3. A detector with efficiency $\eta < 1$ is equivalent to losses $\sqrt{\eta}$; the discrimination is still unambiguous but succeeds only with probability $p = 1 - e^{-\eta|\alpha|^2} < p_{USD}$ (note that this is still the optimal procedure, under the constraint that one has to use such imperfect detectors). If dark counts are present, the detector may fire even if there was no photon; therefore the discrimination is no longer unambiguous

Exercise 3.3

1. We have to compute

   $$\chi(\rho_0, \rho_1) = S(\rho) - \frac{1}{2}[S(\rho_0) + S(\rho_1)] \text{ with } \rho = \frac{1}{2}(\rho_0 + \rho_1)$$

Now ρ_0 and ρ_1 are both incoherent mixtures of two orthogonal states with the same weights; therefore

$$S(\rho_0) = S(\rho_1) = -(1-\varepsilon)\log(1-\varepsilon) - \varepsilon\log\varepsilon \equiv h(\varepsilon)$$

Moreover

$$\rho = (1-\varepsilon)^2 |e_1\rangle\langle e_1| + \varepsilon(1-\varepsilon)|e_2\rangle\langle e_2| + \varepsilon(1-\varepsilon)|e_3\rangle\langle e_3| + \varepsilon^2 |e_4\rangle\langle e_4|$$

whence $S(\rho) = 2h(\varepsilon)$. All in all, $\chi(\rho_0, \rho_1) = h(\varepsilon)$.

2. One can see the relation between Alice and Eve as a channel, in which Alice's bit value a has been encoded in a state ρ_a. Therefore, the Holevo bound represents the maximal information that Eve might extract about Alice's bit.

4. Lecture Quantum correlations (I): the failure of alternative descriptions

This and the following two lectures are devoted to the violation of Bell's inequalities and related topics. This means that we shall focus on a restricted family of quantum phenomena: the establishment of *correlations between distant partners through separated measurement of entangled particles*. This kind of phenomenon by no means exhausts the possibilities of quantum physics. However, it is there that the discrepancy between classical and quantum physics manifests itself in the most straightforward way (this Lecture). As such, one might expect this feature to be useful for some quantum information tasks: this is indeed the case, but curiously enough, this awareness is only rather recent (Lecture 6). Lecture 5 will be an excursion into an extended theoretical framework that may be very promising or may just be a wrong track, but is funny and worth at least having heard about.

4.1. Correlations at a distance

4.1.1. Generic setup under study

The kind of experiment we are considering is sketched in Fig. 4.1.

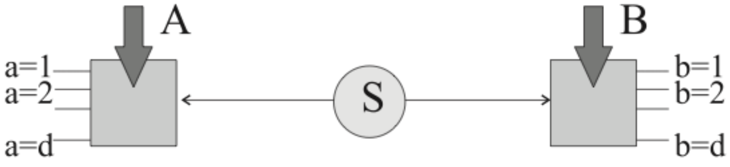

Figure 4.1: The setup to be kept in mind for Bell-type experiments: a source S distributed two quantum systems to separated locations. On each location, the physicist is free to choose which measurement to perform (A, B); as a result, they obtain outcomes a, b. Of course, this setup can be generalized to more parties, or to the case where the number of outcomes is different between the parties.

A source sends out two particle to two distant locations. In each location, a user chooses a possible measurement (A, B) and registers the outcome (a, b). The procedure is repeated a large number of times. Later, the two users come together, compare their results and derive the probability distribution

$$P_{AB}(a, b) \tag{4.1}$$

This probability distribution is often written as a conditional probability $P(a, b|A, B)$. Ultimately, it?s a choice of notation that matters little; I use the notation (4.1) to stress that the origin of the statistics on (a, b) is the quantum randomness we want to query, while the statistics on (A, B) are of a very different nature (just the choice of the users on how often to perform each measurement).

A few crucial remarks:

(i) There is a conceptual distinction between the *users* and the physicists who have constructed the whole setup. To perform the task, the users do not even have to know what state the source has prepared, what physical systems are sent, which measurements are being performed. For the users, the measurement is just a knob they can freely set at any position; A and B refer to the label of the scale in their instruments. In what follows, the names of Alice and Bob will be always referring to the users.

(ii) The number of possible outcomes a and b plays an important role, but the labeling of the outcomes does not matter: the probability distribution $P(a, b)$ fully characterizes the process of measuring A and B, irrespective of whether the outcomes are real numbers, multiples of $\hbar$, complex numbers, colors... Because of this, we are free to choose the most convenient labeling without loss of generality. For instance, if the number of outcomes is two on both sides, then the correlation coefficient is defined as

$$E_{AB} = P_{AB}(a = b) - P_{AB}(a \neq b) \tag{4.2}$$

This definition is unambiguous and always valid. If in addition, one chooses the labeling $a, b \in \{-1, +1\}$, then one obtains the handy relation $E_{AB} = \langle ab \rangle_{AB}$ the average of the product of the outcomes.

4.1.2. Classical mechanisms for correlations

We are dealing with a family of probability distributions: Alice picks up measurement A, Bob picks up measurement B, and the outcomes are guaranteed to be distributed according to the probability distribution P_{AB}. Now, in general, $P_{AB}(a, b) \neq P_A(a)P_B(b)$ where $P_A(a) = \sum_b P_{AB}(a, b)$ and $P_B(b) = \sum_a P_{AB}(a, b)$ are the *marginal distributions.* In other words, random variables distributed according to P_{AB} are correlated. The question that we are going to study is: can these correlations be ascribed to a classical mechanism? The question seems vague, but it becomes less vague once one realizes that there are only *two* classical mechanisms for distributing correlations.

The first and most obvious is *communication*: for instance, the information about Alice's choice of measurement A is available to the particle measured by Bob. This mechanism can be checked by arranging $space-likeseparation$: loosely speaking, if the two particles reach the measurement devices at the same time and the choice of measurement is done at the very last moment, a signal informing a particle about what is happening in the other location should travel faster than light. If the correlations persist in this configuration, the mechanism of communication becomes highly problematic.

The second mechanism consists in using *pre-established strategies*: each particle might have left the source with a set of instructions, specifying how it should behave in any measurement. Interestingly, this mechanism can explain the behavior of the singlet state in the cases where Alice and Bob choose to perform the same measurement (P_{MM}). Indeed, assume that the particles are carrying lists of pre-determined results $\lambda_A = \{a_M\}_M$ and $\lambda_B = \{b_M\}_M$ such that $a_M = -b_M$ for each measurement M: one always gets $P_{MM}(a \neq b) = 1$, and the local randomness can be easily accounted for by varying the lists at each run. However, it is the milestone result by John Bell in 1964 to prove that the *whole* family of probabilities predicted by quantum physics cannot be reproduced with pre-established strategies [Bell 1964].

We embark now on a more detailed study of each of the two classical mechanisms and their failure to reproduce the results observed in experiments.

4.2. Pre-established strategies ("Local variables")

4.2.1. The model

By definition, a pre-established strategy is some hypothetical information λ that the particles carry out with themselves from the source. Each particle is supposed to produce its outcome taking into account only this λ and the measurement to which it is submitted (plus some possible information encountered along the path, that we neglect for simplicity here). In other words, for given λ, the two random processes are supposed to be independent: $P_{AB}(a,b|\lambda) = P_A(a,|\lambda)P_B(b|\lambda)$. The only freedom left is the possibility of changing the information λ at each run. If λ is drawn from a distribution $\rho(\lambda)$, the observed probability distribution will be

$$P_{AB}(a,b|\lambda) = \int d\lambda\, \rho(\lambda) P_A(a,|\lambda) P_B(b|\lambda) \qquad (4.3)$$

Here comes an important mathematical characterization:

Theorem. $P_{AB}(a,b|\lambda)$ can be obtained by pre-established strategies if and only if it can be written as a convex sum of local deterministic strategies.

A *deterministic strategy* is a strategy in which, for each possible measurement, the result is determined. A *local* deterministic strategy is defined by

$$P_A(a,|\lambda) = \delta_{a=f(A,\lambda)} \text{ and } P_B(b|\lambda) = \delta_{b=g(B,\lambda)}$$

The "if" implication is therefore trivial. The "only if" implication stems from the fact that any classical random process can be mathematically decomposed as a convex sum of deterministic processes. In other words, for each λ, there exists an additional random variable $\mu = \mu(\lambda)$ with distribution $\rho'(\mu)$ such that

$$P_A(a, |\lambda) = \int d\mu\, \rho'(\mu)\delta_{a=f(A,\lambda,\mu)}$$

One can therefore just "enlarge" the definition of the local variable to $\lambda' = (\lambda, \mu(\lambda))$.

Note that no restriction of the family of pre-established strategies under study is being made. The theorem is a purely mathematical result. It is useful because it allows deriving results by arguing with deterministic strategies, that are very easy to handle; if the result is stable under convex combination, it is automatically guaranteed to hold for *all* possible pre-established strategies, deterministic or not.

4.2.2. Two remarks

Before continuing, it is worth while making two remarks.

About terminology

Historically, λ was called "local hidden variable". While the expression "local variable" is ultimately convenient, the adjective "hidden" is definitely superfluous and even misleading: quantum physics is at odds with local variables, irrespective of whether they are supposed to be hidden or not. For instance, the local variable may be a description of the total quantum state [Gisin 2009]. Much more recently, in the interaction between physicists and computer scientists, the name of "shared randomness" has also become fashionable to denote pre-established strategies.

Regarding the fact itself, that quantum correlations cannot be attributed to pre-established strategies and local parameters, the linguistic debate is even more involved. Maybe the most precise expression would be "falsification of crypto-determinism", as proposed e.g., by Asher Peres [Peres 1995]; but it is hardly used. The two most common expressions found in the literature are "quantum non-locality" and "violation of local realism". Both have their shortcomings, some people in the field have rather strong feeling against one or the other[11]. You must above all keep in mind that there are plenty of

[11] For instance, in the language of quantum field theory (which historically pre-dates Bell's inequalities), "non-locality" had been used to mean what we call here "signaling": if

unfortunate expressions in science - the essence of "quantum" physics is the superposition of states, certainly not the discreteness! - but there is little danger in using them, as long as one knows what their meaning is. I shall use "non-locality", without the "quantum", and refer to "non-local correlations" as a shortcut for "probability distributions that cannot be reproduced by pre-established agreement".

Pre-established values for single systems

If quantum systems are considered as a whole, i.e., not consisting of separate sub-systems, one can always find a local variable model that reproduces the correlations. or dimensions larger than two, the Kochen-Specker theorem proves that the local variable model must take into account the whole measurement (it must be "contextual"); but contextuality is not a problem for local variable theories.

Though its interest may be rather limited, for the sake of illustration, let me introduce a well-known explicit local variable model that reproduces the quantum predictions for one qubit. The goal is to reproduce the statistics of the quantum state

$$\rho = \frac{1}{2}(\mathbb{1} + \vec{m}\cdot\vec{\sigma})$$

under all possible von Neumann measurements, i.e.,

$$P(\pm\vec{a}) = \frac{1}{2}(\mathbb{1} + \vec{m}\cdot\vec{a})$$

The local variable is a vector

$$\vec{\lambda} = [\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta]$$

on the surface of the unit sphere; for each realization, $\vec{\lambda}$ is drawn randomly with uniform measure. The rule for the outcome of a measurement is deterministic: the outcome is

$$r(\vec{a},\vec{\lambda}) = \mathrm{sign}[(\vec{m}-\vec{\lambda})\cdot\vec{a}]$$

you stick to this meaning, quantum theory and quantum field theory are "local", because they are no-signaling. As another example, I tend not to like "violation of local realism", because in philosophy "realism" is the school of thought that accepts that there is something outside our brain and we can have some convenient knowledge about it. In particular then, a "philosophical realist" is someone who accepts that "local realism is violated".

Then one can prove that

$$\langle r\rangle(\vec{a}) = \int_{S^2} r(\vec{a},\vec{\lambda})\sin\theta d\theta d\phi = \vec{m}\cdot\vec{a}$$

which is exactly the quantum expectation value (the proof is simple using a geometrical visualization: the integral is the difference between the surface of two spherical hulls).

4.2.3. CHSH inequality: derivation

Let us now present the derivation of the most famous Bell inequality, the one derived by Clauser, Horne, Shimony and Holt and therefore known as CHSH [Clauser et al., 1969]. This inequality is defined by the fact that both Alice and Bob can make only two possible measurements (we label them A, A' for Alice, B, B' for Bob) and the outcomes are binary (we choose the labeling $a, b \in \{+1, -1\}$)

Let us first consider a local deterministic strategy λ_D: here, it is just any list $\lambda_D = (a, a', b, b')$ specifying the two outcomes of Alice and the two outcomes of Bob. If this list is defined, then the number

$$S(\lambda_D) = (a + a')b + (a - a')b'$$

is also defined. By inspection, it is obvious that $S(\lambda_D)$ can only take the values +2 or −2. If we now take a convex combination of such strategies with distribution ρ, it is obvious that

$$\langle S\rangle = \int d\lambda_D \rho(\lambda_D) S(\lambda_D)$$

must lie between −2 and +2. Moreover, since the average of a sum is the sum of the averages, and since with our labeling $\langle ab\rangle$ is the correlation coefficient, we have found

$$|\langle S\rangle| = |E_{AB} + E_{A'B} + E_{AB'} - E_{A'B'}| \le 2 \tag{4.4}$$

This is the *CHSH inequality*: it holds for all convex combinations of local deterministic strategies and therefore, by virtue of the theorem above, it holds for all pre-established strategies. An important remark here: the derivation of the inequality is particularly simple when the outcomes are labeled +1 and −1; this is how we did it, and we shall keep this convention in this whole lecture (for the topic of the next lecture, another labeling will prove more convenient). However, the inequality itself is independent of this choice: if one chooses another labeling, the inequality still holds with the general definition (4.2).

4.2.4. CHSH inequality: violation in quantum physics

In quantum physics, a measurement that can give two outcomes, +1 and −1, is described by an Hermitian operator with those eigenvalues (possibly degenerate). Therefore

$$E_{AB} \to \langle A \otimes B \rangle \tag{4.5}$$

where A and B are two such operators. In other words, in quantum physics $\langle S \rangle$ becomes the expectation value of the *CHSH operator*

$$\mathcal{S} = A \otimes B + A' \otimes B + A \otimes B' - A' \otimes B' \tag{4.6}$$

The largest possible value of $\langle S \rangle$ is therefore[12] the largest eigenvalue of $\mathcal{S}$. As noted by Tsirelson[13], in the case of the CHSH inequality it is easy to find a bound for the maximal eigenvalue [Cirel'son 1980]. Indeed, using the fact that $A^2 = A'^2 = B^2 = B'^2 = \mathbb{1}$, one can easily compute

$$\mathcal{S}^2 = 4\mathbb{1} \otimes \mathbb{1} + [A, A'] \otimes [B, B']$$

The maximal eigenvalue of $[A, A']$ cannot exceed 2, because

$$|\langle [A, A'] \rangle| \leq |\langle AA' \rangle| + |\langle A'A \rangle|$$

and the spectrum of both A and A' contains only +1 and −1. So the maximal eigenvalue of $\mathcal{S}^2$ cannot exceed 8, which implies that in quantum physics

$$|\langle S \rangle| \leq 2\sqrt{2} \tag{4.7}$$

This bound can be saturated already with two-qubit states. Indeed, consider the correlations of the singlet state

$$E_{AB} = -\vec{a} \cdot \vec{b}$$

[12] We are using the well-known fact that the maximal eigenvalue is equal to the largest average value over the set of possible states.

[13] This author used to spell his name as Cirel'son until the mid-eighties.

by choosing

$$\vec{a} = \hat{z}\ ,\ \vec{a}' = \hat{x}\ ,\ \vec{b} = \frac{1}{\sqrt{2}}(\hat{z}+\hat{x})\ ,\ \vec{b}' = \frac{1}{\sqrt{2}}(\hat{z}-\hat{x})$$

one obtains

$$E_{AB} = E_{A'B} = E_{AB'} = -E_{A'B'} = -\frac{1}{\sqrt{2}}$$

whence

$$|\langle S\rangle| \leq 2\sqrt{2}$$

Two important remarks:

- It is remarkable that local variables can be ruled out already in the *simplest* scenario: with the smallest Hilbert space that allows entanglement and with the least possible number of measurements and of outcomes. Indeed, if a party performs only one measurement, one can always find a local variable model that explains the statistics; and a measurement with only one outcome would be obviously trivial.

- As we have just seen, there is not a unique Bell operator, but a family parametrized by the measurement settings. It is trivial to find "bad settings", that don't give any violation even if the state is maximally entangled: for instance, remember that the correlations of the singlet cannot violate any Bell inequality if one requests Alice and Bob to perform the same measurements, i.e., if $A = B, A' = B'$.

4.2.5. Ask Nature: experiments and loopholes

Experiments are reviewed in other series of lectures in this school; comprehensive review articles are also available [Tittel and Weihs 2001, Pan et al. 2008]. Here I just address the following question: to which extent alternative models have been falsified.

The non-locality of quantum correlations is so striking an effect, that it has been scrutinized very closely. Basically two possible *loopholes* have been identified:

- If one does not arrange the timing properly, the detections may be attributed to a sub-luminal signal: this is called *locality loophole*. In order to close this loophole, the events must be ordered in space-time as in Fig. 4.2.

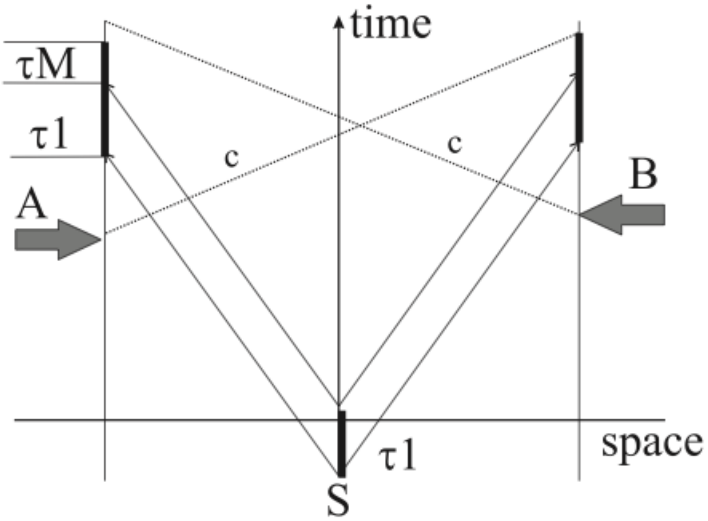

Figure 4.2: Space-time diagram to close the locality loophole: the measurement A must be chosen at a time, such that this information reaches the location of B after the measurement on particle B is completed; and symmetrically. The dotted lines denoted by c represent the light cone in vacuum; the particles, even when they are photons, may propagate at a slower speed (e.g., if they are sent through optical fibers). The emission of a pair happens with some uncertainty τ_1 called single-particle coherence time. The parameter τ_M indicates the time it takes to perform a measurement, i.e., to "collapse" the state: a very problematic notion in this discussion, see text.

Note in particular that the choice of setting on Alice's side must be space-like separated from the end of the measurement on Bob's side, and vice-versa. Now, the "end of a measurement" is one of the most fuzzy notions in quantum theory! Consider a photon impinging on a detector: when does quantum coherence leaves place to classical results? Already when the photon generates the first photo-electron? Or when an avalanche of photo-electrons is produced? Or when the result is registered in a computer? There is even an interpretation (Everett's, also called many-worlds) in which no measurement ever happens, the whole evolution of the universe being just a developing of quantum entanglements. All these options are compatible with our current understanding and practice of quantum theory (a fact expressed sometimes by saying that there is a "quantum measurement problem"). As long as this is the situation, strictly speaking it is impossible to close the locality loophole. However, many physicists adopt the reasonable assumption that the measurement is finished "not too long time" after the particle impinges on the detector[14] and I shall adopt such a position below. In this view, τ_M is of the order of the microsecond, in which case the distance between Alice and Bob should at least be 300m.

- The other possible problem is called *detection loophole*. In all experiments, the violation of Bell's inequalities is measured on the events in which both particles have been detected. Since detectors don?t have perfect efficiency, there is also a large number of events in which only one particle is detected, while the other has been missed. The detection loophole assumes a form of conspiracy, in which the undetected particle "chose" not to be detected after learning to which measurement it was being submitted. In this scenario, if the detection efficiency is not too high, it is pretty simple to produce an apparent violation Bell's inequalities with local variables: when the local variable does not have the desired value, the particle opts not to reply altogether! For CHSH and maximally entangled states, the threshold value of the detection efficiency required to close this loophole is around 80% (see Tutorial).

At the moment of writing, no experiment has closed both loopholes simultaneously[15]; some groups are aiming at it, but the requirements are really demanding. Is it really worth while? After all, all experimental data are perfectly described by our current physical theory and are in agreement with our understanding of how devices (e.g. detectors) work. By the *usual* standards in physics, the violation of Bell's inequalities by separated entangled pairs is experimentally established beyond doubt. Given the implications of this statement for our world view, though, some people think that *much higher standards* should be applied in this case before making a final claim. I leave it to the readers to choose their camp. I just note here that the detection

[14] Some theorists having speculated on a link between gravitation and "collapse", a recent experiment has bothered connecting the detectors to a piezo-electric device: each time a photon impinges is detected, in addition to the signal being registered by a computer, a "large" mass (a few grams) is set into motion [Salart, Baas, van Houwelingen et al. 2008]. If those theoretical speculations have a foundation in nature, this experiment may be extremely meaningful. If not... it had at least the merit of raising the awareness.

[15] The locality loophole can be closed (and has been closed) in experiments with photons; but typical detection efficiencies, including losses, are far from reaching the threshold that would close the detection loophole. In turn, this loophole has been closed in experiments with trapped ions and atoms; but the micrometric separation and long detection times definitely prevent to exclude some sub-luminal communication.

loophole, so artificial in this context, has recently acquired a much more serious status in a different kind of problems: we shall have to study it in detail in Lecture 6.

4.2.6. Entanglement and non-locality

The literature devoted to Bell's inequalities is huge. All possible generalizations of the CHSH inequality have been presented: more than two measurements, more than two outcomes, more than two particles, and all possible combinations. There are still several open questions [Gisin 2007, Scarani and Méthot 2007]. Just to give an idea of this complexity, let me review rapidly the development of one of the main questions: *do all entangled states violate a Bell inequality?* This seems an obvious question, but as we shall see, the answers are rather intricate.

The easy part is the following: *all pure entangled states violate some Bell-type inequality.* This statement is known under the name of *Gisin's theorem*: indeed Gisin seems to be the first to have addressed the question, and he proved the statement for bipartite states [Gisin 1991]; shortly afterwards, Popescu and Rohrlich provided the extension to multipartite states[16] [Popescu and Rohrlich 1992].

Problems begin when one considers *mixed states.* Even before Gisin proved his theorem, Werner had already shown that there exist mixed states such that (i) they are definitely entangled, in the usual sense that any decomposition as a mixture of pure states must contain some entangled state; (ii) the statistics obtained for all possible von Neumann measurements can be reproduced with pre-established strategies [Werner 1998]. Several years later, Barrett showed that the result holds true for the same states even when

[16] The scheme of Popescu and Rohrlich goes as follows: if the system consists of N particles, measurements are performed on $N-2$ of them and the results are communicated to the last two measuring stations: conditioned on this knowledge, the two particles are now in a well-defined bipartite state, so Gisin's original theorem applies. Though it uses communication, the scheme is perfectly valid because the last two parties do *not* communicate with each other: for them, the information communicated by the others acts as pre-established information, which cannot create a violation of Bell's inequalities. For a reason unknown to me, a quite large set of people working in the field are convinced that Gisin's theorem has not been proved in general. Admittedly, for many cases we may not know any compact and elegant inequality that is violated by all pure states; but the scenario of Popescu and Rohrlich does prove the statement in general.

POVMs are taken into account [Barrett 2002]. So *there exist mixed entangled states that do not violate any Bell-type inequality.*

For two qubits, the one-parameter family of Werner states is

$$\rho_W = W\,|\Psi^-\rangle\,\langle\Psi^-| + (1-W)\frac{\mathbb{1}}{4} \tag{4.8}$$

with $0 \leq W \leq 1$. These states are entangled for $W > \frac{1}{3}$. The best extension of Werner's result shows that the statistics of von Neumann measurements are reproducible with local variables for $W \lesssim 0.6595$ [Aćin, Toner and Gisin 2006]. On the other hand, Werner states provably violate some Bell inequality as soon as $W \gtrsim 0.7056$ [Vértesi 2008]. Nobody has been able to close the gap at the moment of writing: this is an open problem, that one may legitimately consider of moderate interest, but that is frustrating in its apparent simplicity[17].

Now, having learned about *distillation of entanglement*, you may legitimately be surprised by Werner's result. Indeed, Werner's states are distillable: why can't one just distill enough entanglement to violate a Bell's inequality? Of course, one can! The previous discussion has been made under the tacit assumption that Alice and Bob do not perform collective measurements, but measure each of their particles individually. This is the way experiments are usually done, but is not the most general scenario allowed by quantum physics. Distillation involves collective measurements, so there is no contradiction between the distillability of Werner's state and Werner's result on local variable models. Actually, it is worth while stressing that the first distillation protocol was invented by Popescu *precisely* in order to extract a violation of Bell's inequalities out of "local" Werner states [Popescu 1995].

It has recently been shown that *all entangled mixed states (including bound-entangled ones), if submitted to suitable multi-copy processing, can produce statistics that violate some Bell-type inequality* [Masanes et al. 2008]. This is somehow a comforting result: in this very general scenario, entanglement and the impossibility of pre-established strategies coincide. Still, in my opinion,

[17] For the anecdote, let us stress that for many years the known bound for violation was $W > 1/\sqrt{2} \sim 0.7071$; this is the bound directly obtained using CHSH. In order to obtain the minor improvement reported above, namely 0.7056, one needs inequalities that use at least 465 settings per party!

Werner's result and all its extensions keep their astonishing character: each pair being entangled and produced independently of the others, why does it take complex collective measurements in order to reveal the non-classicality of the source?

4.3. Superluminal communication

4.3.1. General considerations

The issue of superluminal communication as a possible explanation for quantum correlations is delicate. Several physicists think that it is just not worth while addressing in the first place. While I am definitely convinced that no superluminal communication is indeed going on, instead of shunning this topic completely, I prefer to adopt a more pragmatic view based on the following considerations: while local variables have been directly and conclusively disproved as a possible mechanism, superluminal communication seems to be excluded "only" on the belief that nothing, really nothing, should propagate faster than light. Now, strictly speaking, relativity forbids faster-than-light propagation of signals that can be used by us (because this would open causality loops and allow signaling in the past). And, as far as I know, nobody has really proved that all possible models based on communication are intrinsically inconsistent - some serious authors have actually argued the opposite [Reuse 1984, Caban and Rembielinski 1999]. Therefore, I think it is worth while trying to invent such models and falsify them in experiments.

Still, the family of such models seems to be much more complex than the clear-cut definition of pre-established strategies. Here are a few elements to be taken into account:

- First of all, when dealing with a hypothetical signal traveling faster than light, one has to decide *in which frame this communication is defined*. There are basically two alternatives: either one considers a *global preferred frame*; or one envisages *separate preferred frames for each particle*. Both alternatives have been explored.

- By admission of its very founders, Bohmian mechanics can be seen as a theory in which information (the deformation of the quantum potential due to a measurement) propagates at infinite speed in a global preferred frame [Bohm and Hiley 1993]. Now, however problematic the full Bohmian program may be, Bohmian mechanics is mathematically equivalent to quantum mechanics. Therefore, we know that there is at least one model with superluminal communication that cannot be checked against quantum predictions, because its predictions are exactly the same! In other words, there cannot be an analog of Bell's theorem ruling out all possible models with communication.

- When constructing a model, it is customary to enforce the fact that the users should not be able to signal faster than light (i.e., the fact that superluminal communication must be "hidden"). The models we shall review here are meant to be of this kind; note however that, as soon as one is willing to introduce a preferred frame in physics, there is no compelling reason to enforce no-signaling [Eberhard 1989].

4.3.2. Bounds on the speed in a preferred frame

Let us first suppose that the hypothetical superluminal communication is defined in a *preferred frame*. One has to arrange the detection events to be as simultaneous as possible in this frame; upon observing that the correlations persist, one obtains a lower bound on the speed of this hypothetical communication. Now, we don't know any preferred frame, so which one should one choose? It can be anything, from the local rest frame of the town to the frame in which the cosmic background radiation has no Doppler shift.

It turns out that one does not really have to choose! Indeed, suppose one arranges simultaneity in the rest frame of the town: by virtue of Lorentz transformation[18], simultaneity is automatically guaranteed in all those frames whose relative speed is orthogonal to the direction A-B. If in addition one arranges the line A-B in the East-West orientation, the rotation of the Earth will scan *all* possible frames in twelve hours!

When experimental data are analyzed, the bounds are striking: the hypothetical communication should travel at speeds that exceed 10000c [Scarani et al. 2000, Salart, Baas, Branciard et al. 2008]! This is a very strong suggestion that the mechanism of communication in a preferred frame is not a valid explanation for quantum correlations.

[18] We assume that standard Lorentz transformation applies to the description of classical events, such as detection; we refer to the discussion above for the problem of defining classical events at all.

4.3.3. Before-before arrangement in the local frames

It seems that the previous discussion settles the problem. However, we have another alternative thanks to the inventiveness of Antoine Suarez. This is sometimes known as Suarez-Scarani model, because I happened to help Antoine in formalizing his intuition[19]. The idea is that there may not be a unique preferred frame: rather, each particle, upon detection, would send out superluminal signals in the rest frame of the measurement device [Suarez and Scarani 1997].

Again, one has to define which is the meaningful measurement device: the choice of the setting, the detector... Whichever one chooses though, how does one falsify such a model? By arranging a *before-before* timing. Consider the space-time diagram sketched in Fig. 4.3, representing the situation in which the two measurement devices move away from one another: in that arrangement, each particle arrives before the other one at its respective device! If the model is correct, in this situation one should cease to violate Bell's inequalities. Experiments have been performed and the violation does not disappear [Zbinden et al. 2001, Stefanov et al. 2002, Stefanov et al. 2003]: the Suarez-Scarani model adds to the long list of falsified alternative descriptions of quantum phenomena.

4.4. Leggett's model

Let's leave aside the details and admit the failure of the two classical explanations for correlations: some- thing "non-classical" is present. Still, one may try to save as much as possible of a classical world-view. This is a quite general idea but, as a matter of fact, only one specific model inspired by this line of thought has been proposed so far. This first such proposal is due to Leggett [Leggett 2003], whose initial intuition was rediscovered by the Vienna group [Gröblacher et al. 2007] and has undergone further clarification [Branciard et al. 2008].

Recall that one of the astonishing aspects of entanglement is the fact that a composite system can be in an overall pure state, while none of its components is. One can try to check this statement directly. The problem can be

[19] Some years later, I realized that this formalized version was not very brilliant after all: if one tries to extend it to three particles, it leads to signaling [Scarani and Gisin 2002]! But by then, experiments had already falsified the model anyway [Zbinden et al. 2001].

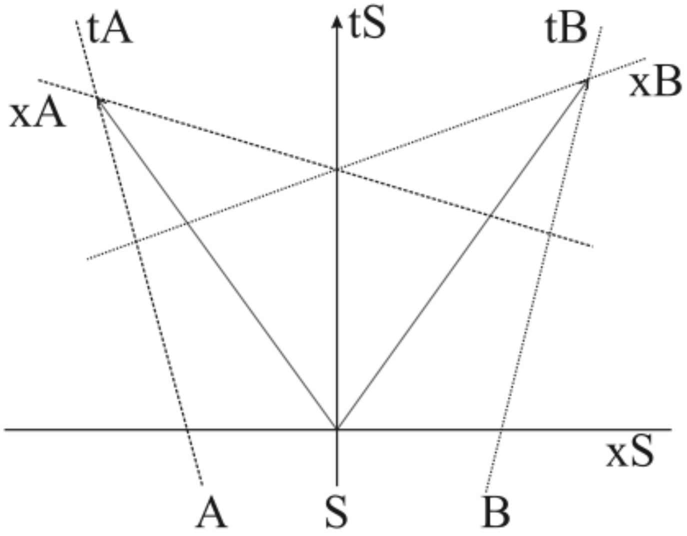

Figure 4.3: Space-time diagram for the before-before experiment (coherence times have been omitted for simplicity, of course they must be taken into account in a detailed analysis). When particle A is detected, in the reference frame of its device particle B had not been detected yet; a symmetrically.

formulated as follows: whether it is possible to find a decomposition

$$P_{AB}(a,b) = \int d\xi \, P_{AB}(a,b|\xi) \tag{4.9}$$

such that $P_{AB}(a,b|\xi)$ is no-signaling but describes correlations that violate Bell's inequalities (the necessary "non-classical" element) but the marginals are like those of a pure, single-particle state. If this is possible, one could have sharp properties for both the composite and the individual systems.

Without entering the details of the derivation, a decomposition like (4.9) fails to recover the quantum probabilities. Experiments have been performed, whose results are in excellent agreement with the latter and falsify therefore the alternative model. In addition, one can prove that, to reproduce the correlations of the singlet state, the marginal of all the $P_{AB}(a,b|\xi)$ cannot be anything else than completely random.

From Leggett's model we have learned that not only the full classical mechanisms are ruled out, but also attempts at sneaking back some elements of classicality in an otherwise non-classical theory seem destined to fail. I dare say that this is another remarkable proof of the “structural accuracy” of quantum theory: not only, as widely known, this theory is most powerful in its numerical predictions: its structure itself has uncovered properties of reality that we could not have imagined otherwise.

4.5. Balance

Let us summarize the extent of the failure of alternative descriptions to quantum physics, inspired by the phenomenon of quantum correlations at a distance:

Pre-established agreement: All possible models based on pre-established agreement (local variables) are incompatible with the quantum predictions. All the experiments are in excellent agreement with the latter; there is a universal consensus that the remaining loopholes are technical issues whose closure is only a matter of time and skill. *Superluminal communication*: No direct check of all models seems to be possible, and at least one of them is known to be compatible with all quantum predictions. However, all the models that have been tested, together with the implausibility of the assumption itself, clearly uphold the view that no communication is involved in the establishment of quantum correlations (in other words, that there is no *time-ordering* between the correlated events).

There is no third classical way: if one considers (as I do) that we have enough evidence to rule out both pre-established agreement and communication, the conclusion is that *no mechanism in space and time* can reproduce quantum correlations. These correlations “happen”, we can predict them, but they are a fundamental, irreducible phenomenon. We don?t have an explanation in terms of more basic phenomena, and, most remarkably, *we know we shall never find one.*

4.6. Tutorials

4.6.1. Problems

Exercise 4.1

This calculations shows that all pure entangled states of two qubits violate the CHSH inequality [Gisin 1998]. Using the Schmidt decomposition, any two-qubit pure state can be written, in a suitable basis, as

$$|\psi(\theta)\rangle = \cos\theta\,|00\rangle + \sin\theta\,|11\rangle$$

with $\cos\theta \geq \sin\theta \geq 0$.

1. Write down the Bell-CHSH operator $\mathcal{S}$ for
$$A = \sigma_z\ ,\ A' = \sigma_x\ ,\ B = \cos\beta\sigma_z + \sin\beta\sigma_x \text{ and } B' = \cos\beta\sigma_z - \sin\beta\sigma_x$$
2. Compute $S(\theta|\beta) = \langle\psi(\theta)|\,\mathcal{S}\,|\psi(\theta)\rangle$, then choose the settings of Bob to achieve the optimal violation
$$S(\theta) = \max_\beta S(\theta|\beta) = 2\sqrt{1+\sin^2 2\theta}$$

Note: it can be proved that the family of settings above is optimal for the family of states under study. Therefore $S(\theta)$ is the maximal violation of CHSH achievable with the state $|\psi(\theta)\rangle$. For CHSH and two qubits, the optimal violation is also known for mixed states [Horodecki et al. 1995].
Exercise 4.2

In this exercise, we study the Greenberger-Horne-Zeilinger (GHZ) argument for non-locality [Greenberger et al. 1989, Mermin 1990]. Consider the three-qubit state

$$|GHZ\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle) \tag{4.10}$$

Compute the four expectation values

$$\langle\sigma_x\otimes\sigma_x\otimes\sigma_x\rangle\,,\ \langle\sigma_x\otimes\sigma_y\otimes\sigma_y\rangle\,,\ \langle\sigma_y\otimes\sigma_x\otimes\sigma_y\rangle\,,\ \langle\sigma_y\otimes\sigma_y\otimes\sigma_x\rangle$$

on this state. Try to find a local deterministic strategy that would have such correlations and obtain a contradiction.

4.6.2. Solutions

Exercise 4.1

Inserting the given settings in the definition of the Bell-CHSH operator, one finds

$$\mathcal{S} = 2\cos\beta\sigma_z\otimes\sigma_z + 2\sin\beta\sigma_x\otimes\sigma_x \tag{4.11}$$

whence

$$\langle\psi(\theta)|\,\mathcal{S}\,|\psi(\theta)\rangle = \cos\beta + \sin\beta\sin 2\theta$$

As is well-known,

$$\max_x(a\cos x + b\sin x) = \sqrt{a^2+b^2}$$

is obtained for

$$\cos x = \frac{a}{\sqrt{a^2+b^2}}$$

therefore

$$S(\theta) = 2\sqrt{1+\sin^2 2\theta} \;\text{ for }\; \cos\beta = \frac{1}{\sqrt{1+\sin^2 2\theta}} \tag{4.12}$$

Note that $S(\theta)$ is always larger than 2, apart from the case $\cos\theta = 0$ of a product state. **Exercise 4.2**

For the GHZ state, one has

$$\begin{aligned} \langle\sigma_x\otimes\sigma_x\otimes\sigma_x\rangle &= 1\\ \langle\sigma_x\otimes\sigma_y\otimes\sigma_y\rangle &= -1\\ \langle\sigma_y\otimes\sigma_x\otimes\sigma_y\rangle &= -1\\ \langle\sigma_y\otimes\sigma_y\otimes\sigma_x\rangle &= -1 \end{aligned} \tag{4.13}$$

i.e., the state is an eigenvector of those operators. Explicitly, the first line means that only four possible triples of results are possible when all particles are measured in the x direction: $(+,+,+),(+,-,-),(-,+,-)$ and $(-,-,+)$; in the other three cases, the other four triples are possible.

In a local deterministic strategy, all the outcomes should be known in advance, so we want to find six numbers

$$\{a_X, a_Y,;b_X,b_Y;c_X,C_Y\} \in \{+1,-1\}^6$$

such that

$$\begin{aligned} a_X b_X C_X &= 1\\ a_X b_Y C_Y &= -1\\ a_Y b_X C_Y &= -1\\ a_Y b_Y c_X &= -1 \end{aligned} \tag{4.14}$$

These relations are however obviously impossible to satisfy (notice for instance that the product of the four lines gives $+1 = -1$). So there is no deterministic point satisfying the correlations of the GHZ state. Since these are perfect correlations, no additional element of randomness can create them: in other words, no convex combination of deterministic points can satisfy the correlations either.

5. Lecture Quantum correlations (II): the mathematics of no-signaling

5.1. The question of Popescu and Rohrlich

We saw in the first lecture that the present definition of quantum physics is rather a description of its formalism. Still, there is no shortage of "typically quantum features": intrinsic randomness, incompatible measurements and uncertainty relations, no-cloning, teleportation, non-local correlations... Can't one find the physical definition of quantum physics among those?

Popsecu and Rohrlich asked this question for *non-locality without signaling* [Popescu and Rohrlich 1994]: does quantum physics define the set of all probability distributions that are possibly non-local but are compatible with the no-signaling condition? The answer is: NO. Consider the following probability distribution for binary input $A, B \in \{0,1\}$ and binary output $a, b \in \{0,1\}$ (compared to the previous lecture, we use again our freedom of choosing the labels in the most convenient way):

$$P_{AB}(a \oplus b = AB) = 1 \quad , \quad P_{AB}(a) = P_{AB}(b) = \frac{1}{2} \tag{5.1}$$

The distribution is obviously no-signaling: the marginal of Alice does not depend on Bob's measurement (nor on Alice's, in this specific case) and the same for Bob. The correlation says that for $(A,B) \in \{(0,0),(0,1),(1,0)\}$ one has $a = b$ (perfect correlation) while for $(A,B) = (1,1)$ one has $a = b \oplus 1$ (perfect anti-correlation). Therefore, if one evaluates the CHSH expression on this distribution, one finds $\langle S \rangle = 4$. In other words, this innocent-looking probability distribution reaches the largest possible value of CHSH, while we have seen that quantum physics cannot go beyond the Tsirelson bound $\langle S \rangle = 2\sqrt{2}$.

The hypothetical resource that would produce the distribution (5.1) has been called *PR-box*. After the Poescu-Rohrlich paper, the PR-box remained rather in the shadow for a few years. Interest in it was revived mainly by two studies. Computer scientists were astonished by the result of van Dam, who noticed that this box would make "communication complexity" tasks trivial [van Dam 2005]. For physicists, the result of Cerf, Gisin, Massar and Popescu is maybe more appealing [Cerf et al. 2005, Degorre et al. 2005]: they proved that the correlations of the singlet can be simulated by local variables plus a single use of the PR-box (thus improving on a previous result by Toner and Bacon [Toner and Bacon 2003], who showed the same but using one bit of communication as non-local resource). This latter work raised the hope that the PR-box would play the same role of elementary building-block for non-local distributions, as the singlet plays for quantum states. This hope was later shattered: there are multi-partite non-local distributions that cannot be obtained even if the partners, pairwise, share arbitrarily many PR-boxes [Barrett and Pironio 2005], and even for bipartite states it seems that full simulation will be impossible beyond the two-qubit case [Bacciagaluppi 2008].

However, the fact that initial hopes are shattered is not new in physics (nor in life, for that matters): creativity is not stopped by that, and indeed, several further interesting results have been obtained along the line of thought started by Popescu and Rohrlich. In order to appreciate them, we have to go a step further. Given the gap between 4 and $2\sqrt{2}$, one can easily guess that the PR-box is not the only no-signaling resource outside quantum physics: there is actually a continuous family of such objects. The next paragraph is devoted to the formal framework in which no-signaling distributions can be studied.

5.2. Formal framework to study no-signaling distributions

5.2.1. A playful interlude

As a warm up, we consider a family of games in which the two players, Alice and Bob, after learning the rules, are sent to two different locations. There, each of them is submitted to some external input and has to react. The game is won if the reactions are in agreement with the rules that have been fixed in advance. Specifically, consider the following[20]

[20] In all the following examples, we assume that there is no correlation between the content of a paper and its acceptance or rejection by a referee. This is not a very strong assumption, especially if the referee has some other interests, like here: winning a game.

Game 1: *each of the players receives a paper to referee. The game is won if, whenever the two players receive the same paper, they produce the same answer (i.e., either both accept or both reject it).*

This game is easy to win: Alice and Bob can just agree in advance that, whatever paper they receive, they will accept it. Admittedly, if many runs of the game are played, this strategy may be a bit boring; but more elaborated ones are possible, for instance: they accept in the first run, reject in the second and third, then accept again in the fourth... Also, the strategy in each round may be much more subtle than a simple "accept all" or "reject all": for instance, in one round they can decide to "accept when ([first author name starts with A-M] OR [has been submitted to IJQI]) AND [does not come from CQT]; reject otherwise". This family of strategies are called pre-established strategies. Obviously, there are uncountably many *pre-established strategies.* Remarkably, though, their set can be bounded, as we already know: it is the convex body, whose extremal points are local deterministic strategies.

Let us now change the rules of the game:

Game 2: *ach of the players receives a paper to referee. The game is won if the two players produce the same answer only when they receive the same paper, and produce different answers otherwise.*

If the set of possible inputs consists only of two papers, the game can still be won by pre-established strategy: the players have just to agree on which paper to accept and which to reject. As soon as there are more than two possible inputs, however, pre-established strategies cannot win the game with certainty. For instance, consider the case where there are three possible papers and write down the conditions for winning: $a_1 = b_1, a_2 = b_2, a_3 = b_3, a_1 \neq \{b_2, b_3\}$ etc. It's easy to convince oneself that no set of six numbers $\{a_1, a_2, a_3\} \times \{b_1, b_2, b_3\}$ can fulfill all the conditions. Since no local deterministic strategy can win the game, no pre-established strategy can.

In order to win such a game, in the classical world one has to use the other resource: *communication.* In view of what we know about quantum physics, let us stress here that communication is, in a sense, an exaggerated resource for Game 2. Indeed, with communication, one can win all possible games, for instance

Game 3: *each of the players receives a paper to referee. The game is won under the condition: Alice accepts her paper if only if the paper received by Bob has been authored in CQT.*

Why does Game 3 seem more extreme than Game 2? Because Game 3 requires Alice to learn *specific information* about the input received by Bob; while in Game 2, the criteria for winning include only *relations* between Alice's and Bob's inputs and outputs. In other words, communication is intrinsically required to win Game 3; while one might hope to win Game 2 without communication - with no-signaling resources.

5.2.2. Local and no-signaling polytopes: a case study

Turning now to the formalism, I think it more useful, for the purpose of this school, to do a fully developed case study, rather than giving the general formulas, that those who are going to work in the field will easily find in the literature. Therefore, we focus on the simplest non-trivial case, the same as for the CHSH inequality and the PR-box: two partners, each with binary input $A, B \in \{0,1\}$ and binary output $a, b \in \{0,1\}$.

Probability space

The first element to be studied is the *dimensionality of the probability space*: how many numbers are required to specify the four no-signaling probability distributions $P_{AB}(a,b)$ completely? In all, there are sixteen probabilities; but since each of the four P_{AB} must be normalized, we can already reduce to twelve. Actually, a more clever reduction to only *eight* parameters can be achieved exploiting the no-signaling condition. Indeed, note first that the three numbers needed to specify P_{AB} can be chosen as being $P_{AB}(a=0)$, $P_{AB}(b=0)$ and $P_{AB}(a=0,b=0)$, since $P_{AB}(a=0,b=1) = P_{AB}(a=0) - P_{AB}(a=0,b=0)$, $P_{AB}(a=1,b=0) = P_{AB}(b=0) - P_{AB}(a=0,b=0)$ and $P_{AB}(a=1,b=1)$ follows by normalization. Furthermore, because of no-signaling, $P_{AB}(a=0) = P_A(a=0)$ for all A and $P_{AB}(b=0) = P_B(b=0)$ for all B. Therefore, we are left with eight probabilities that can be conveniently arranged as a table:

$$\mathbf{P} \;=\; \begin{array}{c|cc} & P_{B=0}(b=0) & P_{B=1}(b=0) \\ \hline P_{A=0}(a=0) & P_{00}(0,0) & P_{01}(0,0) \\ P_{A=1}(a=0) & P_{10}(0,0) & P_{11}(0,0) \end{array} \tag{5.2}$$

For instance, the probability distribution of a PR-box (5.1) and the one associated to the best measurements on a maximally entangled state read respectively

$$\mathbf{P}_{PR} \;=\; \begin{array}{c|cc} & 1/2 & 1/2 \\ \hline 1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \end{array} \;,\; \mathbf{P}_{ME} \;=\; \begin{array}{c|cc} & 1/2 & 1/2 \\ \hline 1/2 & \frac{1=1/\sqrt{2}}{4} & \frac{1=1/\sqrt{2}}{4} \\ 1/2 & \frac{1=1/\sqrt{2}}{4} & \frac{1=1/\sqrt{2}}{4} \end{array} \tag{5.3}$$

A priori, Bell's inequalities will appear naturally later in the construction; however, since we have already derived the CHSH inequality in a different way, we can legitimately study here how the inequality looks like in this notation. Noting that

$$E_{AB} = 1 - 2[P_{AB}(0,1) - P_{AB}(1,0)] = 4P_{AB}(0,0) - 2P_A(a=0) - 2P_B(b=0) + 1$$

we find

$$\langle S \rangle = 4[P_{00}(0,0) + P_{01}(0,0) + P_{10}(0,0) - P_{11}(0,0) - P_A(a=0) - P_B(b=0)] + 2$$

Remembering that the inequality reads $-2 \leq \langle S \rangle \leq 2$, by simply rearranging the terms we find

$$-1 \leq P_{00}(0,0) + P_{01}(0,0) + P_{10}(0,0) - P_{11}(0,0) - P_A(a=0) - P_B(b=0) \leq 0 \tag{5.4}$$

i.e., the inequality known as Clauser-Horne (CH) - which is therefore *strictly equivalent* to CHSH, under the assumption of no-signaling. Written as a table, we have

$$\mathbf{T}_{CH} \;=\; \begin{array}{c|cc} & -1 & 0 \\ \hline -1 & 1 & 1 \\ 0 & 1 & -1 \end{array} \tag{5.5}$$

and the inequality reads

$$-1 \leq \mathbf{T}_{CH} \cdot \mathbf{P} \leq 0 \tag{5.6}$$

where $\cdot$ represent term-by-term multiplication. For instance, $\mathbf{T}_{CH} \cdot \mathbf{P}_{PR} = \frac{1}{2}$ and $\mathbf{T}_{CH} \cdot \mathbf{P}_{ME} = \frac{1}{\sqrt{2}} - \frac{1}{2}$.

Local deterministic points (vertices of the local polytope)

The next step consists in identifying all the *local deterministic strategies*. There are only four deterministic functions $a = f(A)$ from one bit to one bit: $f_1(A) = 0$, $f_1(A) = 0$, $f_1(A) = 0$ and $f_1(A) = 0$. Therefore, there are sixteen local deterministic strategies $D_{AB}^{ij}(a,b) = \delta_{a=f_i(A)}\delta_{b=f_i(B)}$. Let us write down explicitly:

$$\mathbf{D}_{11} = \begin{array}{c|cc} & 1 & 1 \\ \hline 1 & 1 & 1 \\ 1 & 1 & 1 \end{array}, \mathbf{D}_{12} = \begin{array}{c|cc} & 0 & 0 \\ \hline 1 & 0 & 0 \\ 1 & 0 & 0 \end{array}, \mathbf{D}_{13} = \begin{array}{c|cc} & 1 & 0 \\ \hline 1 & 1 & 0 \\ 1 & 1 & 0 \end{array}, \mathbf{D}_{14} = \begin{array}{c|cc} & 0 & 1 \\ \hline 1 & 0 & 1 \\ 1 & 0 & 1 \end{array}$$

$$\mathbf{D}_{21} = \begin{array}{c|cc} & 1 & 1 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}, \mathbf{D}_{22} = \begin{array}{c|cc} & 0 & 0 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}, \mathbf{D}_{23} = \begin{array}{c|cc} & 1 & 0 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}, \mathbf{D}_{24} = \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}$$

$$\mathbf{D}_{31} = \begin{array}{c|cc} & 1 & 1 \\ \hline 1 & 1 & 1 \\ 0 & 0 & 0 \end{array}, \mathbf{D}_{32} = \begin{array}{c|cc} & 0 & 0 \\ \hline 1 & 0 & 0 \\ 0 & 0 & 0 \end{array}, \mathbf{D}_{33} = \begin{array}{c|cc} & 1 & 0 \\ \hline 1 & 1 & 0 \\ 0 & 0 & 0 \end{array}, \mathbf{D}_{34} = \begin{array}{c|cc} & 0 & 1 \\ \hline 1 & 0 & 1 \\ 0 & 0 & 0 \end{array}$$

$$\mathbf{D}_{41} = \begin{array}{c|cc} & 1 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 1 & 1 \end{array}, \mathbf{D}_{42} = \begin{array}{c|cc} & 0 & 0 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 0 \end{array}, \mathbf{D}_{43} = \begin{array}{c|cc} & 1 & 0 \\ \hline 0 & 0 & 0 \\ 1 & 1 & 0 \end{array}, \mathbf{D}_{44} = \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}$$

According to the theorem shown in Lecture 4, we know that the set of local distributions (distributions that can be obtained with pre-established strategies) is the convex set whose extremal points are the deterministic strategies. A convex set with a finite number of extremal point is called "polytope", therefore we shall call this set *local polytope*. The extremal points of a polytope are called *vertices*.

Facets of the local polytope: Bell's inequalities

The vertices of the local polytope define its *facets*, i.e., the planes that bound the set. If a point is below the facet, the corresponding probability distribution can be reproduced with local variables; if a point is above the facet, it cannot. Therefore, facets are the geometric representation of Bell's inequalities! Actually, in addition to Bell's inequalities, there are many *trivial facets* that can never be violated: conditions like $P_{AB}(a,b) = 0$ or $P_{AB}(a,b) = 1$ obviously define boundaries within which every local distribution must be found... because *any* distribution must be found therein! We forget about these trivial facets in what follows. In our simple case, the local polytope is a polygon in an 8-dimensional space, so its facets are 7-dimensional planes. One has to identify the sets of eight points that define one of these planes, then write the equation that define each plane. This can be done by brute force, but we just use some intuition and then quote the known result. The CH inequality indeed defines facets: we have $\mathbf{T}_{CH} \cdot \mathbf{D} = 0$ for $\mathbf{D}_{11}$, $\mathbf{D}_{13}$, $\mathbf{D}_{22}$, $\mathbf{D}_{24}$, $\mathbf{D}_{31}$, $\mathbf{D}_{34}$, $\mathbf{D}_{42}$, $\mathbf{D}_{43}$; and one can check that these eight points indeed define a plane of dimension 7. Above this facet, the most non-local point is the PR-box (5.1).

Also, $\mathbf{T}_{CH} \cdot \mathbf{D} = -1$ for the other eight deterministic points: this facet is "opposite" to the previous one, with the local polytope between the two. The most non-local point above this facet is also a PR-box, the one defined by the rule $a \oplus b = AB \oplus 1$. Note that this PR-box is obtained from the "original" one by trivial local processing: e.g., Alice flips her outcome.

A simple argument of symmetry gives us immediately six other equivalent facets. Indeed, given a Bell inequality, a relabeling of the inputs and/or the outputs provides another Bell inequality. Their number is most easily counted by studying how many different rules for PR-boxes one can find, and these are obviously $a \oplus b = (A \oplus 1)B$, $a \oplus b = A(B \oplus 1)$ and $a \oplus b = (A \oplus 1)(B \oplus 1)$, with the corresponding opposite facets obtained as before by adding 1 on either side.

Now, it can be proved that there are no other Bell's inequalities for this case: only eight versions of CHSH, each with eight extremal points on the corresponding facet[21] and one single PR-box on top. All these facets being equivalent to the others up to trivial relabeling of the inputs and/or the outputs, it is customary to say that *there is only one Bell inequality for the case of two users, each with binary inputs and outputs; namely, CHSH (or CH).*

The no-signaling polytope and the quantum set

We have studied at length the set of local distributions. Now we have to say a few words about the two other meaningful sets, namely the no-signaling distributions and the distributions that can be obtained with quantum physics.

[21]

The image of the probability space is usually drawn as in Fig. 5.1. This drawing is of course a rather poor representation of an 8-dimensional object.

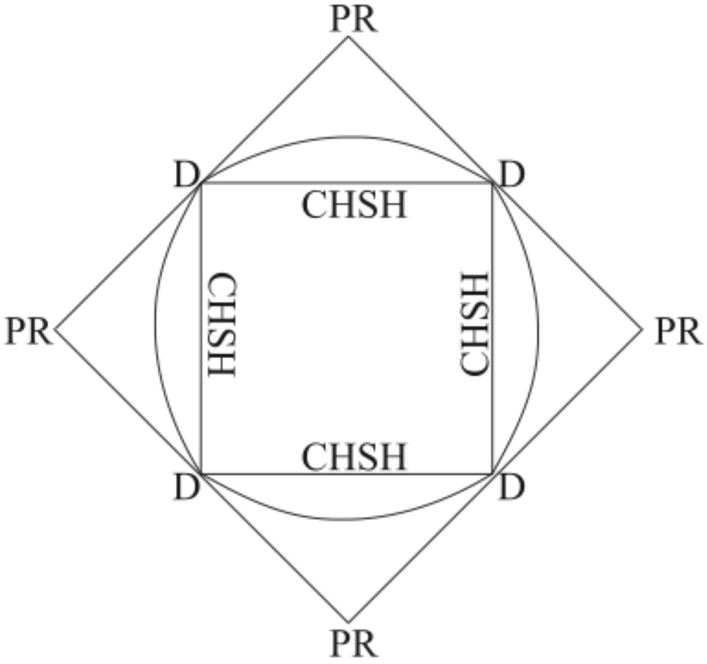

Figure 5.1: Representation of the local polytope, the set of quantum correlations and the no-signaling polytope, for the case under study (2 parties, 2 inputs and 2 outputs per party). The local polytope (inner square) is delimited by versions of the CHSH inequality; the quantum set (round body) exceeds the local polytope, and is contained in the no-signaling polytope (external square), whose extremal points are PR-boxes. Local deterministic points (D) are extremal points of all three sets (not obvious in the drawing, which is just a projection on a two-dimensional slice: the deterministic points do not lie on this slice).

The no-signaling conditions obviously define a convex set (if two distributions are no-signaling, any convex combination will also be no-signaling). It turns out that this set is also a polytope, i.e., it has a finite number of extremal points; obviously, it is called *no-signaling polytope*. All the local deterministic points are also extremal for the no-signaling polytope; but in addition to those, of course, there are some non-local ones, that can in principle be found[22]. For the simple example we studied, the only additional extremal points are the eight PR-boxes defined above [Barrett, Linden et al. 2005].

The quantum set is also convex, but is not a polytope: it has an uncountable number of extremal points. Interestingly, at the moment of writing, the shape of this convex body has not been characterized in full generality yet, not even for the simple case under study here. A necessary condition is the following [Tsirelson 1987, Landau 1988, Masanes 2003]: for any probability distribution coming from quantum physics, the correlation coefficients must satisfy an inequality that reminds of CHSH, namely

$$|\sin^{-1}(E_{AB}) + \sin^{-1}(E_{AB'}) + \sin^{-1}(E_{A'B}) - \sin^{-1}(E_{A'B'})| \leq \pi \tag{5.7}$$

However, this condition is provably not sufficient in general: there are probability distributions that satisfy this inequality but cannot be produced with quantum states [Navascues et al. 2008].

Final remarks

Let us conclude this section by saying that most of the simple features of this case study are *not* maintained as soon as one considers more complex situations, i.e., more than two inputs, or more than two outputs, or more than two parties.

For a general polytope, the task of finding the facets given the vertices is increasingly complex (in fact, it's provably computationally hard). Many Bell's inequalities have by now been found with computers finding facets - and possibly sorting the equivalent ones, otherwise the number is just overwhelming[23].

[22] The facets of the no-signaling polytope are rather easy to characterize: they are the "trivial" facets that satisfy no-signaling. By intersecting them, a computer program finds the extremal points. So the procedure it?s somehow the reverse of the one used to find Bell's inequalities from local deterministic strategies.

[23] Note that "lower" Bell inequalities remain facets even in more general cases [Pironio 2005]: for instance, in any local polytope there are CHSH-like facets that involve only two of the users, two of the inputs and two of the outputs. However, even removing those obvious embeds from lower dimensional polytopes, the number of inequivalent Bell's inequalities grows extremely fast. As an anecdote: Pitowsky and Svozil listed 684 inequalities for the case where the users have each ternary input and binary output [Pitowski and Svozil 2001]. A later inspection by Collins and Gisin [Collins and Gisin 2004] proved

Also the characterization of the no-signaling polytope becomes cumbersome. Once moving away from the simplest case, it is no longer the case that all the extremal no-signaling points are equivalent up to symmetries; nor that only one extremal no-signaling point can be found above each facet. The number of objects increases rapidly and the impossibility of visualizing these high-dimensional structures is a powerful deterrent not to embark on such studies unless driven by a very good reason.

5.3. The question revisited: why $2\sqrt{2}$?

5.3.1. What no-signaling non-local points share with quantum physics

Most of the features that are usually highlighted as "typically quantum" are actually shared by all possible theories that allow non-locality without signaling. This vindicates, in a sense, the intuition of Popescu and Rohrlich: non-locality without signaling is a deep physical principle, to which many observed facts can be attributed.

Here is a rapid list of features that seem to be the share of no-signaling non-local theories, rather than of quantum physics alone:

- *Intrinsic randomness.* Consider first the PR-box: if one enforces the rule $a \oplus b = xy$ (i.e., maximal algebraic violation of CHSH), it is easy to verify that any prescription for the local statistics other than $P(a = 0) = P(b = 0) = \frac{1}{2}$ leads to signaling. In other words, the local outcomes of the PR-box *must* be random in order to satisfy no-signaling. Now, this holds in fact for any non-local probability distribution: as soon as $\mathbf{P}$ violates a Bell inequality (be it compatible or not with quantum physics), the local outcomes cannot be deterministic. We'll see in the next Lecture that this can be quantified to provide guaranteed randomness.

- *No-cloning theorem.* A form of no-cloning theorem can be defined for all non-local no-signaling probability distributions [Masanes et al. 2006, Barnum et al. 2007]. The formulation goes as follows: suppose there exist $\mathbf{P}_{ABB'} = P(a, b, b'|x, y, y')$ such that $\mathbf{P}_{AB}$ and $\mathbf{P}_{AB'}$ are the same

that there are actually only *two* non-equivalent ones: CHSH for sure, and a new one that they baptized I_{3322} but that in fact had been already proposed several years earlier in a forgotten paper [Froissard 1981].

distribution; then $\mathbf{P}_{AB}$ is local. In other words, if $\mathbf{P}_{AB}$ is non-local, one cannot find an extension $\mathbf{P}_{ABB'}$ satisfying no- signaling and such that $\mathbf{P}_{AB'}$ is equal to $\mathbf{P}_{AB}$: the box of B cannot be cloned. The proof is particularly simple in the case of the PR box, see Tutorial.

- *Possibility of secure cryptography.* It seems that the security of cryptography can be proved only on the basis of no-signaling, without invoking at all the formalism of quantum physics [Barrett, Hardy et al. 2005, Aćin, Gisin and Masanes 2006]. The exact statement is slightly more involved, since a few technical problems in the security proofs have not been sorted out yet: see the most recent developments for all details [Masanes 2008, Hänggi et al. 2009].

- *Uncertainty relations, i.e., information-disturbance tradeoff.* Such a tradeoff has been discussed in the context of cryptographic protocols [Scarani et al. 2006].

- *Teleportation and swapping of correlations.* After a first negative attempt [Short et al. 2006], it seems that they can actually be defined as well within the general no-signaling framework [Barnum et al. 2008, Skrzypczyk et al. 2009]. Still work in progress.

5.3.2. And what they do not.

Still, there is overwhelming evidence that our world is well described by quantum physics, while there seems to be no evidence of more-than-quantum correlations. Where does the difference lie? This is an open question. Again, I present a rapid list of what is known.

- *Poor dynamics.* Recall that we are playing with purely kinematical concepts: the $\mathbf{P}$'s are, at least at first sight, the analog of the measurement of an entangled state, not of the state itself! There are suggestions that the allowed dynamics of objects like PR-boxes would be seriously restricted [Barrett 2007, Short and Barrett 2009, Gross et al. 2009].

- *Communication tasks becoming trivial.* A few communication tasks have been found, for which quantum physics does not lead to any significant advantage over classical physics, while some more-than-quantum correlations would start helping (the task becoming ultimately trivial if PR-boxes would be available). As mentioned above, the first such example was *communication complexity* [van Dam 2005, Brassard et al. 2006, Brunner and Skrzypczyk 2009]. Two further tasks were proved to become more efficient than allowed in the quantum world as soon as $S > 2\sqrt{2}$: *non-local distributed computing* [Linden et al. 2007] and a form of *random access codes* [Pawłowski et al. 2009, Allcock et al. 2009]. These last works formulated the principle of "information causality" as a possible criterion to rule more-than-quantum correlations out.

- *Classical limit.* Finally, it was noticed that most stronger-than-quantum correlations would not recover the classical world in the limit of many copies [Navascues and Wunderlich 2009]. This criterion of "macroscopic locality" is related to a family of experiments rather than to an information-theoretical task.

5.4. Tutorials

5.4.1. Problems

Exercise 5.1

Which of the following games can be won by pre-established strategies?

1. Each of the players receives a paper to referee. The game is won if both players produce always the same answer, unless both papers come from CQT, in which case they must produce different answers.

2. Generalization of Game 2 of the lectures: the two players must produce the same output if and only if they received the same input. *Hint:* Consider the cases where the number of possible inputs is smaller, equal or larger than the number of possible outputs.

Exercise 5.2

Figure 5.2 represents a two-dimensional slice of the no-signaling polytope for 2 parties, 2 inputs and 2 outputs per party. Add the missing information. *Hint:* in order to assign the deterministic points, re-write the inequality in the form of a table, as done in the text.

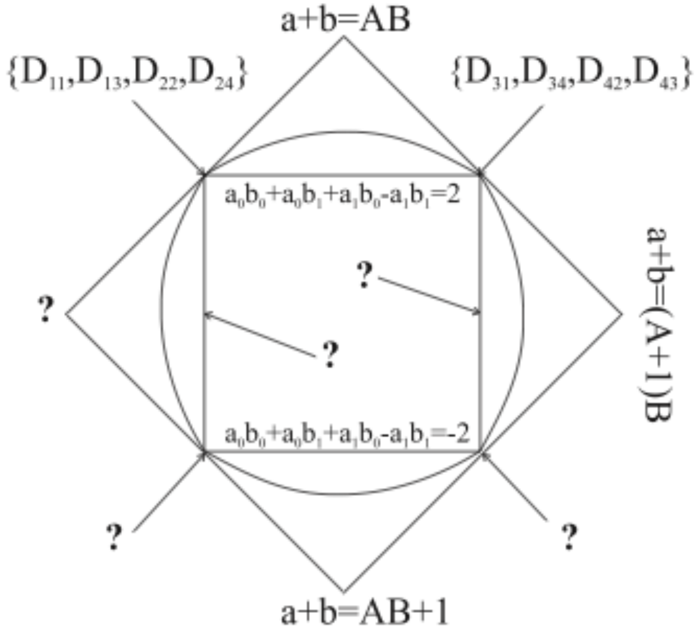
a+b=AB
{D$_{11}$,D$_{13}$,D$_{22}$,D$_{24}$}
{D$_{31}$,D$_{34}$,D$_{42}$,D$_{43}$}
a$_0$b$_0$+a$_0$b$_1$+a$_1$b$_0$-a$_1$b$_1$=2
?
?
a+b=(A+1)B
?
a$_0$b$_0$+a$_0$b$_1$+a$_1$b$_0$-a$_1$b$_1$=-2
?
?
a+b=AB+1

Figure 5.2: Slice of the no-signaling polytope for 2 parties, 2 inputs and 2 outputs per party.

Exercise 5.3

Prove that a tripartite distribution $\mathbf{P}_{ABB'} = P(a,b,b'|x,y,y')$, such that $a \oplus b = xy$ and $a \oplus b' = xy'$, violates the no-signaling constraint. This is the proof of the no-cloning theorem for the specific case of the PR-box.

5.4.2. Solutions

Exercise 5.1

The first game is obviously non-local: it looks like a generalized PR-box with a large number N of inputs (the possible affiliations) and two outcomes. Interestingly, in the limit $N \to \infty$, these correlations can be achieved with quantum states [Barrett et al. 2006].

The second game can be won with local strategies if and only if the number of outputs is larger or equal to the number of inputs. Indeed, if this is case, the players agree to output their input, or a function thereof on which they had agreed upon in advance. If this is not the case, a generalization of the argument given in the lectures leads to the conclusion that the game cannot be won. **Exercise 5.2**

The inequality below the PR-box $a \oplus b = (x+1)y$ is obtained from (5.5) by flipping Alice's input, i.e., by flipping the lines:

$$\mathbf{T}_{CH} = \begin{array}{c|cc} & -1 & 0 \\ \hline 0 & 1 & -1 \\ -1 & 1 & 1 \end{array} \qquad (5.8)$$

The rest follows quite immediately; the full result is given in Fig. 5.3.

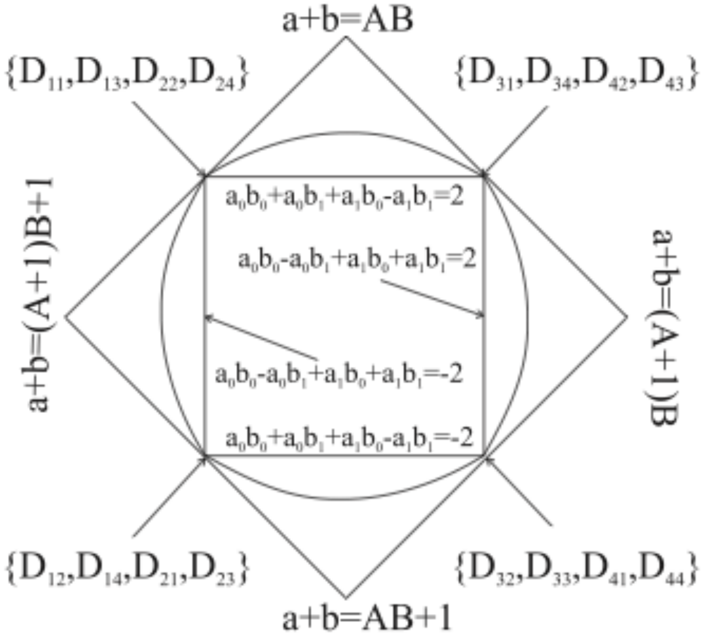

Figure 5.3: Slice of the no-signaling polytope for 2 parties, 2 inputs and 2 outputs per party.

Exercise 5.3

The proof is very simple: the two conditions $a \oplus b = xy$ and $a \oplus b' = xy'$ imply $b \oplus b' = x(y \oplus y')$. So,if both B and B' are given to Bob, he knows b, b', y and y' and can therefore reconstruct Alice's input x. In other words, signaling is possible from A to (B, B').

6. Lecture Quantum correlations (III): the power of Bell

6.1. The model, again

In section 4.2.1, the "local variable" model was presented in its original flavor (though clarified by some more modern terminology than the one that was used in the 1960's to 1980's): namely, a possible alternative description of nature, supposedly along the line Einstein had in mind. Through the check of Bell's inequalities, experiment is called to rule out this particular model. I find it useful to open this last lecture by representing the model under a different light [Gisin 2007], inspired by the present discussions in quantum information science.

The scenario is about two players, Alice and Bob, who are asked to reproduce some quantum statistics without actually measuring the state. Specifically, Alice and Bob are brought together and told that the shall have to reproduce the statistics of a two-particle state ρ under a family of measurements $\{A_j\}$ for Alice and $\{B_k\}$ for Bob. Let's stress that Alice and Bob are given a *complete and exact description on paper* of both the state and the measurements: therefore, they know the statistics $P_{A_j,B_k}(a,b)$ they should obtain for all j and k. While still together, they are also allowed to agree on a *common strategy* denoted s. After this, Alice and Bob are brought far apart from one another, without possibility of communication, and the game starts: run after run, an external party gives Alice a value of j and Bob a value of k, upon which Alice and Bob have to produce outcomes a and b that will satisfy the expected statistics.

Now, Alice and Bob can succeed if and only if the set of $P_{A_j,B_k}(a,b)$ does not violate a Bell-type inequality. While this conclusion should not surprise the reader at this stage, it may be useful to stress that Alice knows quite a lot about Bob's part: she knows the whole state they are supposed to share, she knows the set of measurements Bob will be asked to choose from, and she may have shared a strategy with him. The only thing Alice does not know about Bob is, which k he will receive in each specific run. The same holds for Bob's knowledge about Alice's part. In other words, the "local variable" can be as large as $\lambda = \{``\rho'', ``\{A_j\}'', ``\{B_k\}'', s\}$.

The idea can be presented in yet another way: *if Alice and Bob observe a violation of Bell's inequalities and the correlations cannot be attributed to a signal (e.g., because the emission of the signal and the choice of the measurement are space-like separated), then Alice and Bob must be measuring entangled states.* In other words, it makes a difference whether Alice and Bob actually share entangled particles and measure them, or whether they try to simulate the same statistics with purely classical means.

6.2. Device-independent: quantum information in a black box

The fact that violation of Bell's inequalities implies entanglement has been recognized very early in quantum information: in particular, this was precisely Ekert's intuition when he re-discovered quantum cryptography [Ekert 1991]. The new development, that I personally find very exciting, is the awareness that *the amount of violation of Bell's inequalities makes it possible to derive quantitative statements on the "quantumness" of a black box.* Here follow the specific examples that have been studied to date. Since each case refers to a very limited number of papers, I keep this text quite short, hoping to encourage the reader to read the original papers for all details.

6.2.1. Device-independent quantum cryptography

The first task, to which the idea of device-independent assessment has been applied, is quantum cryptography. This research project amounts precisely at making Ekert's intuition quantitative: if the observed value of the CHSH parameter is S, how much information could the eavesdropper Eve have got? Here is *cryptography at its most paranoid*: the authorized partners Alice and Bob could have bought even the source and the measurement devices from the eavesdropper, provided they control the choices of the measurements!

I direct the reader to the original references for a thorough discussion of the scenario, the results and the open issues [Aćin et al. 2007, Ling et al. 2008, Pironio et al. 2009]. See also [Scarani and Kurtsiefer 2009] for a more personal assessment of the importance of these results for quantum cryptography in general.

6.2.2. Device-independent source characterization

Once the idea of device-independent assessment is understood, a rather obvious task is the direct assessment of the quality of a source: the observed violation of a Bell inequality should be related to the amount of entanglement of the produced state. This question can be given a slightly different turn, inspired by the idea of device-testing [Mayers and Yao 2004, Magniez et al. 2006]: if I use the source that yields a given violation for a specific task, how probable it is that the result differs from the one I would have obtained with an ideal source? As the reader reminds from section 3.1.2, this amounts at finding a bound for the trace distance as a function of the violation of a Bell's inequality. The derivation of such bounds seems very complex even in the simplest case, only partial results are known [Bardyn et al. 2009].

6.2.3. Guaranteed randomness

Yet another intuitive statement that can be turned quantitative: if one violates Bell's inequalities, the individual outcomes must have some intrinsic randomness - because pseudo-randomness is ultimately deterministic and therefore cannot violate Bell's inequalities! Phrased in a different way: if one observes a violation of Bell's inequalities, one can be sure that the underlying process is random.

This can be an extremely exciting development. As of today, in order to certify a random number generator, one has to look in detail into the complicated process and get somehow convinced that something random (or, at the very least, very hard to predict) is happening. On the contrary, Bell's inequalities provide a very straightforward way of certifying randomness, based on the observed statistics only.

At this point, I may disappoint the reader: at the moment of writing, there is no available document on this topic. To my knowledge, the first such study lies deeply buried in the Ph.D. thesis of Roger Colbeck (Cambridge University); for reasons I refrain from commenting here, this study may never be turned into a publication. A much more thorough investigation by Antonio Aćin, Serge Massar and Stefano Pironio has been announced in several conferences and will hopefully be available soon.

6.2.4. Dimension witnesses

Finally, Bell's inequalities can be used to estimate bounds on the dimension of the Hilbert space of the measured particles. More specifically, one can have the following statement: if an observed violation of some inequality exceeds some threshold, then the system that is being measured must be at least of dimension D. The derivation of such statements is technically demanding, in particular because one wants to rule out also POVMs: in order, for instance, to exclude qubits, it is not enough to observe that three outcomes are possible!

As it turns out, the CHSH inequality cannot be used as a dimension witness, because one can already obtain all possible violations with two-qubit states. The first results were obtained independently for inequalities with ternary outcomes [Brunner et al. 2008] and for inequalities with binary outcomes and more than two measurements [Vértesi and Pál 2009]. Note that, in itself, the problem of bounding the dimension of the Hilbert space does not require the use of Bell?s inequalities [Wehner et al. 2008].

6.3. Detection loophole: a warning

In all the previous paragraph, I have written "violation of Bell?s inequality" several times. This expression refers to a *conclusive* violation, i.e., one that is not obtained by post-selection; in other words, *the detection loophole must be closed.* We have already encountered this loophole in Section 4.2.5. There, in the context of experiments, it looked as a very minor point, an absurd conspiracy theory invented by the die-hard of classical mechanisms. Indeed, the vast majority of physicists is certain that the detection loophole will be closed one day: as John Bell himself used to note, it's hard to believe that quantum physics will suddenly be falsified just by increasing the efficiency of our detectors.

However, in device-independent assessment, the situation is quite different. In this scenario, we have not been built devices in order to ask questions to a benevolent (or at least, not malevolent) Nature: rather, we are testing devices built by someone else - and this someone else may be adversarial. Even in cryptography, the vendor of an allegedly good source or trusted random number generator: all may have very good reasons to try and cheat us! In other words, if we are not careful, they may engineer a purely classical device that exhibits "violations of Bell?s inequalities" using the detection loophole. This situation has triggered a revival of interest in the detection loophole. For those who want to know more, there is no review paper available, but the latest work [Vértesi, Pironio and Brunner 2009] summarizes pretty well the situation.

6.4. Challenges ahead

When I entered the field of quantum information in the year 2000, the general atmosphere was rather cold about Bell's inequalities. Indeed, the general argument one could hear was: "we know by now that local hidden variables are not there, it's time to study entanglement theory". Some of us tried to counter such an argument by vague statements like "Bell's inequalities uncover one of the most astonishing quantum features, they must be useful for something" - not very convincing... but true! It took several years and several detours, but at the moment of writing, the role of Bell's inequalities in quantum information has been vindicated: they are the only tool for device-independent quantum information.

One can scarcely imagine quantum information without the notion of "qubit" and still, the assessment on the dimensionality of a quantum system requires a very careful characterization. I would not go as far as suggesting that quantum information will ultimately be formulated without qubits; were it only because there are scenarios in which the physical system is well characterized and there is no reason to be paranoid. But the possibility of device-independent assessment is fascinating, useful... and with plenty of challenges ahead for both theorists and experimentalists!

6.5. Tutorials

6.5.1. Problems

Exercise 6.1

In this exercise, we consider the detection loophole for maximally entangled state and the CHSH inequality; but the construction can be easily generalized to study other situations, e.g., non-maximally entangled states [Eberhard 1993] or asymmetric detection efficiencies [Cabello and Larsson 2007, Brunner et al. 2007].

Suppose for simplicity that Alice and Bob have a heralded pair source, so that they know when a pair of particles is expected to arrive. They are not allowed to discard any event, so they agree on the following: if Alice?s detector fires, she keeps the result of the detection (+1 or −1); if not, she arbitrarily decides that the result is +1. Bob follows the same strategy. Let now η be the efficiency of the detectors (supposed identical for all detectors): then the observed average of CHSH will be

$$\langle S \rangle = \eta^2 S_2 + 2\eta(1-\eta)S_1 + (1-\eta^2)S_0 \tag{6.1}$$

where S_2 is the expected value of CHSH when both Alice's and Bob's detectors have fired, S_1 when only either Alice's or Bob's have fired, and S_0 where none has fired.

1. Supposing everything else is perfect (ideal measurements, no dark counts etc), what are the values of S_2, S_1 and S_0?
2. Prove that $\langle S \rangle > 2$ can be achieved only if $\eta > \frac{2}{\sqrt{2}+1} \approx 82\%$.

6.5.2. Solutions

Exercise 6.1

If everything is perfect, when both photons are detected we have $S_2 = 2\sqrt{2}$. However, $S_1 = 0$: when one party detects and the other does not, there are no correlations between the results; and since the detection comes from half of a maximally entangled state, that bit is completely random. Finally, $S_0 = 2$: by pre-established agreement, both Alice and Bob output +1, thus achieving the local bound. The threshold for η follows immediately.

At the address of experimentalists: the fact that, in case of no detection, the local bound is achieved, allows to get rid of the no-detection events altogether (these events are not even defined for most real sources, that are not heralded).

Bibliography

[Aćin, Gisin and Masanes 2006] Aćin, A., N. Gisin, L. Masanes, Phys. Rev. Lett. 97, 120405 (2006)

[Aćin, Toner and Gisin 2006] Aćin, A., B. Toner, N. Gisin, Phys. Rev. A 73, 062105 (2006)

[Aćin et al. 2007] Aćin, A., N. Brunner, N. Gisin, S. Massar, S. Pironio, V. Scarani, Phys. Rev. Lett. 98, 230501 (2007).

[Allcock et al. 2009] Allcock, J., N. Brunner, M. Pawłowski, V. Scarani, arXiv:0906.3464 (2009)

[Audenaert et al. 2007] Audenaert, K.M.R., J. Calsamiglia, L. Masanes, R. Muñoz-Tapia, A. Aćin, E. Bagan, F. Verstraete, Phys. Rev. Lett. 98, 160501 (2007)

[Bacciagaluppi 2008] Bacciagaluppi, G., arXiv:0811.3444 (2008)

[Bae and Aćin 2006] Bae, J., A.Aćin, Phys. Rev. Lett. 97, 030402 (2006)

[Bardyn et al. 2009] Bardyn, C.-E., T.C.H. Liew, S. Massar, M. McKague, V. Scarani, arXiv:0907.2170 (2009)

[Barnum et al. 2007] Barnum, H., J. Barrett, M. Leifer, A. Wilce, Phys. Rev. Lett. 99, 240501 (2007)

[Barnum et al. 2008] Barnum, H., J. Barrett, M. Leifer, A. Wilce, arXiv:0805.3553 (2008)

[Barrett 2002] Barrett, J., Phys. Rev. A 65, 042302 (2002)

[Barrett and Pironio 2005] Barrett, J., S. Pironio, Phys. Rev. Lett. 95, 140401 (2005)

[Barrett, Linden et al. 2005] Barrett, J., N. Linden, S. Massar, S. Pironio, S. Popescu, D. Roberts, Phys. Rev. A 71, 022101 (2005)

[Barrett, Hardy et al. 2005] Barrett, J., L. Hardy, A. Kent, Phys. Rev. Lett. 95, 010503 (2005)

[Barrett et al. 2006] Barrett, J., A. Kent, S. Pironio, Phys. Rev. Lett. 97, 170409 (2006)

[Barrett 2007] Barrett, J., Phys. Rev. A 75, 032304 (2007)

[Bell 1964] Bell, J.S., Physics 1, 195 (1964)

[Bennett et al. 1993] Bennett, C.H., G. Brassard, C. Crépeau, R.Jozsa, A. Peres, W.K. Wootters, Phys. Rev. Lett. 70, 1895(1993)

[Bennett et al. 1999] Bennett, C.H., D. DiVincenzo, T. Mor, P.W. Shor, J.A. Smolin, B.M. Terhal, Phys. Rev. Lett. 82, 5385 (1999)

[Bohm and Hiley 1993] Bohm, D., B.J. Hiley, The Undivided Universe (Routledge, New York, 1993)

Brassard et al. 2006] Brassard, G., H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger, Phys. Rev. Lett. 96, 250401 (2006)

[Branciard et al. 2008] Branciard, C., N. Brunner, N. Gisin, C. Kurtsiefer, A. Lamas-Linares, A. Ling, V. Scarani, Nature Physics 4, 681 (2008)

[Breuer and Petruccione 2002] Breuer, H.-P., F. Petruccione, The theory of open quantum systems (Oxford University Press, Oxford, 2002)

[Briegel et al. 1998] Briegel, H.-J., W. Dür, J.I. Cirac, P. Zoller, Phys. Rev. Lett. 81, 5932 (1998)

[Brunner et al. 2007] Brunner, N., N. Gisin, V. Scarani, C. Simon, Phys. Rev. Lett. 98, 220403 (2007)

[Brunner et al. 2008] Brunner, N., S. Pironio, A. Aćin, N. Gisin, A. A. M?ethot, V. Scarani, Phys. Rev. Lett. 100, 210503 (2008)

Brunner and Skrzypczyk 2009] Brunner, N., P. Skrzypczyk, Phys. Rev. Lett. 102, 160403 (2009)

[Bruss et al. 1998] Bruss, D., D.P. DiVincenzo, A. Ekert, C.A. Fuchs, C. Macchiavello, J.A. Smolin, Phys. Rev. A 57, 2368 (1998)

[Bužek and Hillery 1996] Bužek, V., M. Hillery, Phys. Rev. A 54, 1844 (1996)

[Bužek et al. 1999] Bužek, V., M. Hillery, R.F. Werner, Phys. Rev. A60, R2626 (1999)

[Caban and Rembielinski 1999] Caban, P., J. Rembielinski, Phys. Rev. A 59, 4187 (1999)

[Cabello and Larsson 2007] Cabello, A., J.-A. Larsson, Phys. Rev. Lett. 98, 220402 (2007)

[Caves et al. 2004] Caves, C.M., C.A. Fuchs, K. Manne, J.M. Renes, Found. Phys. 34, 193 (2004)

[Cerf et al. 2005] Cerf, N.J., N. Gisin, S. Massar, S. Popescu, Phys. Rev. Lett. 94, 220403 (2005)

[Cerf and Fiurášek 2006] Cerf, N.J., J. Fiurášek, Progress in Optics, vol. 49, Edt. E. Wolf (Elsevier, 2006), p. 455

[Chefles 2000] Chefles, A., Contemp. Phys. 41, 401 (2000)

[Christandl et al. 2009] Christandl, M., R. Koenig, R. Renner, Phys. Rev. Lett. 102, 020504 (2009)

[Cirel'son 1980] Cirel'son, B.S., Lett. Math. Phys. 4, 93 (1980)

[Clauser et al. 1969] Clauser, J.F., M.A. Horne, A. Shimony, R.A. Holt, Phys. Rev. Lett. 23, 880 (1969)

[Collins and Gisin 2004] Collins, D., N. Gisin, J. Phys. A: Math. Gen. 37, 1175 (2004)

[Cover and Thomas 2006] Cover, T.M., J.A. Thomas, Elements of information theory (Wiley-Interscience, 2nd edition 2006)

[Degorre et al. 2005] Degorre, J., S. Laplante, J. Roland, Phys. Rev. A 72, 062314 (2005)

[Devetak 2005] Devetak, I., IEEE Trans. Inf. Theory 51, 44 (2005).

[Dieks 1982] Dieks, D., Phys. Lett. 92A, 271 (1982)

[Dieks 1988] Dieks, D., Phys. Lett. A 126, 303 (1988)

[Duan et al. 2001] Duan, L.M., M.D. Lukin, J.I. Cirac, P. Zoller, Nature 414, 413 (2001)

[Eberhard 1989] Eberhard, P.H., in: W. Schommers (Ed.), Quantum Theory and Pictures of Reality (Springer, Berlin, 1989)

[Eberhard 1993] Eberhard, P.H., Phys. Rev. A 47, R747 (1993).

[Ekert 1991] Ekert, A.K., Phys. Rev. Lett. 67, 661 (1991)

[Froissard 1981] Froissard, M., Nuovo Cim. B 64, 241 (1981)

[Gisin 1991] Gisin, N., Phys. Lett. A 154, 201 (1991)

[Gisin and Massar 1997] Gisin, N., S. Massar, Phys. Rev. Lett. 79, 2153 (1997)

[Gisin 1998] Gisin, N., Phys. Lett. A 242, 1 (1998)

[Gisin 2007] Gisin, N., quant-ph/0702021 (2007)

[Gisin 2009] Gisin, N., arXiv:0901.4255 (2009)

[Gleason 1957] Gleason, A.M., J. Math. Mech. 6, 885 (1957)

[Greenberger et al. 1989] Greenberger, D.M., M. Horne, A. Zeilinger, in: E. Kafatos (ed.), Bells Theorem, Quantum Theory, and Conceptions of the Universe (Kluwer, Dordrecht, 1989), p. 69

[Gröblacher et al. 2007] Gröblacher, S., T. Paterek, R. Kaltenbaek, Č. Brukner, M. Żukowski, M. Aspelmeyer, A. Zeilinger, Nature 446, 871 (2007)

[Gross et al. 2009] Gross, D., M. Mu?ller, R. Colberck, O.C.O. Dahlsten, arXiv:0910.1840 (2009)

[Hänggi et al. 2009] Hänggi, E., R. Renner, S. Wolf, arXiv:0906.4760 (2009)

[Hastings 2009] Hastings, M.B., Nature Physics 5, 255 (2009).

[Hausladen and Wootters 1994] Hausladen, P., W. K. Wootters, J. Mod. Opt. 41, 2385 (1994).

[Hausladen et al. 1996] Hausladen, P., R. Jozsa, B. Schumacher, M. Westmoreland, W.K. Wootters, Phys. Rev. A 54, 1869 (1996).

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

[Herbert 1982] Herbert, N., Found. Phys. 12, 1171 (1982)

[Herzog and Bergou 2004] Herzog, U., J.A. Bergou, Phys. Rev. A 70, 022302 (2004)

[Horodecki et al. 1995] Horodecki, R., P. Horodecki, M. Horodecki, Phys. Lett. A 200, 340 (1995)

[Horodecki, Horodecki et al. 2005] Horodecki, K., M. Horodecki, P. Horodecki, J. Oppenheim, Phys. Rev. Lett. 94, 160502 (2005).

[Horodecki, Oppenheim and Winter 2005] Horodecki, M., J. Oppenheim, A. Winter, Nature 436, 673 (2005)

[Horodecki et al. 2009] Horodecki, R., P. Horodecki, M. Horodecki, K. Horodecki, Rev. Mod. Phys. 81, 865 (2009)

[Huttner et al. 1996] Huttner, B., A. Muller, J.D. Gautier, H. Zbinden, N. Gisin, Phys. Rev. A 54, 3783 (1996)

[Ivanovic 1987] Ivanovic, I.D., Phys. Lett. A 123, 257 (1987)

[Jaeger and Shimony 1995] Jaeger, G. A. Shimony, Phys. Lett. A 197, 83 (1995)

[Kempe et al. 2000] Kempe, J., C. Simon, G. Weihs, Phys. Rev. A 62, 032302 (2000)

[Koenig et al. 2008] Koenig, R., R. Renner, C. Schaffner, arXiv:0807.1338 (2008)

[Kofler et al. 2006] Kofler, J., T. Paterek, C. Brukner, Phys. Rev. A 73, 022104 (2006)

[Landau 1988] Landau, L., Found. Phys. 18, 449 (1988)

[Le Bellac 2006] Le Bellac, M., A Short Introduction to Quantum Information and Quantum Computation (Cambridge University Press, Cambridge, 2006)

[Leggett 2003] Leggett, A.J., Found. Phys. 33, 1469 (2003)

[Li et al. 2009] Li, K., A. Winter, X. Zou, G. Guo, arXiv:0903.4308 (2009)

[Lindblad 1976] Lindblad, G., Comm. Math. Phys. 48, 119 (1976)

[Linden et al. 2007] Linden, N., S. Popescu, A.J. Short, A. Winter, Phys. Rev. Lett. 99, 180502 (2007).

[Ling et al. 2008] Ling, A., M.P. Peloso, I. Marcikic, V. Scarani, A. Lamas-Linares, C. Kurtsiefer, Phys. Rev. A 78, 020301(R) (2008).

[Magniez et al. 2006] Magniez, F., D. Mayers, M. Mosca, H. Ollivier, Self-testing of Quantum Circuits. In Proceedings of ICALP2006, Part I, M. Bugliesi et al. (Eds.), Lecture Notes in Computer Science 4051, pp. 72-83 (2006) [quant-ph/0512111].

[Mandel 1983] Mandel, L., Nature 304, 188 (1983)

[Masanes 2003] Masanes, L., quant-ph/0309137 (2003)

[Masanes et al. 2006] Masanes, L., A. Ac??n, N. Gisin, Phys. Rev. A. 73, 012112 (2006)

[Masanes 2008] Masanes, L., arXiv:0807.2158 (2008)

[Masanes et al. 2008] Masanes, L., Y.-C. Liang, A. Doherty, Phys. Rev. Lett. 100, 090403 (2008)

[Mayers and Yao 2004] Mayers, D., A. Yao, Quant. Inf. Comput., 4, 273 (2004).

[Mermin 1990] Mermin, N.D., Am. J. Phys. 58, 731 (1990).

[Milonni and Hardies 1982] Milonni, P.W., M.L. Hardies, Phys. Lett. 92A, 321 (1982)

[Navascues et al. 2008] Navascues, M., S. Pironio, A. Ac??n, New J. Phys. 10, 073013 (2008)

[Navascues and Wunderlich 2009] Navascues, M., H. Wunderlich, arXiv:0907.0372 (2009)

[Nielsen and Chuang 2000] Nielsen, M., I. Chuang, Quantum Computation and Quantum Information (Cambridge University Press, Cambridge, 2000)

[Pan et al. 2008] Pan, J.-W., Z.-B. Chen, M. Żukowski, H. Weinfurter, A. Zeilinger, arXiv:0805.2853 (2008)

[Pawłowski et al. 2009] Pawłowski, M., T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, M. Z? ukowski, Nature 461, 1101 (2009)

[Peres 1988] Peres, A., Phys. Lett. A 128, 19 (1988)

[Peres 1995] Peres, A., Quantum Theory: Concepts and Methods (Kluwer, Dordrecht, 1995)

[Peres 2002] Peres, A., quant-ph/0205076 (2002)

Pironio 2005] Pironio, S., J. Math. Phys. 46, 062112 (2005)

[Pironio et al. 2009] Pironio, S., A. Ac??n, N. Brunner, N. Gisin, S. Massar, V. Scarani, New J. Phys. 11, 045021 (2009).

[Pitowski and Svozil 2001] Pitowski, I., K. Svozil, Phys. Rev. A 64, 014102 (2001)

[Popescu and Rohrlich 1992] Popescu, S., D. Rohrlich, Phys. Lett. A 166, 293 (1992)

[Popescu and Rohrlich 1994] Popescu, S., D. Rohrlich, Found. Phys. 24, 379 (1994)

[Popescu 1995] Popescu, S., Phys. Rev. Lett. 74, 2619 (1995)

[Raynal and Lütkenhaus 2005] Raynal, P., N. Lu?tkenhaus, Phys. Rev. A 72, 022342 (2005)

[Renner 2007] Renner, R., Nature Physics 3, 645 (2007)

[Reuse 1984] Reuse, F., Ann. Phys. 154, 161 (1984)

[Sakurai 1993] Sakurai, J.J., Modern Quantum Mechanics (Addison Wesley, revised edition 1993)

[Salart, Baas, van Houwelingen et al. 2008] Salart, D., A. Baas, J. van Houwelingen, N. Gisin, H. Zbinden, Phys. Rev. Lett. 100, 220404 (2008)

[Salart, Baas, Branciard et al. 2008] Salart, D., A. Baas, C. Branciard, N. Gisin, H. Zbinden, Nature 454, 861 (2008)

[Sangouard et al. 2009] Sangouard, N., C. Simon, H. De Riedmatten, N. Gisin, arXiv:0906.2699 (2009)

[Scarani et al. 2000] Scarani, V., W. Tittel, H. Zbinden, N. Gisin, Phys. Lett. A 276, 1 (2000)

[Scarani et al. 2002] Scarani, V., M. Ziman, P. Štelmachovi?c, N. Gisin, V. Bŭzek, Phys. Rev. Lett. 88, 090705 (2002)

[Scarani and Gisin 2002] Scarani, V., N. Gisin, Phys. Lett. A 295, 167 (2002)

[Scarani et al. 2005] Scarani, V., S. Iblisdir, N. Gisin, A. Aćin, Rev. Mod. Phys. 77, 1225-1256 (2005)

[Scarani et al. 2006] Scarani, V., N. Gisin, N. Brunner, Ll. Masanes, S. Pino, A. Aćin, Phys. Rev. A 74, 042339 (2006)

[Scarani and Méthot 2007] Scarani, V., A.A. Méthot, Quantum Inf. Comput. 7, 157 (2007)

[Scarani et al. 2009] Scarani, V., H. Bechmann-Pasquinucci, N.J. Cerf, M. Dŭsek, N. Lütkenhaus, M. Peev, Rev. Mod. Phys. 81, 1301 (2009)

[Scarani and Kurtsiefer 2009] Scarani, V., C. Kurtsiefer, arXiv:0909.2601 (2009)

[Short and Barrett 2009] Short, A.J., J. Barrett, arXiv:0909.2601 (2009)

[Short et al. 2006] Short, A.J., N. Gisin, S. Popescu, Phys. Rev. A 73, 012101 (2006)

[Simon et al. 2000] Simon, C., G. Weihs, A. Zeilinger, Phys. Rev. Lett 84, 2993 (2000)

[Skrzypczyk et al. 2009] Skrzypczyk, P., N. Brunner, S. Popescu, Phys. Rev. Lett. 102, 110402 (2009).

[Smith and Yard 2008] Smith, G., J. Yard, Science 321, 1812 (2008)

[Smith and Smolin 2009] Smith, G., J.A. Smolin, Phys. Rev. Lett. 102, 010501 (2009)

[Stefanov et al. 2002] Stefanov, A., H. Zbinden, N. Gisin, A. Suarez, Phys. Rev. Lett. 88, 120404 (2002)

[Stefanov et al. 2003] Stefanov, A., H. Zbinden, N. Gisin, A. Suarez, Phys. Rev. A 67, 042115 (2003)

[Stucki et al. 2005] Stucki, D., N. Brunner, N. Gisin, V. Scarani, H. Zbinden, Appl. Phys. Lett. 87, 194108 (2005)

[Suarez and Scarani 1997] Suarez, A., V. Scarani, Phys. Lett. A 232, 9 (1997)

[Tittel and Weihs 2001] Tittel, W., G. Weihs, Quantum Inf. Comput. 1, 3 (2001)

[Toner and Bacon 2003] Toner, B.F., D. Bacon, Phys. Rev. Lett. 91, 187904 (2003)

[Tsirelson 1987] Tsirelson, B., J. Sov. Math. 36, 557 (1987)

[van Dam 2005] van Dam, W., quant-ph/0501159 (2005)

[Vértesi 2008] Vértesi, T., Phys. Rev. A 78, 032112 (2008)

[Vértesi and Pál 2009] Vértesi, T., K.F. Pál, Phys. Rev. A 79, 042106 (2009)

[Vértesi, Pironio and Brunner 2009] Vértesi, T., S. Pironio, N. Brunner, arXiv:0909.3171 (2009)

[Wehner et al. 2008] Wehner, S., M. Christandl, A.C. Doherty, Phys. Rev. A 78, 062111 (2008)

[Werner 1989] Werner, R.F., Phys. Rev. A 40, 4277 (1989)

[Werner 1998] Werner, R.F., Phys. Rev. A 58, 1827 (1998)

[Wootters and Żurek 1982] Wootters, W.K., W.H. Żurek, Nature 299, 802(1982)

[Yurke and Stoler, 1992] Yurke, B., D. Stoler, Phys. Rev. A 46, 2229 (1992)

[Zbinden et al. 2001] Zbinden, H., J. Brendel, N. Gisin, W. Tittel, Phys. Rev. A 63, 022111 (2001)

[Żukowski et al. 1993] Żukowski, M., A. Zeilinger, M.A. Horne, A.K. Ekert, Phys. Rev. Lett. 71, 4287 (1993)