Chapter 1

Open Quantum Systems

In quantum mechanics, the state of a system is represented by a vector $|\psi\rangle$ in a Hilbert space. The wavefunction $|\psi\rangle$ evolves through unitary operations and is finally subjected to a projective measurement to obtain experimental predictions. This description is not satisfactory to tackle real world quantum systems that are unavoidably coupled to other degrees of freedom, the so-called «environment». A full quantum description of the system and its environment is in general not possible and would be of little interest. This has lead to the development of the notion of «open quantum systems», where one considers the behavior of a small quantum system embedded in a larger Hilbert space. The goal of this chapter is to review some results that have been obtained for such systems and are widely used in quantum information.

We introduce this chapter by considering a basic example to illustrate how the coupling to an environment may drastically modify the evolution of a simple quantum system in the form of a single qubit. Let us first consider that the qubit evolves unitarily as an isolated system through [1]

$$|\uparrow\rangle \xrightarrow{U_0} \frac{1}{\sqrt{2}}\left(|\uparrow\rangle + |\downarrow\rangle\right) \xrightarrow{U_\phi} \frac{1}{\sqrt{2}}\left(|\uparrow\rangle + e^{i\phi}|\downarrow\rangle\right)$$
$$\xrightarrow{U_0} \frac{1}{2}\left[\left(|\uparrow\rangle + |\downarrow\rangle\right) + e^{i\phi}\left(-|\uparrow\rangle + |\downarrow\rangle\right)\right] = e^{i\phi/2}\left(-i\sin\frac{\phi}{2}|\uparrow\rangle + \cos\frac{\phi}{2}|\downarrow\rangle\right).$$

This sequence of unitary operations realizes a Ramsey interferometer, where U_0 corresponds to a $\pi/2$ pulse and U_ϕ to free evolution. By measuring the final spin in the $|\uparrow\rangle$, $|\downarrow\rangle$ basis, we observe Ramsey fringes: the probability to obtain the spin up outcome is a sine function of ϕ. We now modify this sequence and add a coupling to the environment in between the first $\pi/2$ pulse and the free evolution period. We suppose that the spin and the environment, which is initially in $|0\rangle_E$, evolve unitarily such that $|\downarrow\rangle|0\rangle_E \to |\downarrow\rangle|e_\downarrow\rangle_E$ and $|\uparrow\rangle|0\rangle_E \to |\uparrow\rangle|e_\uparrow\rangle_E$. The full sequence becomes

$$|\uparrow\rangle|0\rangle_E \xrightarrow{\tilde{U}_0} \frac{1}{\sqrt{2}}\left(|\uparrow\rangle|e_\uparrow\rangle_E + |\downarrow\rangle|e_\downarrow\rangle_E\right) \xrightarrow{U_\phi} \frac{1}{\sqrt{2}}\left(|\uparrow\rangle|e_\uparrow\rangle_E + e^{i\phi}|\downarrow\rangle|e_\downarrow\rangle_E\right)$$
$$\xrightarrow{U_0} \frac{1}{2}\left[\left(|\uparrow\rangle + |\downarrow\rangle\right)|e_\uparrow\rangle_E + e^{i\phi}\left(-|\uparrow\rangle + |\downarrow\rangle\right)|e_\downarrow\rangle_E\right] = \frac{1}{\sqrt{2}}\left[|\uparrow_x\rangle|e_\uparrow\rangle_E - e^{i\phi}|\downarrow_x\rangle|e_\downarrow\rangle_E\right].$$

[1] The two unitary operations U_0 and U_ϕ are given in matrix form by

$$U_0 = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix} \quad U_\phi = \begin{pmatrix}1 & 0\\ 0 & e^{i\phi}\end{pmatrix}.$$

The final measurement can be decomposed as a projective measurement of the environment followed by a projective measurement of the spin. There are two extreme cases depending on the angle between the two environment states $|e_\downarrow\rangle_E$ and $|e_\uparrow\rangle_E$. If the states are identical, then the environment state can be factored out and the projective measurement of the environment does not modify the state of the spin: Ramsey oscillations with full contrast are still observed.

But, if the environment states are orthogonal and therefore perfectly distinguishable, the first measurement prepares the spin in $|\uparrow_x\rangle = (|\uparrow\rangle + |\downarrow\rangle)/\sqrt{2}$, if the environment is projected in $|e_\uparrow\rangle_E$, or in $|\downarrow_x\rangle = (|\uparrow\rangle - |\downarrow\rangle)/\sqrt{2}$ otherwise. Both outcomes happen with probability 1/2. The final measurement then leads to spin up and down outcomes with probability 1/2, independently of the value of ϕ, and the Ramsey oscillations are washed out. This exemplifies some important features of an open quantum system:

- Once the environment is measured, the state of the system is, in general, in a mixture of states. It cannot be represented as a ket of the Hilbert space but by an operator called density matrix.

- The evolution of a quantum system in the presence of an environment is, in general, not unitary and is called a quantum map. Here, we have obtained a map where the coherence is completely lost: the initial pure state $|\uparrow\rangle$ is mapped onto a fully mixed state. This happens because the qubit and the environment have become maximally entangled. This loss of coherence of the qubit is irreversible. There is no quantum operation acting on the qubit alone that can restore the coherence and the fringe contrast.

In the following, we build on this basic example to deduce some properties of density matrices and quantum maps. Here, we have supposed that the coupling to the environment happens at once. But in general, the system and the environment are always coupled and decoherence happens continuously. We will see how to obtain the master equation that governs the evolution of the density matrix in this situation. In the last section, we will show how a quantum map may also be interpreted as a generalized measurement.

1.1 The density matrix

We consider a quantum system in an Hilbert space $\mathscr{H}_S$. It is coupled to an environment, and the total Hilbert space is $\mathscr{H}_S \otimes \mathscr{H}_E$. A generic state $|\psi\rangle$ of the system and the environment can be written as

$$|\psi\rangle = \sum_{i,\mu} a_{i\mu} |i\rangle |\mu\rangle_E$$

where $|i\rangle$ and $|\mu\rangle_E$ form a basis of $\mathscr{H}_S$ and $\mathscr{H}_E$. Because we are only interested in the properties of the system S, we wish to compute the expectation value of observables $M \otimes \mathbb{1}_E$ that act only on $\mathscr{H}_S$

$$\langle M \rangle = \langle \psi | M \otimes \mathbb{1}_E | \psi \rangle = \sum_{i,j,\mu} a_{i\mu} a^*_{j\mu} \langle j | M | i \rangle$$

Defining the density matrix as the partial trace over the environment of the operator $|\psi\rangle \langle\psi|$

$$\rho = \mathrm{Tr}_E[|\psi\rangle \langle\psi|], \tag{1.1}$$

we obtain $\rho = \sum_{ij\mu} a_{i\mu} a^*_{j\mu} |i\rangle\langle j|$, and the expectation value of M can be written as $\langle M \rangle = \mathrm{Tr}[\rho M]$. Introducing the unnormalized states $|\phi_\mu\rangle = \sum_i a_{i\mu} |i\rangle$, we can rewrite $\rho = \sum_\mu |\phi_\mu\rangle\langle\phi_\mu|$. From this definition of ρ, we see that the density matrix has the following properties:

- ρ is a positive semidefinite operator, because ρ is hermitian and for any state $|\varphi\rangle$ of the system, one has $\langle\varphi|\rho|\varphi\rangle = \sum_\mu |\langle\varphi|\phi_\mu\rangle|^2 \geq 0$,
- ρ has unit trace, because $\mathrm{Tr}[\rho] = \sum_{i,\mu} |a_{i\mu}|^2 = 1$.

These properties imply that ρ can be diagonalized as

$$\rho = \sum_\mu p_\mu |\varphi_\mu\rangle\langle\varphi_\mu| \tag{1.2}$$

where the $|\varphi_\mu\rangle$ are orthogonal states and the p_μ are probabilities, $0 \leq p_\mu \leq 1$, with the normalization condition $\sum_\mu p_\mu = 1$. This form of ρ is called a statistical mixture, meaning that the system can be considered to be in state $|\varphi_\mu\rangle$ with probability p_μ.

1.1.1 Purification of a density matrix

Let us now forget about the environment and consider that the system is described by a density matrix ρ, that we define as an hermitian operator on $\mathscr{H}_S$ fulfilling the properties that we just derived. Can such a matrix ρ can be written as a trace over a pure state in a larger space as in (1.1)? The answer is yes and is called a purification of ρ. Such a purification can be constructed as follows. Starting from (1.2), we define the state $|\psi\rangle$ as

$$|\psi\rangle = \sum_\mu \sqrt{p_\mu} |\varphi_\mu\rangle |\mu\rangle_E$$

where $|\mu\rangle_E$ are orthogonal states chosen in a Hilbert space whose dimension is larger or equal to the number of non-zero eigenvalues p_μ. We then obtain $\mathrm{Tr}_E[|\psi\rangle\langle\psi|] = \sum_\mu p_\mu |\varphi_\mu\rangle\langle\varphi_\mu| = \rho$. We see that such a purification is not unique (see exercise).

1.2 Quantum map

We are now going to extend the construction of the previous section to describe the evolution of an open quantum system as resulting from the unitary evolution of a global state, including the environment, followed by partial trace. By construction, this operation is a linear transformation, called a quantum map or channel, which maps a density matrix to another valid density matrix. We thus define the action of a quantum map Λ on a density matrix ρ as

$$\rho \xrightarrow{\Lambda} \mathrm{Tr}_E[U(\rho \otimes \omega_E)U^\dagger] \tag{1.3}$$

where U is a unitary operator on $\mathscr{H}_S \otimes \mathscr{H}_E$ and ω_E is the density matrix describing the initial state of the environment. In the following we will consider ω_E to be a pure state $\omega_E = |0\rangle\langle 0|_E$ without loss of generality. The important hypothesis is that the system and the environment are initially in a product state [1, 2].

1.2.1 Kraus form and dilation

In order to obtain the action of the map in terms of operators acting only on ρ, we decompose the unitary matrix as a block matrix

$$U = \sum_{\mu\nu} U_{\mu\nu} \otimes |\mu\rangle \langle\nu|_E \qquad U = \begin{pmatrix} U_{00} & U_{01} & \dots \\ U_{10} & \ddots & \\ \vdots & & \end{pmatrix}$$

where the $U_{\mu\nu}$ are operators acting in $\mathscr{H}_S$ and the vectors $|\mu\rangle$ form a basis of $\mathscr{H}_E$. The action of Λ can then be written as

$$\rho \xrightarrow{\Lambda} \sum_{\mu\mu_1\mu_2\nu_1\nu_2} U_{\mu_1\nu_1}\, \rho\, U^\dagger_{\mu_2\nu_2} \langle\mu|\mu_1\rangle \langle\nu_1|0\rangle \langle 0|\nu_2\rangle \langle\mu_2|\mu\rangle = \sum_\mu U_{\mu 0}\, \rho\, U^\dagger_{\mu 0} \tag{1.4}$$

Because U is unitary, we have that $\langle 0|\, U^\dagger U\, |0\rangle_E = \mathbb{1}_S$ and thus

$$\sum_\mu U^\dagger_{\mu 0} U_{\mu 0} = \mathbb{1} \tag{1.5}$$

The two equations (1.4) and (1.5) define a map in the Kraus form, as first introduced by Kraus in 1971. The operators $U_{\mu 0}$ are generally not unitary, except when there is only one Kraus operator in the decomposition. If the decomposition contains a single unitary operator, it corresponds to a unitary evolution of the system. This is the only case of reversible map. If there are more than one Kraus operator in the decomposition, the map describes an irreversible evolution. Each term of the decomposition corresponds to the environment being in a given final state.

A map defined in the Kraus form can be recasted as the contraction of a unitary evolution in a lager Hilbert space. From the knowledge of the $U_{\mu 0}$ operators, we construct the first block column of the U matrix. Because of (1.5), the column vectors forming this rectangular matrix are orthogonal and normed to one. We can extend this set to obtain a full basis of $\mathscr{H}_S \otimes \mathscr{H}_E$ and thus a unitary matrix U. In order to do so, the minimal dimension of the environment space must be at least the number of Kraus operators $U_{\mu 0}$ that define the map. This operation of expressing a map in terms of a unitary evolution in a larger space followed by partial trace is called a dilation.

In the Kraus form, it is easy to see that Λ preserves the trace and the positivity of ρ. Such a map is called a positive trace preserving map. But Λ satisfies an even stronger property: it is completely positive. This means that by extending the map to a larger Hilbert space and defining the action of the extended map as $\mathbb{1} \otimes \Lambda$, we obtain a map that is also positive. This is physically justified by considering the fact that a map acting on a system while doing nothing to any other system should map the density matrix of the global system to a valid (positive) density matrix. This property is important because there exists maps that are positive but not completely positive and thus not physical. In the next chapter, we will see an example of such a map and how it can be used to detect entanglement.

Let us check that a map in the Kraus form is completely positive, the action of the extended map $\mathbb{1} \otimes \Lambda$ on a density matrix ρ is

$$(\mathbb{1} \otimes \Lambda)[\rho] = \sum_\mu (\mathbb{1} \otimes K_\mu)\rho(\mathbb{1} \otimes K^\dagger_\mu) = \sum_\mu [(\mathbb{1} \otimes K_\mu)\sqrt{\rho}][\sqrt{\rho}(\mathbb{1} \otimes K^\dagger_\mu)]$$

where the K_μ's are the Kraus operators defining Λ and $\sqrt{\rho}$ is well defined because ρ is positive. The last expression is of the from $\sum_\mu A_\mu A^\dagger_\mu$ and is therefore positive.

1.2.2 Choi matrix

We have arrived at the conclusion that tracing over the environment after unitary evolution generates maps that are Completely Positive Trace Preserving (CPTP) linear applications. We can now wonder if the converse is true: can any CPTP application be written as (1.3)? The answer is yes as explicitly shown by Choi in 1975 [3]. Choi first showed that any CPTP application can be written in Kraus form, which can then be dilated to an unitary evolution in a larger space.

The main observation of Choi is that a CPTP map Λ can be represented by a matrix which is positive. There are different ways to represent a map by a single matrix. In a Hilbert space of dimension n, a density matrix contains n^2 terms, and a map must therefore be represented by a $n^2 \times n^2$ complex matrix. After having chosen a basis $|i\rangle$ of the Hilbert space, such a matrix can be obtained, for example, by flattening the density matrix as a vector $\vec{\rho}$ and looking for the matrix $\mathscr{M}$ such that the action of the map is obtained by the matrix product $\mathscr{M}.\vec{\rho}$. This representation is often used in numerical calculations. But other choices are possible, such as the Choi matrix C_Λ, which is defined as the block matrix where the (i,j) block is equal to $\Lambda[|i\rangle\langle j|]$

$$C_\Lambda = \sum_{i,j} |i\rangle\langle j| \otimes \Lambda[|i\rangle\langle j|]. \tag{1.6}$$

The C_Λ matrix can also be seen as the matrix obtained by acting with the extended map $\mathbb{1} \otimes \Lambda$ on the density matrix $|\psi_+\rangle\langle\psi_+|$, where $|\psi_+\rangle$ is the unnormalized state $\sum_i |i\rangle \otimes |i\rangle$.

Because the matrix is constructed block-wise rather than column-wise, the action of the map is not just a matrix product [4]. We will now show that $\Lambda[\rho] = \mathrm{Tr}_1[(\rho^T \otimes \mathbb{1})C_\Lambda]$, where Tr_1 is the partial trace over the first (leftmost) space of the tensor product. We have

$$\begin{aligned}
\mathrm{Tr}_1[(\rho^T \otimes \mathbb{1})C_\Lambda] &= \sum_{i,j} \mathrm{Tr}_1[(\rho^T \otimes \mathbb{1})(|i\rangle\langle j| \otimes \Lambda[|i\rangle\langle j|])] \\
&= \sum_{i,j} \mathrm{Tr}_1[\rho^T |i\rangle\langle j| \otimes \Lambda[|i\rangle\langle j|]] \\
&= \sum_{i,j,k} \langle k| \rho^T |i\rangle \langle j|k\rangle \Lambda[|i\rangle\langle j|] \\
&= \sum_{i,j} \rho_{ij} \Lambda[|i\rangle\langle j|] = \Lambda[\rho].
\end{aligned} \tag{1.7}$$

Supposing that Λ is CPTP, the matrix C_Λ, which is obtained by acting on the positive operator $|\psi_+\rangle\langle\psi_+|$ with the extended map $\mathbb{1} \otimes \Lambda$, must be positive. Therefore, it can be diagonalized as $C_\Lambda = \sum_\mu \lambda_\mu |\psi_\mu\rangle\langle\psi_\mu|$ with real positive eigenvalues $\lambda_\mu \geq 0$. Defining $|\phi_\mu\rangle = \sqrt{\lambda_\mu}\,|\psi_\mu\rangle$, we obtain $C_\Lambda = \sum_\mu |\phi_\mu\rangle\langle\phi_\mu|$. Each $|\phi_\mu\rangle$ can be decomposed as $|\phi_\mu\rangle = \sum_{ij} \alpha_{ij}^{(\mu)} |i\rangle \otimes |j\rangle$, we then introduce the operator $K_\mu = \sum_{i,j} \alpha_{ij}^{(\mu)} |j\rangle\langle i|$. The action of $\mathbb{1} \otimes K_\mu$ on $|\psi_+\rangle$ is given by

$$\begin{aligned}
(\mathbb{1} \otimes K_\mu)|\psi_+\rangle &= \sum_{ijk} \alpha_{ij}^{(\mu)} (\mathbb{1} \otimes |j\rangle\langle i|)(|k\rangle \otimes |k\rangle) \\
&= \sum_{ij} \alpha_{ij}^{(\mu)} |i\rangle \otimes |j\rangle \\
&= |\phi_\mu\rangle\,.
\end{aligned}$$

Using (1.7), we can now write the action of the map as

$$\begin{aligned}
\Lambda[\rho] &= \sum_{\mu} \mathrm{Tr}_1[(\rho^T \otimes \mathbb{1})\,|\phi_\mu\rangle\langle\phi_\mu|] \\
&= \sum_{\mu} \mathrm{Tr}_1[(\rho^T \otimes \mathbb{1})(\mathbb{1} \otimes K_\mu)\,|\psi_+\rangle\langle\psi_+|\,(\mathbb{1} \otimes K_\mu^\dagger)] \\
&= \sum_{ij\mu} \mathrm{Tr}_1[(\rho^T \otimes \mathbb{1})(\mathbb{1} \otimes K_\mu)(|i\rangle\langle j| \otimes |i\rangle\langle j|)(\mathbb{1} \otimes K_\mu^\dagger)] \\
&= \sum_{ij\mu} \mathrm{Tr}_1[\rho^T\,|i\rangle\langle j| \otimes K_\mu\,|i\rangle\langle j|\,K_\mu^\dagger] \\
&= \sum_{ij\mu} \langle j|\,\rho^T\,|i\rangle\,K_\mu\,|i\rangle\langle j|\,K_\mu^\dagger \\
&= \sum_{\mu} K_\mu \rho K_\mu^\dagger,
\end{aligned}$$

which is in Kraus form. A short calculation shows that the preservation of the trace $\mathrm{Tr}[\Lambda[|i\rangle\langle j|]] = \delta_{ij}$ translates into $\sum_\mu K_\mu^\dagger K_\mu = \mathbb{1}$.

This leads us to the main conclusion of this chapter. The most generic evolution of an open quantum is a linear map that is completely positive and preserves the trace. We have shown that such a map can always be written as a Kraus form. We note that the decomposition in Kraus operators of a quantum map is not unique. Alternatively, the map can also be «dilated» to a unitary evolution in a larger Hilbert space, followed by a partial trace. Again this dilation is not unique. Because the existence of the dilation can be interpreted as the consequence of a more general theorem due to Stinespring, it is often referred to as a Stinespring's dilation.

1.2.3 Example of quantum maps for a qubit

Common quantum maps for a single qubit are obtained by considering Kraus forms with two Kraus operators

$$K_0 = \sqrt{1-p}\,\mathbb{1} \quad K_1 = \sqrt{p}\,\sigma_i \quad \text{with} \quad 1 \geq p \geq 0$$

where σ_i is one of the three Pauli matrices. Because $\sigma_i^2 = \mathbb{1}$, we have that $K_0^\dagger K_0 + K_1^\dagger K_1 = \mathbb{1}$. The associated map acts on the density matrix of the qubit as

$$\rho \xrightarrow{\Lambda} (1-p)\rho + p\,\sigma_i \rho \sigma_i$$

These channels can be seen as inducing bit flip errors if $i = X$, phase flip error if $i = Z$ and both phase and bit flip error if $i = Y$ with an error probability given by p. Note that if $p = 1$, then the error is certain and is actually not an error anymore. In other words, if $p = 0$ or $p = 1$ the set of Kraus operators contains a single element and the evolution is thus unitary and reversible. Otherwise, the evolution is irreversible.

1.3 Master equation in the Lindblad form

In the case of a closed system, the integration of the Schrödinger equation between an initial and a final time leads to a unitary evolution operator. We can now wonder which differential

equation leads to quantum maps when it is integrated over time. By analogy to statistical mechanics, this equation is called a master equation, which we write formally as

$$\frac{d\rho(t)}{dt} = \mathscr{L}[\rho(t)]$$

where $\mathscr{L}$ is a linear operator acting on ρ. This is not the most general evolution equation for ρ, because the term on the right hand side only depends on ρ at time t and not at previous times, which corresponds to a Markovian hypothesis. In 1976, Lindblad, and independently Gorini, Kossakowski and Sudarshan, have shown that the dynamic generated by such a Markovian equation corresponds to a quantum map if $\mathscr{L}$ takes a special form, which is now often called the Lindblad form. The converse is also true, any Markovian equation in the Lindblad form leads to a quantum map evolution.

1.3.1 Phenomenological derivation

In order to obtain the form of $\mathscr{L}$, we consider that the evolution of ρ during an infinitesimal time dt is a quantum map that we represent by its Kraus decomposition K_0, K_1, ... We suppose that K_0 is close to the identity, $K_0 = \mathbb{1} - L_0$, and that the other K_i correspond to events with a small probability $K_i = \sqrt{\epsilon_i} A_i$, where A_i is of order unity and $\epsilon_i \ll 1$. The normalization condition for the Kraus decomposition gives

$$L_0^\dagger + L_0 = \sum_i \epsilon_i A_i^\dagger A_i$$

where we neglect the $L_0^\dagger L_0$ term which is of higher order. We can thus write

$$L_0 = \frac{1}{2} \sum_i \epsilon_i A_i^\dagger A_i + i\tilde{H}$$

where $\tilde{H}$ is an hermitian operator. The density matrix at $t + dt$ is then given by

$$\begin{aligned} \rho(t+dt) &= \rho(t) - L_0\rho(t) - \rho(t)L_0^\dagger + \sum_i \epsilon_i A_i \rho(t) A_i^\dagger \\ &= \rho(t) - i[\tilde{H}, \rho(t)] + \sum_i \epsilon_i \left(-\frac{1}{2} A_i^\dagger A_i \rho(t) - \frac{1}{2} \rho(t) A_i^\dagger A_i + A_i \rho(t) A_i^\dagger \right) \end{aligned} \tag{1.8}$$

We now suppose that $\epsilon_i = \gamma_i dt$ where γ_i is the rate for the event associated to the decoherence channel i, we also define $H = \hbar\tilde{H}/dt$ and obtain the following master equation

$$\frac{d\rho}{dt} = \frac{1}{i\hbar}[H, \rho(t)] + \sum_i \gamma_i \left(A_i \rho(t) A_i^\dagger - \frac{1}{2} \{A_i^\dagger A_i, \rho(t)\} \right) \tag{1.9}$$

The first term corresponds to the unitary evolution induced by the Hamiltonian H. The right terms induce a non-unitary evolution and are said to be in the Lindblad form, where the A_i are called jump operators. The key hypothesis to obtain this master equation is to assume that the expression (1.8) is valid at any time t, which corresponds to a Markovian hypothesis.

1.3.2 Microscopic derivation

We now follow a different and more physically motivated derivation of the Lindblad form. We consider that the quantum system is coupled to an environment, which, in this context, is often called a bath. The system and the bath evolve with the total Hamiltonian $H = H_S + H_B + H_{SB}$, where H_S is the Hamiltonian of the system, H_B the Hamiltonian of the bath, and H_{SB} couples the system and the bath. We apply a unitary transform to go in the interaction picture and look for the equation of motion of the system density matrix $\rho_S(t)$. The Hamiltonian in the interaction picture is given by $H_I(t) = U^\dagger(t) H_{SB} U(t)$ where $U(t) = \exp(iH_S t) \otimes \exp(iH_B t)$. The total density matrix evolves as

$$\frac{d\rho}{dt} = \frac{1}{i\hbar}[H_I(t), \rho]$$

which can be formally integrated as

$$\rho(t) = \rho(0) + \frac{1}{i\hbar}\int_0^t [H_I(s), \rho(s)]\, ds$$

Inserting back this expression in the evolution equation and tracing over the bath, one obtains

$$\frac{d\rho_S}{dt} = -\frac{1}{\hbar^2}\int_0^t \mathrm{Tr}_B[H_I(t), [H_I(s), \rho(s)]]\, ds + \frac{1}{i\hbar}\mathrm{Tr}_B[H_I(t), \rho(0)]$$

Without any loss of generality, we can suppose that the last term is zero. Changing the integration variable from s to $t - s$ in the first term, we obtain

$$\frac{d\rho_S}{dt} = -\frac{1}{\hbar^2}\int_0^t \mathrm{Tr}_B[H_I(t), [H_I(t-s), \rho(t-s)]]\, ds \tag{1.10}$$

Born approximation

We now suppose that the coupling to the bath is weak and perform a first order expansion in time dependent perturbation theory of the evolution equation. We thus replace $\rho(t)$ in the integral by its value in the absence of bath coupling and obtain

$$\frac{d\rho_S}{dt} = -\frac{1}{\hbar^2}\int_0^t \mathrm{Tr}_B[H_I(t), [H_I(t-s), \rho_S(t) \otimes \rho_B]]\, ds \tag{1.11}$$

where ρ_B is the density matrix of the bath, which is supposed to be time independent. Note that ρ_S is evaluated at t and not $t-s$ as in (1.10), which is part of the Born approximation.

Markovian approximation

The equation (1.11) is an integro-differential equation, which is very difficult to handle. In order to obtain a first order differential equation, we use the short bath memory, or Markov approximation. We suppose that the integrand rapidly drops to zero when s increases and extend the upper bound of the integral to ∞

$$\frac{d\rho_S}{dt} = -\frac{1}{\hbar^2}\int_0^\infty \mathrm{Tr}_B[H_I(t), [H_I(t-s), \rho_S(t) \otimes \rho_B]]\, ds \tag{1.12}$$

For convenience, we rewrite this equation as

$$\frac{d\rho_S}{dt} = \frac{1}{\hbar^2}\int_0^\infty \mathrm{Tr}_B[H_I(t-s)(\rho_S(t) \otimes \rho_B)H_I(s) - H_I(t)H_I(t-s)(\rho_S(t) \otimes \rho_B)]\, ds + \mathrm{hc} \tag{1.13}$$

Secular approximation

We now suppose that the system bath coupling may be written $H_{SB} = A \otimes B$, where A and B are hermitian operators acting on the system and the bath. We denote by $|\varepsilon\rangle$ the eigenstates of H_S and expand A as $A = \sum_{\varepsilon,\varepsilon'} A_{\varepsilon\varepsilon'} |\varepsilon\rangle\langle\varepsilon'|$. In the interaction frame, we obtain

$$A(t) = \sum_{\varepsilon\,\varepsilon'} A_{\varepsilon\varepsilon'} |\varepsilon\rangle\langle\varepsilon'| \, e^{-i(\varepsilon-\varepsilon')t/\hbar} = \sum_{\omega} A_\omega e^{-i\omega t}$$

where the sum over ω is over all the possible transition frequencies $(\varepsilon - \varepsilon')/\hbar$ that appear in the spectrum of H_S. The operators A_ω are in general not hermitian. We now insert $H_I(t-s) = \sum_\omega e^{-i\omega(t-s)} A_\omega \otimes B(t-s)$ and $H_I(t) = \sum_\omega e^{i\omega t} A_\omega^\dagger \otimes B^\dagger(t)$ in (1.13) and arrive to

$$\frac{d\rho_S}{dt} = \frac{1}{\hbar^2} \int_0^\infty \sum_{\omega\,\omega'} e^{-i\omega(t-s)} e^{i\omega' t} \operatorname{Tr}_B[\, A_\omega \rho_S(t) A_{\omega'}^\dagger \otimes B(t-s)\rho_B B^\dagger(t) \\ - A_{\omega'}^\dagger A_\omega \rho_S(t) \otimes B^\dagger(t) B(t-s)\rho_B]\, ds + \text{hc}$$

We introduce the bath correlation function

$$\Gamma(\omega) = \frac{1}{\hbar^2} \int_0^\infty e^{i\omega s} \operatorname{Tr}_B[\, B^\dagger(t) B(t-s)\rho_B]\, ds$$

and obtain

$$\frac{d\rho_S}{dt} = \sum_{\omega\,\omega'} e^{-i(\omega-\omega')t}\, \Gamma(\omega) \left(A_\omega \rho_S A_{\omega'}^\dagger - A_{\omega'}^\dagger A_\omega \rho_S \right) + \text{hc} \tag{1.14}$$

The secular approximation consists in neglecting all the terms that rapidly oscillate in the sum over ω and ω' and keep only the constant terms, which leads to

$$\frac{d\rho_S}{dt} = \sum_{\omega} \Gamma(\omega) \left(A_\omega \rho_S A_\omega^\dagger - A_\omega^\dagger A_\omega \rho_S \right) + \text{hc} \tag{1.15}$$

We now separate the real and imaginary parts of the bath correlations as

$$\Gamma(\omega) = \frac{1}{2}\gamma(\omega) + i\delta(\omega)$$

The two functions $\gamma(\omega)$ and $\delta(\omega)$ verify the Kramers-Kronig relations. Inserting this expression of Γ in the last equation, we arrive at a master equation in the Lindblad form

$$\frac{d\rho_S}{dt} = \sum_{\omega} -i\delta(\omega)[A_\omega^\dagger A_\omega, \rho_S] + \gamma(\omega) \left(A_\omega \rho_S A_\omega^\dagger - \frac{1}{2}\{A_\omega^\dagger A_\omega, \rho_S\} \right)$$

The coupling to the bath has two effects:

- The Hamiltonian of the system is shifted by a quantity $\sum_\omega \hbar\delta(\omega) A_\omega^\dagger A_\omega$. This term is often called the Lamb shift term by analogy to the shift of the atomic levels that arises from the coupling of the atom to its electromagnetic environment.
- Jump operators appear, which induce transitions between the eigenstates of H_S. In the secular approximation, the energy exchanged between the system and the bath during a jump is perfectly defined. The rate associated to the jump where a quantum of energy $\hbar\omega$ is exchanged between the system and the bath is given by

$$\gamma(\omega) = \Gamma(\omega) + \Gamma^*(\omega) = \frac{1}{\hbar^2} \int_{-\infty}^\infty e^{i\omega s} \langle B^\dagger(t) B(t-s)\rangle\, ds$$

The probability for a jump in a time interval dt depends on $\gamma(\omega)$ and on the state of the system, it is given by $\gamma(\omega)dt\ \mathrm{Tr}_S[A_\omega^\dagger A_\omega \rho_S]$.

Jumps with positive ω correspond to energy flowing from the system to the bath, for example, a spontaneous emission event during which an atom emits a photon into vacuum. Jumps with negative ω corresponds to an absorption by the system of a quantum of energy coming from the bath. If the bath is at thermal equilibrium, the ratio of the absorption and emission rates is given by the Kubo-Martin-Schwinger formula

$$\gamma(-\omega) = e^{-\beta\omega}\gamma(\omega)\,,$$

where β is the inverse temperature of the bath. At zero temperature, as often considered in quantum optics, only emission processes are allowed and $\gamma(\omega < 0) = 0$.

The secular approximation is often the least justified of the three approximations that we used to obtain the Lindblad form, but it can be avoided. Other approaches, such as coarse graining, which do not use the secular approximation and therefore have a broader range of applicability, also lead to master equations in the Lindblad form (see for example [5]).

1.3.3 Example of master equations

Master equations in the Lindblad form are used extensively in quantum optics to describe light-matter coupling at the single photon level. For example, the evolution of a two-level atom under laser excitation is governed by the following master equation

$$\frac{d\rho}{dt} = -i\frac{\Omega}{2}[\sigma_+ + \sigma_-, \rho] + \gamma\left(\sigma_-\rho\sigma_+ - \frac{1}{2}\{\sigma_+\sigma_-, \rho\}\right)$$

where $\sigma_+ = |e\rangle\langle g|$ and $\sigma_- = |g\rangle\langle e|$. The first term describes the coupling of the atom dipole to the external laser field, which drives the atomic transition with Rabi frequency Ω, while the second term describes the spontaneous emission of the atom from the excited to the ground state with a rate γ. Here, the laser is considered to be at resonance with the atomic transition. This master equation is known as the optical Bloch equation and can be solved analytically.

The evolution of a qubit in the presence of continuous phase and bit flip errors is described by a master equation

$$\frac{d\rho}{dt} = -i[H, \rho] + \gamma_z\sigma_z\rho\sigma_z + \gamma_x\sigma_x\rho\sigma_x - (\gamma_z + \gamma_x)\rho$$

where γ_z (γ_x) is the phase (bit) flip error rate. In the case where H is sufficiently simple, this master equation is similar to the Bloch equation and can also be solved exactly.

Another well-known example of exactly solvable master equation is the one describing the evolution of a non-linear cavity under laser excitation [6]. The evolution of the density matrix of the resonator mode with annihilation operator a is given by

$$\frac{d\rho}{dt} = -i[-\delta a^\dagger a + \eta(a + a^\dagger) + U a^\dagger a^\dagger a a, \rho] + \kappa\left(a\rho a^\dagger - \frac{1}{2}\{a^\dagger a, \rho\}\right)$$

The term proportional to η describes the laser pump, which is detuned by δ from the cavity resonance, U sets the non-linearity, and κ is the cavity loss rate. The jump probability per unit time is given by κ times the mean photon number $\langle a^\dagger a\rangle$. Historically, this model was

introduced to describe an optical cavity filled with a Kerr medium. It has regained a lot of interest with the advent of the superconducting transmon qubit, which is well described by a non-linear oscillator.

Master equations are also getting more and more used in condensed matter physics to describe electronic transport in mesoscopic systems. Such a master equation takes the following general form

$$\frac{d\rho}{dt} = -i[H,\rho] + \sum_k \gamma_k^+ \left(c_k^\dagger \rho c_k - \frac{1}{2}\{c_k c_k^\dagger, \rho\} \right) + \sum_k \gamma_k^- \left(c_k \rho c_k^\dagger - \frac{1}{2}\{c_k^\dagger c_k, \rho\} \right)$$

where H is the Hamiltonian of the conductor. The two Lindblad terms describe charge injection (depletion) with rates γ_k^+ (γ_k^-) in the conductor from reservoirs at a given temperature and chemical potential.

1.3.4 Quantum trajectories

It is in general difficult to obtain an analytical solution of the master equation, and various methods have been developed in order to solve it numerically. Standard methods to solve coupled linear first order differential equations can be used, with the difficulty that the size of the system rapidly with the dimension of the Hilbert space. If d is the dimension of the space, the density matrix contains d^2 elements and the master equation is therefore a $d^2 \times d^2$ system. In order to avoid manipulating such large matrices, a widely used alternative is the Monte Carlo wave-function method [7]. To explain the method, we first rewrite (1.8) as

$$\rho(t+dt) = \rho(t) - \frac{i}{\hbar}(H_{\text{eff}}\rho(t) - \rho(t)H_{\text{eff}}^\dagger)dt + \sum_i \gamma_i A_i \rho(t) A_i^\dagger dt \tag{1.16}$$

where we introduce the effective non-hermitian Hamiltonian $H_{\text{eff}} = H - i\sum_i (\hbar\gamma_i/2) A_i^\dagger A_i$. We now decompose the density matrix at time t as a mixture of pure states $\rho(t) = \sum_\phi p_\phi |\phi(t)\rangle\langle\phi(t)|$. The evolution of one of the pure state $|\phi(t)\rangle$ from t to dt is obtained through the following algorithm:

- with a probability $p_i = \gamma_i dt \langle\phi(t)| A_i^\dagger A_i |\phi(t)\rangle$, we suppose that a quantum jump occurs and that $|\phi(t+dt)\rangle = A_i |\phi(t)\rangle / \| A_i |\phi(t)\rangle \|$
- with a probability $1 - \sum_i p_i$, we suppose that no jump occurs and that the wave-function evolves with the Hamiltonian H_{eff}. We then obtain $|\phi(t+dt)\rangle = (1 - iH_{\text{eff}}dt/\hbar)|\phi(t)\rangle / \|(1 - iH_{\text{eff}}dt/\hbar)|\phi(t)\rangle\|$.

The density matrix at $t+dt$ is obtained by weighting the different outcomes with their corresponding probabilities

$$\begin{aligned}
&(1 - \sum_i p_i) \frac{(1 - iH_{\text{eff}}dt/\hbar)|\phi(t)\rangle\langle\phi(t)|(1 + iH_{\text{eff}}^\dagger dt/\hbar)}{\langle\phi(t)|(1 + iH_{\text{eff}}^\dagger dt/\hbar)(1 - iH_{\text{eff}}dt/\hbar)|\phi(t)\rangle} + \sum_i p_i \frac{A_i|\phi(t)\rangle\langle\phi(t)|A_i^\dagger}{\langle\phi(t)|A_i^\dagger A_i|\phi(t)\rangle} \\
&= (1 - \sum_i p_i) \frac{(1 - iH_{\text{eff}}dt/\hbar)|\phi(t)\rangle\langle\phi(t)|(1 + iH_{\text{eff}}^\dagger dt/\hbar)}{\langle\phi(t)|(1 + i(H_{\text{eff}}^\dagger - H_{\text{eff}})dt/\hbar)|\phi(t)\rangle} + \sum_i \gamma_i A_i |\phi(t)\rangle\langle\phi(t)| A_i^\dagger \, dt \\
&= (1 - iH_{\text{eff}}dt/\hbar)|\phi(t)\rangle\langle\phi(t)|(1 + iH_{\text{eff}}^\dagger dt/\hbar) + \sum_i \gamma_i A_i |\phi(t)\rangle\langle\phi(t)| A_i^\dagger \, dt \\
&= |\phi(t)\rangle\langle\phi(t)| - \frac{i}{\hbar}(H_{\text{eff}}|\phi(t)\rangle\langle\phi(t)| - |\phi(t)\rangle\langle\phi(t)| H_{\text{eff}}^\dagger)dt + \sum_i \gamma_i A_i |\phi(t)\rangle\langle\phi(t)| A_i^\dagger \, dt
\end{aligned}$$

By averaging the last line over the different states $|\phi(t)\rangle\langle\phi(t)|$ with probabilities p_ϕ, we recover (1.16). The density matrix at times t can therefore be obtained by considering different trajectories obtained through the stochastic evolution of pure states as described above. A single quantum trajectory consists of periods of Hamiltonian evolution of the wave-function separated by quantum jumps, which happen randomly. This approach requires much less memory than when directly solving the master equation. But, many trajectories have to be computed and then averaged in order to obtain a good estimate of $\rho(t)$.

1.4 Generalized measurement (POVM)

We finish this chapter by generalizing the von Neumann projective measurement to a broader class of measurements. We follow the same approach that we used to build a quantum map and consider that the system and the environment evolves unitarily. We do not trace over the environment and obtain the following density matrix

$$\rho \otimes |0\rangle \langle 0|_E \xrightarrow{U} \sum_{\mu\nu} U_{\mu 0}\, \rho\, U_{\nu 0}^\dagger \otimes |\mu\rangle \langle \nu|_E$$

We can now proceed to a projective measurement over the environment. This can be formalized by considering the set of orthogonal projectors $\Pi_\mu = \mathbb{1} \otimes |\mu\rangle \langle\mu|_E$, which sum up to identity $\sum_\mu \Pi_\mu = \mathbb{1} \otimes \mathbb{1}_E$. The probability to observe the outcome μ is

$$p_\mu = \mathrm{Tr}\Big[U_{\mu 0}\rho U_{\mu 0}^\dagger\Big] = \mathrm{Tr}[\rho E_\mu] \text{ with } E_\mu = U_{\mu 0}^\dagger U_{\mu 0}$$

The operators E_μ are positive and verify $\sum_\mu E_\mu = \mathbb{1}$, they form what is called a Positive Operator Valued Measure (POVM). Given a set of E_μ, we can find Kraus operators such that $E_\mu = U_{\mu 0}^\dagger U_{\mu 0}$ (considering for example $U_{\mu 0} = \sqrt{E_\mu}$) and thus any POVM can be dilated to a unitary evolution in a larger Hilbert space followed by a projective measurement. This result is known in the literature as the Naimark's dilation theorem (or Neumark's dilation theorem). It can also be seen as a consequence of the Stinespring's theorem. After measurement, the density matrix of the system is $\rho_\mu = U_{\mu 0}\rho U_{\mu 0}^\dagger / p_\mu$. The POVM generalize projective measurements but have different properties:

- The E_μ are, in general, not orthogonal, therefore the number of operators (or measurement outcomes) is not limited to the dimension of the Hilbert space. For the same reason, repeated measurements can lead to different outcomes as opposed to projective measurements.
- The knowledge of the E_μ is not sufficient to define the state of the system after the measurement. This is because the same POVM can be realized by different sets of Kraus operators. If one is interested in both the measurement result and the final state of the system, then the measurement must be specified trough Kraus operators rather than POVM.

1.4.1 Example of POVMs for a qubit

A standard way to construct a POVM for a qubit is to consider the set of N operators

$$E_i = \frac{1}{N}(\mathbb{1} + \vec{p}_i.\vec{\sigma}) \text{ with } \sum_{i=1}^{N} \vec{p}_i = \vec{0} \tag{1.17}$$

where the $\vec{p}_i$ are 3D vectors with $||\vec{p}_i|| \leq 1$. We will see in the next chapter that such a POVM is useful to distinguish non-orthogonal qubit states. A well chosen POVM can perform better than a projective measurement, meaning that the correct result can be obtained with with higher probability.

1.5 Summary

- The state of an open quantum system is described by a density matrix ρ which is a positive semi-definite operator with unit trace. The eigenvalues of ρ correspond to the probabilities to observe the system in the corresponding eigenstate. A state is pure if ρ has a unique non zero eigenvalue, which is equal to one. Otherwise, ρ describes a mixed state. A density matrix can always be written (in a non unique way) as $\rho = \mathrm{Tr}_E[|\psi\rangle\langle\psi|]$, one says that $|\psi\rangle$ is a purification of ρ.

- The most general evolution of a density matrix corresponds to a completely positive trace preserving (CPTP) linear application, called a quantum map or channel. The action of a quantum map on ρ can be described in various ways, including the Kraus form or the use of the Choi matrix. It can be dilated to a unitary evolution in a larger Hilbert space. Such dilation is not unique. A quantum map is not reversible unless it corresponds to a unitary evolution.

- A markovian master equation for the density matrix results in an evolution, which is a quantum map, if and only if it can be written in the Lindblad form. The non Hamiltonian terms in the master equation can be interpreted as quantum jumps that happen randomly during the evolution of the quantum state.

- At the end of the evolution, a generalized measurement (POVM) can be performed to obtain the state of the system. The probability of the outcome μ is given by the mean value of a positive operator E_μ. In general, the different outcomes are not mutually exclusive. A generalized measurement can be dilated in a non unique way to a projective measurement in a larger Hilbert space.

Exercises

- Consider the map acting on a qubit generated by the two Kraus operators $\mathbb{1}/\sqrt{2}$ and $\sigma_3/\sqrt{2}$. Find a dilation of this map by considering an environment consisting of a second qubit. Explicitly write the 4×4 unitary matrix that is involved in the dilation.

- Consider the same map as above. The state ρ of a qubit can be written as a column vector with four terms and the map as a matrix acting on this vector. Choose a basis and find the matrix representation of the map. Compare it to the Choi matrix of the map. What are the eigenvalues of these matrices?

- Consider a purifications $|\psi\rangle$ in $\mathscr{H}_S \otimes \mathscr{H}_E$ of a given density matrix ρ in $\mathscr{H}_S$. Consider a unitary transform U_E acting in $\mathscr{H}_E$. Show that $(\mathbb{1} \otimes U_E)|\psi\rangle$ also purifies ρ. Show that the converse is true, meaning that two purifications are always related by a unitary transform in $\mathscr{H}_E$.

- Consider a POVM as defined in (1.17) with $N = 4$ and the $\vec{P}_i$ being unit vectors pointing along x,$-x$,y and $-y$. Compute the probability of the measurement outcomes for a qubit prepared in $|\uparrow_x\rangle$, $|\downarrow_x\rangle$, $|\uparrow_y\rangle$ or $|\downarrow_y\rangle$. Compare with a projective measurement in the x or y basis or any other basis of the qubit. What is the advantage of the POVM?

References

[1]P. Pechukas, "Reduced Dynamics Need Not Be Completely Positive", Physical Review Letters **73**, 1060–1062 (1994).

[2]C. A. Rodríguez-Rosario, K. Modi, A. M. Kuah, A. Shaji, and E. C. Sudarshan, "Completely positive maps and classical correlations", Journal of Physics A: Mathematical and Theoretical **41** (2008).

[3]M. D. Choi, "Completely positive linear maps on complex matrices", Linear Algebra and Its Applications **10**, 285–290 (1975).

[4]S. Milz, F. A. Pollock, and K. Modi, "An Introduction to Operational Quantum Dynamics", Open Systems and Information Dynamics **24** (2017).

[5]E. Mozgunov and D. Lidar, "Completely positive master equation for arbitrary driving and small level spacing", Quantum **4**, Publisher: Verein zur Förderung des Open Access Publizierens in den Quantenwissenschaften, 227 (2020).

[6]P. D. Drummond and D. F. Walls, "Quantum theory of optical bistability. I. Nonlinear polarisability model", Journal of Physics A: Mathematical and General **13**, Publisher: IOP Publishing, 725–741 (1980).

[7]J. Dalibard, Y. Castin, and K. Mølmer, "Wave-function approach to dissipative processes in quantum optics", Physical Review Letters **68**, Publisher: American Physical Society, 580–583 (1992).

Chapter 2

Entanglement

Entanglement is maybe the most important concept introduced by quantum mechanics. Because of entanglement, physicists must abandon the idea that the physical world is described by a local causal theory. Bell was the first to realize that the non-locality of quantum mechanics could actually be tested experimentally using his famous inequality. We will start this chapter by deriving a special form of Bell's inequalities, which are called the Clauser, Horne, Shimony and Holt (CHSH) inequalities. These are the entanglement inequalities that have been most often tested experimentally.

Today, the concept of entanglement is well accepted and entanglement is seen as a resource to perform tasks that would be impossible to a classical machine. Entanglement is the fuel of the (yet to come) quantum technologies. It is required for quantum algorithms to be efficient and quantum metrology devices to be more precise. A natural goal for the quantum engineer is therefore to quantify the amount of entanglement stored in a given state. However, this question turns out to be very difficult and, up to now, there is no simple way to measure how much entangled a state is, or even if it is entangled or not.

In quantum information, the natural unit of entanglement is the «ebit» that corresponds to two qubits in a maximally entangled state. There are four canonical states with this property, which are called Bell states

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

Alice and Bob have n ebits if they share $2n$ qubits holding n copies of a Bell state. They can use this resource for quantum communication: for example, if Alice wishes to teleport one qubit to Bob, the process consumes one ebit (see next chapter). Given a state $|\psi\rangle$, a practical information for Alice and Bob would be to know how many ebits they can obtain from the n copies of $|\psi\rangle$. If the state is pure, we will see that there is a simple answer but that the situation for mixed state is more complicated.

2.1 CHSH inequalities

The CHSH inequalities apply to an experiment where Alice and Bob share some boxes with the following properties. Each box has two buttons, and when one button is pressed, the box outputs a binary result. The results obtained by Alice and Bob are described by

the joint probability distribution $P(m_A\, m_B|s_A\, s_B)$ where $m_i \in \{0,1\}$ is the measurement outcome and $s_i \in \{0,1\}$ is the button that was pressed (s_i stands for setting on side i). To test whether the boxes are ruled by quantum mechanics or by classical probability laws, we look at the correlations between the measurement results obtained by Alice and Bob. We introduce the probability $q(s_A, s_B)$, which is the probability that Alice and Bob measure the same result $q(s_A, s_B) = P(00|s_A\, s_B) + P(11|s_A\, s_B)$. Taking all possible settings, we obtain a four dimension vector $\vec{q}$ defined as

$$\vec{q} = \begin{bmatrix} q(0,0) \\ q(0,1) \\ q(1,0) \\ q(1,1) \end{bmatrix}$$

We now want to find the volume spanned by $\vec{q}$ when the probability distribution P is «classical», or more precisely when it follows from a local hidden variable (LHV) theory. The LHV assumption means that the outcome m_i of a box follows a probability distribution that depends only on the local setting s_i. In addition, this probability distribution may depend on the value of some hidden variables. Under this LHV assumption, the joint probability $P(m_A\, m_B|s_A\, s_B)$ becomes

$$P(m_A\, m_B|s_A\, s_B) = \sum_i \lambda_i P_A^{(i)}(m_A|s_A) P_B^{(i)}(m_B|s_B), \quad \lambda_i \in [0,1], \quad \sum_i \lambda_i = 1 \tag{2.1}$$

where λ_i is the probability that the hidden variables lead to the distribution with index i. Each probability distribution $P_A^{(i)}$ and $P_B^{(i)}$ can be further decomposed over probability distributions with certain outcomes. There are four of them

$$\pi_1 : \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 1 & 1 \end{array} \quad \pi_2 : \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1 & 1 \\ 1 & 0 & 0 \end{array} \quad \pi_3 : \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array} \quad \pi_4 : \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1 & 0 \\ 1 & 0 & 1 \end{array}$$

Each table shows the probability of the $m = 0$ and $m = 1$ outcomes as two column vectors, one for each setting s. We obtain

$$P(m_A\, m_B|s_A\, s_B) = \sum_{ijk} \mu_{ijk} \pi_j(m_A|s_A) \pi_k(m_B|s_B), \quad \mu_{ijk} \in [0,1], \quad \sum_{ijk} \mu_{ijk} = 1$$

The sum over i is over the different values of the hidden variables and the two sums over j and k are over the different distributions with certain outcomes. The vector $\vec{q}$ is then given by

$$\begin{aligned} \vec{q} &= \sum_{ijk} \mu_{ijk} \begin{bmatrix} \pi_j(0|0)\pi_k(0|0) + \pi_j(1|0)\pi_k(1|0) \\ \pi_j(0|1)\pi_k(0|0) + \pi_j(1|1)\pi_k(1|0) \\ \pi_j(0|0)\pi_k(0|1) + \pi_j(1|0)\pi_k(1|1) \\ \pi_j(0|1)\pi_k(0|1) + \pi_j(1|1)\pi_k(1|1) \end{bmatrix} \\ &= \nu_0 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \nu_1 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} + \nu_2 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} + \nu_3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} + \nu_4 \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + \nu_5 \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \nu_6 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \nu_7 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \end{aligned}$$

where $\nu_i \in [0,1]$ and $\sum_i \nu_i = 1$. The vector $\vec{q}$ is thus inside a convex polytope, whose vertices are given by the eight points obtained from all possible combination of the probability

distributions with certain outcomes[1]. A coordinate is one if $\pi_j(0|s_A) = \pi_k(0|s_B)$ and zero otherwise. It is easy to see that the number of null coordinates per vertex must be 0,2 or 4. We then notice that the coordinates of the first four points satisfy $x_1 + x_2 + x_3 - x_4 = 0$ while the ones of the last four points verify $x_1 + x_2 + x_3 - x_4 = 2$. These two equations define two parallel hyperplanes. From this observation, we conclude that every point inside the polytope is also inside the slab delimited by these two hyperplanes. We therefore obtain the following inequalities for the coordinates of $\vec{q}$

$$0 \leq q(0,0) + q(0,1) + q(1,0) - q(1,1) \leq 2 \tag{2.2}$$

We can repeat the same reasoning by noticing that the two parallel hyperplanes defined by $x_i + x_j + x_k - x_l = 0$ and $x_i + x_j + x_k - x_l = 2$ always contain each four points for any combination of the coordinates. We then obtain three other sets of two inequalities

$$\begin{aligned} 0 \leq q(0,1) + q(1,0) + q(1,1) - q(0,0) \leq 2 \\ 0 \leq q(1,0) + q(1,1) + q(0,0) - q(0,1) \leq 2 \\ 0 \leq q(1,1) + q(0,0) + q(0,1) - q(1,0) \leq 2 \end{aligned}$$

Geometrically, the polytope allowed by LHV theories is an octahedron in dimension four[2]. The eight CHSH inequalities together with the eight trivial inequalities $0 \leq q(i,j) \leq 1$ define this hyperoctahedron, as first shown by Froissart in 1981 [1].

The original CHSH inequalities are written in terms of slightly different quantities $E_{s_A s_B} = P(01|s_A\, s_B) + P(10|s_A\, s_B) - P(00|s_A\, s_B) - P(11|s_A\, s_B)$ which can be rewritten $E_{s_A s_B} = 1 - 2q(s_A, s_B)$. The inequalities (2.2) then become

$$C = E_{00} + E_{01} + E_{10} - E_{11} \qquad |C| \leq 2 \tag{2.3}$$

as originally obtained by Clauser, Horne, Shimony and Holt [2]. Three other inequalities can be obtained by permuting the expectation values. Let us summarize this important result: classical theories, in the sense of local hidden variables theories, predict measurement correlations that must necessarily satisfy the CHSH inequalities. If the outcome of an experiment performed by Alice and Bob violates this inequality, then the LHV assumption must be rejected.

2.1.1 Maximal violation of the CHSH inequalities

In order to violate the CHSH inequality (2.3), Alice and Bob can share a pair of qubits in a Bell state and carefully choose their measurement settings. If Alice and Bob measure σ_A and σ_B, the two possible outcomes on each side are -1 and 1. The probability that enters in the CHSH inequality is $P(\sigma_A = 1, \sigma_B = -1) + P(\sigma_A = -1, \sigma_B = 1) - P(\sigma_A = 1, \sigma_B = 1) - P(\sigma_A = -1, \sigma_B = -1)$, which is equal to $-\langle \sigma_A \sigma_B \rangle$. We therefore obtain

$$\langle \sigma_A^{(0)} \sigma_B^{(0)} \rangle + \langle \sigma_A^{(0)} \sigma_B^{(1)} \rangle + \langle \sigma_A^{(1)} \sigma_B^{(0)} \rangle - \langle \sigma_A^{(1)} \sigma_B^{(1)} \rangle = -C$$

where the (0) and (1) indices identify the two possible measurement settings. We will consider that Alice and Bob share the $|\Phi_+\rangle$ state. This state has the property that

$$\langle \Phi_+ | \, \sigma_X \otimes \sigma_X \, | \Phi_+ \rangle = \langle \Phi_+ | \, \sigma_Z \otimes \sigma_Z \, | \Phi_+ \rangle = 1 \quad \langle \Phi_+ | \, \sigma_X \otimes \sigma_Z \, | \Phi_+ \rangle = 0$$

[1] From the 16 distributions, we obtain only 8 points because each point appears twice.

[2] There is an isometry that transforms the eight vertices to [±1 0 0 0], [0 ± 1 0 0], [0 0 ± 1 0] and [0 0 0 ± 1], which correspond to the eight canonical vertices of the hyperoctahedron. This regular convex polytope is also called the hexadecachoron or 16-cell, it is dual to the hypercube.

If Alice and Bob both measure along σ_X and σ_Z, they cannot violate the CHSH inequality. But, if Alice chooses σ_X and σ_Z as her two measurement settings, while Bob chooses $(\sigma_X + \sigma_Z)/\sqrt{2}$ and $(\sigma_X - \sigma_Z)/\sqrt{2}$, they obtain

$$\langle \sigma_A^{(0)} \sigma_B^{(0)} \rangle = \langle \sigma_A^{(0)} \sigma_B^{(1)} \rangle = \langle \sigma_A^{(1)} \sigma_B^{(0)} \rangle = 1/\sqrt{2} \quad \langle \sigma_A^{(1)} \sigma_B^{(1)} \rangle = -1/\sqrt{2} \quad \Rightarrow |C| = 2\sqrt{2}.$$

In [3], Tsirelson showed that this is the maximal violation allowed by quantum mechanics, this bound is known as the Tsirelson (or Cirel'son) limit.

One can wonder if this value can be derived in the same manner as the CHSH inequalities by supposing some fundamental properties of the physical theory underlying the behavior of the boxes. One natural assumption is the «non-signaling» assumption, which says that the setting of one box cannot influence instantaneously the outcome of the other box, because information cannot travel faster than light. Therefore, if Alice and Bob are sufficiently far apart and carefully time their experiment so that their measurements correspond to events in disjoint cones of space-time, the boxes outcome on one side cannot influence the outcome on the other side. This translates into the following non-signaling conditions for the probability distribution

$$\sum_{m_B} P(m_A\ m_B | s_A\ s_B = 0) = \sum_{m_B} P(m_A\ m_B | s_A\ s_B = 1)\ \forall s_A$$
$$\sum_{m_A} P(m_A\ m_B | s_A = 0\ s_B) = \sum_{m_A} P(m_A\ m_B | s_A = 1\ s_B)\ \forall s_B$$

In [4], Popescu and Rohrlich introduced the probability distribution

$$P(m_A\ m_B | s_A\ s_B) = \begin{cases} 1/2 \text{ if } m_A \oplus m_B = s_A s_B \\ 0 \text{ otherwise} \end{cases}$$

where $\oplus$ is the exclusive or operator, which gives 1 if m_A and m_B are different and 0 otherwise. This probability distribution satisfies the non-signaling condition and leads to stronger correlation than quantum mechanics, reaching $|C| = 4$. The non-signaling constraint is therefore not sufficient to recover the results of quantum mechanics. Recently, other constraints have been proposed to derive the Tsirelson bound, but the question is still debated [5].

2.2 Entanglement of bipartite pure states

For bipartite pure states, entanglement can be quantified straightforwardly. We will first introduce the Schmidt decomposition of a pure state $|\psi\rangle$, from which the entropy of entanglement S_E is defined. We will then show that if Alice and Bob share n copies of $|\psi\rangle$, they can distillate $n \log(S_E)$ ebits.

2.2.1 Schmidt decomposition

A generic bipartite pure state can be expanded in a basis as

$$|\psi\rangle = \sum_{ij} a_{ij} |i\rangle_A |j\rangle_B$$

The Schmidt decomposition is obtained when the expansion basis is chosen to be an eigenbasis of the reduced density matrices. This eigenbasis is directly obtained from the SVD decomposition of the a_{ij} matrix

$$a_{ij} = \sum_k u_{ik}\sigma_k v^*_{kj} \;\Rightarrow\; |k\rangle_A = \sum_i u_{ik}\,|i\rangle_A\,,\; |k\rangle_B = \sum_j v^*_{kj}\,|j\rangle_B$$

where σ_k are the real positive singular values of A, and U and V are unitary matrices defining a basis change in $\mathscr{H}_A$ and $\mathscr{H}_B$. In the new basis, we obtain the following Schmidt decomposition

$$|\psi\rangle = \sum_k \sqrt{p_k}\,|k\rangle_A\,|k\rangle_B$$

where $\sqrt{p_k} = \sigma_k$. The p_k are eigenvalues of the reduced density matrices ρ_A and ρ_B with associated eigenvectors $|k\rangle_A$ and $|k\rangle_B$

$$\rho_A = \mathrm{Tr}_B[|\psi\rangle\langle\psi|] = \sum_k p_k\,|k\rangle\langle k|_A \quad \rho_B = \mathrm{Tr}_A[|\psi\rangle\langle\psi|] = \sum_k p_k\,|k\rangle\langle k|_B\,. \tag{2.4}$$

A pure state is separable if and only if (iff) its Schmidt decomposition contains only one term, in which case ρ_A and ρ_B are pure states. The number of non-zero eigenvalues p_k is called the Schmidt rank and is strictly larger than one iff the state is entangled. On the contrary, we say that a pure state is maximally entangled if $p_k = 1/d$ where d is the smaller dimension of $\mathscr{H}_A$ and $\mathscr{H}_B$. If Alice and Bob apply local unitary operation in the form of $U_A \otimes U_B$, they obtain states with the same Schmidt decomposition. Alice and Bob cannot increase the Schmidt rank with local operations.

2.2.2 Entropy of entanglement and entanglement distillation

The von Neumann entropy of a density matrix ρ is defined as $S = -\,\mathrm{Tr}[\rho\log\rho]$. From the Schmidt decomposition of $|\psi\rangle$, we see that

$$S(\rho_A) = S(\rho_B) = -\sum_k p_k \log p_k \equiv S_E,$$

where the log symbol refers to the base two logarithm function. In information theory, the expression $-\sum_k p_k \log p_k$ defines the Shannon entropy, we will come back to it in the next chapter. The quantity S_E is called entropy of entanglement. Intuitively, we expect that it quantifies the amount of entanglement contained in the state. One defines an entanglement measure as a positive quantity that fulfills (at least) the following two criteria:

- An entanglement measure is a sufficient criterion for entanglement: it is zero for a separable state.
- The measured entanglement cannot be increased by local operations and classical communication. This property is called monotonicity under LOCC.

We have seen that the entropy of entanglement satisfies these two criteria for local operations that are unitary evolutions. It is not too difficult to show that other local operations, involving projective measurements and evolution conditioned on the measurement results cannot increase S_E, even if Alice and Bob discuss their measurements results on the phone. Entropy of entanglement is therefore an entanglement measure.

We will now show that S_E is also an operational measure of how many ebits can be distilled from $|\psi\rangle$. In order to simplify the calculation, we suppose that $|\psi\rangle$ is a two qubit state. Its Schmidt decomposition then reduces to two terms

$$|\psi\rangle = \sqrt{1-p}\,|0\rangle_A\,|0\rangle_B + \sqrt{p}\,|1\rangle_A\,|1\rangle_B$$

and the entanglement entropy is $S_E = -p\log p - (1-p)\log(1-p)$. We now suppose that Alice and Bob share n copies of this state, the total state is

$$|\psi\rangle^{\otimes n} = \sum_s \sqrt{p(s)}\,|s\rangle_A\,|s\rangle_B\,,$$

where the sum goes over the 2^n possible binary strings of n bits. If Alice and Bob would project their share of the state, they would obtain the same binary string s (perfect correlation) with a probability $p(s)$. We rather suppose that Alice performs a projective measurement to obtain the number of one bits in her state. If Alice obtains the outcome m, the global state is projected to

$$|\psi_m\rangle = \sum_{s_m} |s_m\rangle_A\,|s_m\rangle_B$$

where the sum goes over the $\binom{n}{m}$ binary strings of length n with m bits set to one. When n is large, the probability distribution of the measurement outcome is sharply peaked around $m = np$ and the number of strings s_m such that $m = np$ can be approximated by

$$\log\binom{n}{np} \sim -n\left[p\log p + (1-p)\log(1-p)\right] = nS_E$$

This expression can be obtained from the Stirling formula, but we will see in the next chapter that the Shannon entropy $-\sum_k p_k \log p_k$ is actually defined such that it counts the number of strings that are the most probable.

The $|s_m\rangle$ form an orthogonal set spanning a Hilbert space of dimension $d = \binom{n}{m} \approx 2^q$ with $q = nS_E$. To make the statement more precise, we choose q such that $2^q < d < 2^{q+1}$, and we know that $q \approx nS_E$. Alice and Bob now apply a unitary transformation U that sends the $|s_m\rangle$ states to the direct sum of two subspaces, one of dimension 2^q, spanned by all the binary strings of q bits, and the second of dimension $d - 2^q$. By doing so, they prepare

$$(U\otimes U)\,|\psi_m\rangle = \left(\frac{|0\rangle_A\,|0\rangle_B + |1\rangle_A\,|1\rangle_B}{\sqrt{2}}\right)^{\otimes q} \otimes |\psi_{\text{garbage}}\rangle_{AB}$$

where the garbage state $|\psi_{\text{garbage}}\rangle_{AB}$ still contains some entanglement and should be properly recycled. At the end of the operation, Alice and Bob have distilled their initial state to a register of $q \approx nS_E$ ebits. As expected, q is maximal for $p = 1/2$, in which case they obviously obtain n ebits since the initial state $|\psi\rangle$ is already a maximally entangled Bell state. Because Alice and Bob only used local operations, we expect that no entanglement was created during the distillation process, thus S_E is an operational measure of the amount of entanglement contained in $|\psi\rangle$. The proof can be extended to the case where $|\psi\rangle$ is more than a two qubit state.

Let us summarize the results that we have obtained for the entanglement of a bipartite pure state. We have first shown that an entangled state can be used to violate the CHSH inequalities. We have then shown that the Schmidt decomposition provides a natural entanglement criterion. It also allows us to define an entropy of entanglement, which is equivalent to another entanglement measure, the distillable entanglement. Things are even better, because one can show that

- The violation of a CHSH inequality is a sufficient and necessary criterion for the entanglement of bipartite pure states, *i.e.* states having a Schmidt rank strictly larger than one.

- All entanglement measures are equivalent for bipartite pure states, therefore the amount of entanglement is fully characterized by the entropy of entanglement.

2.3 Entanglement of bipartite mixed states

Because there is no equivalent of the Schmidt decomposition for mixed states, we say that a bipartite mixed state is entangled iff it is not a separable state. And we say that a state ρ is separable, iff it can be written as a mixture of tensor product states

$$\rho = \sum_i p_i \, \rho_A^{(i)} \otimes \rho_B^{(i)}, \;\; p_i \in [0,1], \;\; \sum_i p_i = 1 \tag{2.5}$$

Unfortunately, the situation is much more complicated than for pure states, because one can show that:

- Some mixed entangled bipartite states do not violate the CHSH inequalities (e.g. Werner states). The violation of a CHSH inequality is only a sufficient entanglement criterion.

- Some mixed bipartite entangled states cannot be distilled, they are said to be bound entangled. Different entanglement measures may rank entangled states differently.

Therefore, one must abandon the idea of an universal entanglement measure. Furthermore, even knowing whether a state is entangled or not remains a difficult problem, because there is no simple operative criterion to answer this question. Still, an important result was obtained by the Horodeckis that connects entanglement and non positivity of quantum maps.

2.3.1 Non completely positive maps and PPT criteria

In 1996, the Horodeckis showed that a bipartite state ρ is separable iff $(\mathbb{1} \otimes \Lambda)[\rho]$ is positive for any positive map Λ [6]. One direction is obvious, if we consider ρ to be a product state, we obtain

$$\rho = \rho_A \otimes \rho_B \Rightarrow (\mathbb{1} \otimes \Lambda)[\rho] = \rho_A \otimes \Lambda[\rho_B] \geq 0 \, . \tag{2.6}$$

The same is true if ρ is a mixture of product states as defined in 2.5. We will prove the other direction in the next paragraph after introducing the notion of entanglement witness.

With this result, the characterization of entanglement reduces to the characterization of positive but not completely positive (nCP) maps (see chapter 1). The most natural map Λ with this property is the transpose operation. The action of $(\mathbb{1} \otimes \Lambda)$ is called the partial transpose operation and is often denoted ρ^Γ. If we write ρ as a block matrix, we obtain

$$\rho = \sum_{i,j} |i\rangle\langle j| \otimes \rho_{ij} \;\; \Rightarrow \;\; \rho^\Gamma = \sum_{i,j} |i\rangle\langle j| \otimes \rho_{ij}^T \tag{2.7}$$

If the partial transpose of a state is not positive, then the state is entangled. This is called the positive partial transpose (PPT) criterion, which was discovered by Peres in 1996. Let

us apply it to the $|\Phi_+\rangle$ Bell state

$$\rho = |\Phi_+\rangle\langle\Phi_+| = \frac{1}{2}\begin{pmatrix}1&0&0&1\\0&0&0&0\\0&0&0&0\\1&0&0&1\end{pmatrix} \Rightarrow \rho^\Gamma = \frac{1}{2}\begin{pmatrix}1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{pmatrix}$$

By permuting the two central columns of ρ^Γ, we see that $\det 2\rho^\Gamma = -1$, and ρ is not PPT and therefore entangled.

The magic is that this criterion is also a necessary criterion if the dimension $d_A \times d_B$ of the space is four (two qubits) or six (one qubit, one qutrit). This is because, in small dimensions, a positive map can be decomposed as a sum of one CP map and a second CP map composed with the transpose map [6]. For a two qubit state, the criterion even further simplifies to the following statement: a two qubit state ρ is entangled iff $\det \rho^\Gamma < 0$. In higher dimensions, there are entangled states that are not detected by the PPT criterion, it has been shown that such states cannot be distilled.

2.3.2 Entanglement witness

In the previous chapter, we have seen that a map Λ is a linear application that can be represented by a matrix. A possible choice for such matrix is the Choi matrix C_Λ. Here, we suppose that the map acts on Bob's space and the Choi matrix is a $d_B^2 \times d_B^2$ matrix, where d_B is the dimension of $\mathscr{H}_B$. In order to simplify the discussion, we will suppose that $d_A = d_B$ so that we can identify $\mathscr{H}_A \otimes \mathscr{H}_B$ with $\mathscr{H}_B \otimes \mathscr{H}_B$, but the results are still valid if this is not the case.

In chapter 1, we have shown that a CP map corresponds to a positive C_Λ. We can now wonder what happens if the map is positive but not CP. In this case, the Choi matrix is not positive, but just hermitian. The resulting map is called an entanglement witness, which has positive expectation value over all separable states

$$\Lambda \geq 0 \;\Leftrightarrow\; \mathrm{Tr}[C_\Lambda \rho_{\mathrm{sep}}] \geq 0 \tag{2.8}$$

where ρ_{sep} is any separable state. In order to prove this equivalence, we first consider that ρ_{sep} is a product state $|\mu\rangle\langle\mu| \otimes |\nu\rangle\langle\nu|$ of two pure states. Multiplying by C_Λ and taking the trace, we obtain

$$\begin{aligned}\mathrm{Tr}[C_\Lambda\,(|\mu\rangle\langle\mu| \otimes |\nu\rangle\langle\nu|)] &= \sum_{i,j} \mathrm{Tr}[(|i\rangle\langle j| \otimes \Lambda[|i\rangle\langle j|])\,(|\mu\rangle\langle\mu| \otimes |\nu\rangle\langle\nu|)] \\ &= \sum_{i,j} \langle j|\mu\rangle\, \mathrm{Tr}[|i\rangle\langle\mu| \otimes (\Lambda[|i\rangle\langle j|]\,|\nu\rangle\langle\nu|)] \\ &= \sum_{i,j} \langle j|\mu\rangle\,\langle\mu|i\rangle\, \mathrm{Tr}[\Lambda[|i\rangle\langle j|]\,|\nu\rangle\langle\nu|]\,,\end{aligned}$$

where, on the first line, we use the definition of C_Λ given in chapter 1. We define the complex conjugate $|\mu^*\rangle$ of $|\mu\rangle$ as $|\mu^*\rangle = \sum_i \langle\mu|i\rangle\,|i\rangle$ and obtain

$$\begin{aligned}\mathrm{Tr}[C_\Lambda\,(|\mu\rangle\langle\mu| \otimes |\nu\rangle\langle\nu|)] &= \mathrm{Tr}[\Lambda[|\mu^*\rangle\langle\mu^*|]\,|\nu\rangle\langle\nu|] \\ &= \langle\nu|\,\Lambda[|\mu^*\rangle\langle\mu^*|]\,|\nu\rangle\;.\end{aligned}$$

From the last equality, we obtain that $\mathrm{Tr}[C_\Lambda\rho_{\mathrm{sep}}] \geq 0$ is equivalent to $\Lambda \geq 0$. We first suppose that $\mathrm{Tr}[C_\Lambda\rho_{\mathrm{sep}}] \geq 0$ and consider a density matrix ρ that we diagonalize as $\rho = \sum p_\mu\,|\mu\rangle\langle\mu|$.

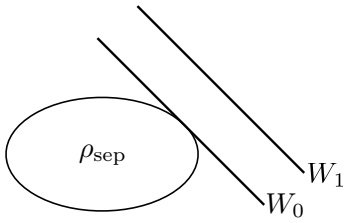


Figure 2.1: An entanglement witness separates the space of density matrices in two halves, where one half contains the convex set of separable states. The witness W_0 is optimal because it is tangent to the set of separable states.

For any $|\nu\rangle$, we obtain that $\langle\nu|\,\Lambda[\rho]\,|\nu\rangle = \sum p_\mu \,\mathrm{Tr}[C_\Lambda\,(|\mu^*\rangle\langle\mu^*|\otimes|\nu\rangle\langle\nu|)] \geq 0$, meaning that $\Lambda[\rho]\geq 0$ and therefore $\Lambda\geq 0$. For the other direction, we consider a separable state ρ_{sep} and decompose it as a convex sum of density matrices $|\mu\rangle\langle\mu|\otimes|\nu\rangle\langle\nu|$, which is always possible by diagonalizing the density matrices $\rho_A^{(i)}$ and $\rho_B^{(i)}$ appearing in (2.5). We then use that $\Lambda\geq 0$ to obtain that each term $\mathrm{Tr}[C_\Lambda\,(|\mu\rangle\langle\mu|\otimes|\nu\rangle\langle\nu|)]$ is positive and finally that $\mathrm{Tr}[C_\Lambda\rho_{\text{sep}}]\geq 0$.

If we forget for a moment about positive maps, we can define an entanglement witness as an observable W such that $\mathrm{Tr}[W\rho_{\text{sep}}]\geq 0$ for all separable states ρ_{sep}. A witness defines an hyperplane in the space of density matrices such that the set of separable states is entirely on one side of the plane as shown in the figure 2.1. For any entangled state ρ, there exists a witness that detects ρ, meaning $\mathrm{Tr}[W\rho]<0$. This comes from the fact that the set of separable states is convex as seen from the definition 2.5. We have actually followed this path when we derived the CHSH inequalities. The observable used by Alice and Bob to violate one of the CHSH inequality is an entanglement witness.

We are now ready to prove that for any entangled state ρ, there exists a positive map Λ such that $(\mathbb{1}\otimes\Lambda)[\rho]$ is not positive. We start from the witness W that detects ρ and consider the map Λ such that $C_\Lambda = W$. Because W is a witness, $\mathrm{Tr}[W\rho_{\text{sep}}]\geq 0$ and we have shown that Λ is therefore positive. We now have to show that $(\mathbb{1}\otimes\Lambda)[\rho]$ is not positive. As already introduced in chapter 1, we consider the state $|\psi_+\rangle$, which is the maximally entangled state $\sum_i |i\rangle\otimes|i\rangle$ without normalization. We have

$$C_\Lambda = \sum_{i,j} |i\rangle\langle j|\otimes\Lambda[|i\rangle\langle j|] = (\mathbb{1}\otimes\Lambda)[|\psi_+\rangle\,\langle\psi_+|]\,, \tag{2.9}$$

from which we obtain

$$\mathrm{Tr}[\rho W] = \mathrm{Tr}[\rho(\mathbb{1}\otimes\Lambda)[|\psi_+\rangle\,\langle\psi_+|]] \;\; = \mathrm{Tr}\Big[|\psi_+\rangle\,\langle\psi_+|\,(\mathbb{1}\otimes\Lambda^\dagger)[\rho]\Big] = \langle\psi_+|\,(\mathbb{1}\otimes\Lambda^\dagger)[\rho]\,|\psi_+\rangle\;,$$

where we have introduced the conjugate map $\Lambda^\dagger$ such that $\mathrm{Tr}\Big[\rho_1\Lambda^\dagger[\rho_2]\Big] = \mathrm{Tr}[\rho_2\Lambda[\rho_1]]$ for any ρ_1 and ρ_2. If W detects ρ, then $\mathrm{Tr}[\rho W]<0$ and $(\mathbb{1}\otimes\Lambda^\dagger)[\rho]$ is not positive. It is easy to show that $\Lambda^\dagger$ is positive iff Λ is positive, therefore $\mathrm{Tr}[\rho W]<0$ also implies that $(\mathbb{1}\otimes\Lambda)[\rho]$ is not positive. This shows that, if $(\mathbb{1}\otimes\Lambda)[\rho]$ is positive for any positive map Λ, then ρ is not entangled, which finishes the proof of the result obtained by the Horodeckis.

2.4 Summary

We summarize the main findings of this chapter and include some other results that might be helpful to grasp the concept of entanglement. All of them and much more can be found

in the review [7].

- A state is entangled over two parts or more, iff it cannot be written as a separable state
$$\rho = \sum_i p_i \, \rho_A^{(i)} \otimes \rho_2^{(i)} \otimes \dots \rho_N^{(i)}, \; p_i \in [0,1], \; \sum_i p_i = 1$$
- A theory that predicts a violation a CHSH inequality is incompatible with local realism. Quantum mechanics is such a theory.
- The constraint of no faster than light communication is not sufficient to recover the maximal violation of the CHSH inequalities predicted by quantum mechanics.
- A pure bipartite state is entangled if and only if it violates a CHSH inequality.
- Some mixed entangled bipartite states do not violate any CHSH inequality (e.g. Werner states).
- An entanglement measure is a sufficient criterion for entanglement that is monotonous under LOCC. There are many forms of entanglement measures, which generally do not agree. If we consider two states ρ and σ such that a first entanglement measure gives $E_1(\rho) > E_1(\sigma)$, a different measure may give $E_2(\rho) < E_2(\sigma)$. Some entangled states may not be detected by some measures.
- All measures of entanglement of a bipartite pure state are equivalent and proportional to the entropy of entanglement.
- Some mixed entangled bipartite states are not distillable, they are said to be bound entangled.
- A mixed or pure bipartite state ρ is entangled if and only if there exists a positive map Λ such that $(\mathbb{1} \otimes \Lambda)[\rho]$ is not positive.
- In 2×2 or 2×3 dimensions, it is enough to consider the positivity of the partial transpose to obtain a necessary and sufficient entanglement criterion. In higher dimension, the PPT criterion does not detect all entangled states.
- States with positive partial transpose are not distillable. The converse is an open question.
- Some separable states lead to measurement outcomes with non-classical correlations (quantum discord).

Exercises

- Consider the Horodecki state, which is the mixed state $p \left|\Phi_+\right\rangle \left\langle\Phi_+\right| + (1-p) \left|00\right\rangle \left\langle 00\right|$ with $\left|\Phi_+\right\rangle$ the Bell state $\left|\Phi_+\right\rangle = (\left|00\right\rangle + \left|11\right\rangle)/\sqrt{2}$. Is the state entangled ? Do you think a CHSH inequality can be violated for any p ?
- Consider the Werner state, which is the mixed state $(1-F)\mathbb{1}_4/3 + (4F-1)/3 \left|\Psi_-\right\rangle \left\langle\Psi_-\right|$. At which condition is the state entangled ?

- In 1997, Wooters showed that the concurrence is an entanglement measure for two qubit states. For a pure state, the concurrence is defined as $C = |\left\langle \phi \middle| \bar{\phi} \right\rangle|$, where $\left|\bar{\phi}\right\rangle = (\sigma_y \otimes \sigma_y) |\phi^*\rangle$ and $|\phi^*\rangle$ is the complex conjugate of $|\phi\rangle$ written in the standard z basis. Compute $\left|\bar{\phi}\right\rangle$ for a state of two spins pointing along $+z$, $-z$, $+x$, $-x$, ... What is the effect of the $\sigma_y \otimes \sigma_y$ operator ? What is the concurrence of these states ? Compute $\left|\bar{\phi}\right\rangle$ for the different Bell states, what is their concurrence ?

- For a mixed state ρ, the concurrence is defined as

$$C = \min \sum_i p_i C(|\phi_i\rangle)$$

  where the minimum is taken over all possible decomposition of $\rho = \sum_i p_i |\phi_i\rangle \langle\phi_i|$. We introduce the state $\bar{\rho} = (\sigma_y \otimes \sigma_y)\rho^*(\sigma_y \otimes \sigma_y)$, where ρ^* is the complex conjugate of ρ written in the z basis. Wooters has shown that the concurrence can be expressed as

$$C = \max [0, \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4]$$

  where the λ_i's are the square roots of the eigenvalues of $\rho\bar{\rho}$ sorted in descending order. Compute the concurrence of the Horodecki and Werner states.

References

[1]M. Froissart, "Constructive generalization of Bell's inequalities", Il Nuovo Cimento B **64**, 241–251 (1981).

[2]J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed Experiment to Test Local Hidden-Variable Theories", Physical Review Letters **23**, 880–884 (1969).

[3]B. S. Cirel'son, "Quantum generalizations of Bell's inequality", Letters in Mathematical Physics **4**, 93–100 (1980).

[4]S. Popescu and D. Rohrlich, "Quantum nonlocality as an axiom", Foundations of Physics **24**, 379–385 (1994).

[5]D. Rohrlich, "A Reasonable Thing That Just Might Work", in *Quantum nonlocality and reality*, edited by M. Bell and S. Gao (Cambridge University Press, Cambridge, 2016), pp. 295–304.

[6]M. Horodecki, P. Horodecki, and R. Horodecki, "Separability of mixed states: Necessary and sufficient conditions", Physics Letters, Section A: General, Atomic and Solid State Physics **223**, 1–8 (1996).

[7]R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, "Quantum entanglement", Reviews of Modern Physics **81**, 865–942 (2009).

Chapter 3

Quantum Information and Communication

Given a message consisting of letters x whose frequencies follow the probability distribution $p(x)$, the Shannon entropy is defined as

$$H(X) = -\sum_x p(x) \log p(x) \,, \tag{3.1}$$

where the log function is the base 2 logarithm. In 1948, Shannon introduced this quantity in his famous paper «Mathematical foundations of communication». Story says that this is von Neumann who advised Shannon to name this quantity entropy: «You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage» [1]. The Shannon entropy is central in the theory of information and establish a deep link between statistical physics and information theory. Extending these concepts to quantum systems is still an active field of research with some fundamental open questions.

In this chapter, we will give a brief overview of Shannon's two theorems, the source coding and the channel capacity theorems, and try to see how they may or may not be transposed to the quantum world. The natural replacement for the Shannon entropy is the von Neumann entropy, which will play a central role. We will try to emphasize the similarities as well as the differences between these two quantities. A close analogue to Shannon's source coding theorem is the Schumacher's quantum source coding theorem, which tells how much a message, where each letter is a pure quantum state, may be compressed. The second Shannon's theorem gives the capacity, or transmission rate, of a communication channel between an emitter and a receiver, taking into account that noise may disturb the transmission. Finding a counterpart to this theorem, when the communication channel becomes a quantum channel, appears to be a difficult problem with no definitive answer. But, we will see that, for a particular type of quantum channel, the capacity of the channel is given by a quantity called the Holevo bound. We will not mathematically prove this result but rather show that it is deeply connected to the validity of the second law of thermodynamics. To make a long story short, we will arrive at the conclusion that quantum channels do not have any advantage over classical channels in terms of transmission rate.

We will end the chapter by describing two protocols that are peculiar to quantum systems: the dense coding protocol and quantum teleportation. These protocols suppose that the emitter and the receiver share some entangled states prior to the communication.

3.1 Classical and quantum source coding theorems

3.1.1 Shannon source coding theorem

Shannon coding theorem applies to classical messages $x_1x_2\dots x_n$, where each letter x_i is chosen from an alphabet consisting of $\aleph$ different symbols. The first Shannon theorem states that a message of n letters can be compressed into a binary message consisting of nH bits with a decoding error that tends to zero as n tends to infinity. This compression rate is optimal, more compression would deteriorate the message, and reachable, meaning that coding schemes reaching this compression rate exist. We will not give a rigorous proof of the theorem but rather limit ourselves to simple arguments supporting Shannon's findings. Given a message $x_1x_2\dots x_n$, the probability for the message to be emitted is

$$P(x_1x_2\dots x_n) = p(x_1)p(x2)\dots p(x_n) \Rightarrow \log P(x_1x_2\dots x_n) = \sum_i \log p(x_i)\,.$$

If we consider the letters to be independent random variables, we obtain

$$\langle \log P(x_1x_2\dots x_n)\rangle = n\langle \log p(x)\rangle = n\sum_x p(x)\log p(x) = -nH(X)\,,$$

where the average $\langle\rangle$ is taken over all possible messages of length n. If we build an histogram of $P(x_1x_2\dots x_n)$, we obtain, when n is large, a distribution that is peaked around a value $P_{\rm typ}$, which corresponds to the probability of the most probable messages. We can then safely exchange log and $\langle\rangle$ and we see that the Shannon entropy is proportional to the log of the probability $P_{\rm typ}$ of the most probable messages, which are called the typical messages. We have that $P_{\rm typ} = 2^{-nH(X)}$. We now argue that it is sufficient to code for the typical messages because the probability for a message to be outside the set of typical messages will tend to zero for large n. This can be rigorously proven mathematically, but this is a common assumption in statistical physics. The law of large number says that the probability for $P(x_1x_2\dots x_n)$ to be close to its average value tends to one when n is large. For example, if x only takes two values 0 or 1 with probabilities $1-p$ and p, we know, from the theory of random walks, that, when n is large, the probability to obtain a sequence with m 1's is a gaussian centered around $m = np$ with a standard deviation $\sigma = \sqrt{np(1-p)}$. All messages such that $m = np \pm \alpha\sigma$ have a total probability close to one for α on the order of unity, they are the typical messages. The number of typical messages $N_{\rm typ}$ is such that $N_{\rm typ}P_{\rm typ} = 1$, from which we obtain

$$N_{\rm typ} = 2^{nH(X)}\,. \tag{3.2}$$

If x only takes two values 0 or 1 with probabilities p and $1-p$, this expression can also be obtained using the Stirling formula applied to the binomial coefficients, which gives $\log \binom{n}{pn} \approx nH(p)$.

A more rigorous statement of the theorem is to say that a message is δ typical if its probability is such that

$$2^{-n(H(X)+\delta)} \le P(x_1x_2\dots x_n) \le 2^{-n(H(X)-\delta)}\,.$$

Then, given a small ϵ, there exist a length n_0 such that the probability for a message of length $n \ge n_0$ to be typical is larger than $1-\epsilon$. Equivalently, the number of δ typical messages tends to $2^{nH(X)}$ when n tends to infinity and δ tends to zero.

A possible compression scheme consists in attributing a code formed of $nH(X)$ bits to every $2^{nH(X)}$ typical messages. If all the letters happen with the same probability, we have

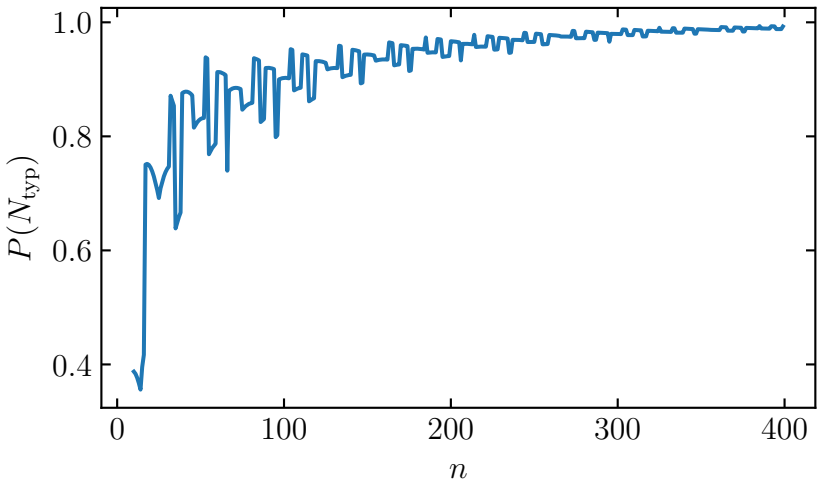


Figure 3.1: Evolution of the probability of a message to be in the set of typical messages as a function of the message length n. We consider binary messages with a probability $p = 0.1$ to have a bit set to one. The probability $P(N_{\text{typ}})$ is the probability of the $N_{\text{typ}} = 2^{n[H(p)+\epsilon]}$ most probable messages, here shown for $\epsilon = 0.1$.

$H(X) = \log \aleph$, where $\aleph$ is the number of letters in the alphabet, and the number of typical messages is equal to the number of possible messages $\aleph^n$. But if some letters are more likely than others, the entropy decreases and some messages become very unlikely and can be safely discarded (see figure 3.1). The Shannon entropy represents our ignorance about the message, the greater it is, the more bits are needed to specify the message.

3.1.2 Schumacher quantum source coding theorem

We now consider the case of a quantum message $|\phi\rangle = |x_1\rangle \otimes |x_2\rangle \otimes \ldots |x_n\rangle$, which is a product state of pure states $|x_i\rangle$. The states $|x_i\rangle$ play the role of the letters. The main difference with the previous situation is that the letters may not be orthogonal states and thus not fully distinguishable. If the $|x_i\rangle$ are orthogonal, one can guess that the classical result of Shannon will be valid and that $nH(X)$ qubits will encode the message optimally. If it is not the case, we will show that the Shannon entropy has to be replaced by the von Neumann entropy, which can only be smaller.

The quantum state of a message with n letters is $|\phi\rangle = |x_1\rangle \otimes |x_2\rangle \otimes \ldots |x_n\rangle$, where, as in the classical case, the letter state $|x\rangle$ appears with probability $p(x)$. Averaging over all possible messages, we obtain the density matrix

$$\rho_m = \rho^{\otimes n} \text{ with } \rho = \sum_x p(x) \, |x\rangle \langle x| \, ,$$

where ρ is the density matrix of a single letter. We introduce the eigenvectors $|y\rangle$ of ρ with eigenvalue $p(y)$. The von Neumann entropy of ρ is given by the Shannon entropy $H(Y)$:

$$S(\rho) = -\operatorname{Tr}[\rho \log \rho] = H(Y) \leq H(X) \, ,$$

where the equality is obtained iff the $|x\rangle$ states are orthogonal (see exercise 1). We now adapt the reasoning of Shannon to the quantum case. A message $|y_1\rangle \otimes |y_2\rangle \otimes \ldots \otimes |y_n\rangle$ is

an eigenstate of ρ_m with the eigenvalue $p(y1)p(y2)\dots p(y_n)$. The most likely eigenvalue is Λ such that $\log\Lambda = -nH(Y)$. The typical states are the states whose eigenvalue is close to Λ. The sum of the eigenvalues of typical states tend to one when n is large, so the number of typical states is $N_{\text{typ}} = 1/\Lambda = 2^{nH(Y)}$. These states span a subspace of dimension $2^{nH(Y)}$, if we introduce Π_{typ} the projector on this subspace, we have that $\text{Tr}[\Pi_{\text{typ}}\rho_m] = 1-\epsilon$ with $\epsilon \to 0$ as $n \to \infty$.

Therefore, Alice only needs $nH(Y)$ qubits to faithfully encode her message. The encoding is a two step process where Alice first projects the message $|\phi\rangle$ on the subspace of typical messages and then encode the projected state onto $nH(Y)$ qubits with a unitary operation U. If the projection does not succeed, she sends some arbitrary state $|\phi_{\text{fail}}\rangle$ to Bob. The density matrix of the encoded message is

$$U\Pi_{\text{typ}}|\phi\rangle\langle\phi|\Pi_{\text{typ}}U^\dagger + \langle\phi|\mathbb{1}-\Pi_{\text{typ}}|\phi\rangle|\phi_{\text{fail}}\rangle\langle\phi_{\text{fail}}| \,.$$

Bob decodes the message and obtain

$$\rho_B = \Pi_{\text{typ}}|\phi\rangle\langle\phi|\Pi_{\text{typ}} + \langle\phi|\mathbb{1}-\Pi_{\text{typ}}|\phi\rangle|\phi_{\text{garb}}\rangle\langle\phi_{\text{garb}}| \,.$$

The fidelity of the transmission is given by

$$F(\phi) = \langle\phi|\rho_B|\phi\rangle = |\langle\phi|\Pi_{\text{typ}}|\phi\rangle|^2 + \langle\phi|\mathbb{1}-\Pi_{\text{typ}}|\phi\rangle|\langle\phi|\phi_{\text{garb}}\rangle|^2 \geq |\langle\phi|\Pi_{\text{typ}}|\phi\rangle|^2 \,.$$

The probability that the state $|\phi\rangle$ is successfully projected on the subspace of typical messages is $\langle\phi|\Pi_{\text{typ}}|\phi\rangle = 1-\epsilon_\phi$. We have $F(\phi) \geq 1-2\epsilon_\phi$, from which we obtain that the average fidelity is bounded by

$$F \geq 1-2\epsilon \,. \tag{3.3}$$

At least $nS(\rho)$ qubits are needed to code the messages without error when n is large. And this compression rate is reachable as demonstrated by Schumacher in [2].

3.2 Capacity of a classical communication channel

We now consider the case of the transmission of a classical message $x_1x_2\dots x_n$ through a communication channel that outputs the message $y_1y_2\dots y_n$. If the channel does not introduce errors, the two messages are identical. If transmission errors happen, the two strings will differ but should still be correlated. The behavior of the channel is fully described by the conditional probability $p(y|x)$, which is the probability to observe y given x.

3.2.1 Mutual information

The mutual information is defined by how much our ignorance on the emitted message is reduced when we read the transmitted message

$$I(X:Y) = H(X) - H(X|Y) \,. \tag{3.4}$$

The joint probability distribution of the letters in the emitted and received messages verify the following relations

$$p(xy) = p(y|x)p(x) = p(x|y)p(y) \;\Rightarrow\; H(XY) = H(Y|X) + H(X) = H(X|Y) + H(Y) \,.$$

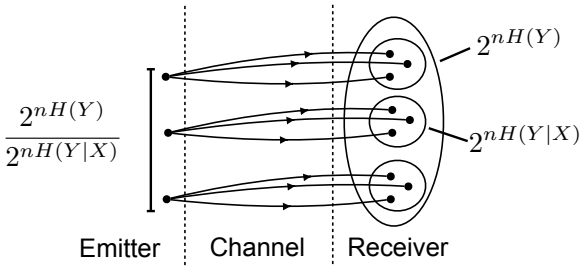
The mutual information can thus be rewritten

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

If the emitted and received messages have no correlation, meaning that $p(xy) = p(x)p(y)$ then $H(XY) = H(X) + H(Y)$ and the mutual information is zero. This corresponds to the case of a very lossy channel. In the opposite case of a perfectly faithful channel, the received message is perfectly correlated to the emitted one and $p(xy) = p(x) = p(y)$. We then obtain $H(XY) = H(X) = H(Y)$ and the mutual information is maximal $I(X:Y) = H(X) = H(Y)$. The amount of information gained when the message is received is equal to the entropy of the message before going through the communication channel, which has no detrimental effect. In between these two extreme cases, the mutual information is a positive quantity bounded by $H(X)$ and $H(Y)$.

3.2.2 Shannon channel capacity theorem

The second Shannon theorem establishes that the capacity of a channel to transmit information is given by the maximal mutual information $I(X:Y)$ that can be obtained between the emitter and the receiver. In order to prove this result, we count the number of messages that can be faithfully transmitted over the channel. Because of the noise over the channel, a message $x_1x_2\ldots x_n$ on the emitter side is transmitted to a message $y_1y_2\ldots y_n$ with probability $p(y_1|x_1)p(y_2|x_2)\ldots p(y_n|x_n)$. The typical number of such messages is $2^{nH(Y|X)}$. By averaging over the initial messages, we obtain that the total number of typical messages is $2^{nH(Y)}$ can be partitioned as shown below:



The receiver can distinguish between the different emitted messages if their typical sets after transmission do not overlap. We therefore obtain that the number of messages that can be faithfully transmitted is

$$\frac{2^{nH(Y)}}{2^{nH(Y|X)}} = 2^{nI(X:Y)}\,.$$

The transmission of n letters results in messages that are coded over $nI(X:Y)$ bits and $I(X:Y)$ is thus the transmission rate in bits per sent letter. This rate depends on $p(y|x)$ but also on how the message is encoded on the emitter side. The capacity of the channel is the optimal transmission rate obtained by maximizing $I(X:Y)$ over the possible distributions $p(x)$:

$$C = \max_X I(X:Y)\,. \tag{3.5}$$

This relation constitutes the Shannon noisy channel capacity theorem.

3.3 Quantum mutual information

3.3.1 Relative entropy

Given the importance of mutual information to quantify the capacity of a channel, it is natural to wonder how this notion can be extended to the quantum case. We will do so by taking a step back and notice that the mutual information is a specific case of a more general quantity called the relative entropy. The classical relative entropy between two distributions $p(x)$ and $q(x)$ quantifies how different they are. It is defined as

$$D(p(x)||q(x)) = \sum_x p(x)\log p(x) - p(x)\log q(x) = \sum_x p(x)\log\frac{p(x)}{q(x)}\,.$$

It is null if the two distributions are identical and it is always positive because $\log x^{-1} \geq 1-x$ for $x \geq 0$:

$$D(p(x)||q(x)) \geq \sum_x p(x)\left(1-\frac{q(x)}{p(x)}\right) = 0\,.$$

The mutual information can be defined as the relative entropy between the joint distribution $p(xy)$ and the uncorrelated distribution $p(x)p(y)$ obtained from the product of the marginal distributions $p(x) = \sum_y p(xy)$ and $p(y) = \sum_x p(xy)$. We have

$$\begin{aligned} D(p(xy)||p(x)p(y)) &= \sum_{xy} p(xy)\log p(xy) - p(xy)\log p(x) - p(xy)\log p(y) \\ &= \sum_{xy} p(xy)\log p(xy) - \sum_x p(x)\log p(x) - \sum_y p(y)\log p(y) \\ &= H(X) + H(Y) - H(XY) = I(X:Y)\,. \end{aligned}$$

In the quantum case, we define the quantum relative entropy between two density matrices ρ and σ as

$$D(\rho||\sigma) = \mathrm{Tr}[\rho\log\rho - \rho\log\sigma]\,. \tag{3.6}$$

If ρ and σ commute, they can be co-diagonalized in the same basis and their relative entropy is given by the classical relative entropy between the two sets of eigenvalues. The quantum relative entropy fulfills two important properties: positivity and monotonicity.

Positivity of the quantum relative entropy (Klein inequality)

We will now show that

$$D(\rho||\sigma) = \mathrm{Tr}[\rho\log\rho - \rho\log\sigma] \geq 0 \tag{3.7}$$

We write ρ and σ in their eigenbasis as $\rho = \sum_x p(x)\,|\phi_x\rangle\langle\phi_x|$ and $\sigma = \sum_x q(x)\,|\psi_x\rangle\langle\psi_x|$. and introduce the probability distribution $\tilde{q}(x) = \sum_{x'} |\langle\phi_x|\psi_{x'}\rangle|^2 q(x')$. One easily checks that $\sum_x \tilde{q}(x) = 1$. The relative entropy between ρ and σ is

$$\begin{aligned} D(\rho||\sigma) &= \mathrm{Tr}[\rho\log\rho - \rho\log\sigma] \\ &= \sum_x p(x)\log p(x) - p(x)\langle\phi_x|\log\sigma\,|\phi_x\rangle \\ &= \sum_x p(x)[\log p(x) - \sum_{x'} |\langle\phi_x|\psi_{x'}\rangle|^2 \log q(x')] \\ &\geq \sum_x p(x)[\log p(x) - \log\sum_{x'} |\langle\phi_x|\psi_{x'}\rangle|^2 q(x')] \\ &\geq \sum_x p(x)[\log p(x) - \log\tilde{q}(x)] = D(p(x)||\tilde{q}(x)) \geq 0\,. \end{aligned}$$

where we have used the concavity of the log function and $\sum_{x'} |\langle\phi_x|\psi_{x'}\rangle|^2 = 1$.

Monotonicity of the quantum relative entropy

Another important result is that $D(\rho||\sigma)$ decreases under the action of a quantum channel. For any quantum channel (CPTP map), the following inequality holds

$$D(\Lambda[\rho]||\Lambda[\sigma]) \leq D(\rho||\sigma) \tag{3.8}$$

This result follows from a lengthy and technical linear algebra calculation, a proof can be found for example in the chapter five of [3]. We will see in section 3.6 that these two properties are very fundamental, because they generalize at the quantum level some important properties of distribution functions that are required for the second law of thermodynamics to be valid.

3.3.2 Quantum mutual information from relative entropy

Following the classical definition, we define the quantum mutual information as

$$I^{A:B}(\rho) = D(\rho||\rho_A \otimes \rho_B) \text{ with } \rho_A = \mathrm{Tr}_B\,\rho,\ \rho_B = \mathrm{Tr}_A\,\rho\,. \tag{3.9}$$

By construction, it is a positive quantity that vanishes when the state is separable, $\rho = \rho_A \otimes \rho_B$. We can rewrite it in terms of the von Neumann entropy of the whole system and the reduced entropy of each part as

$$\begin{aligned} I^{A:B}(\rho) &= \mathrm{Tr}[\rho\log\rho - \rho\log\rho_A\otimes\rho_B] \\ &= \mathrm{Tr}[\rho\log\rho - \rho(\log\rho_A\otimes\mathbb{1}) - \rho(\mathbb{1}\otimes\log\rho_B)] \\ &= S(\rho_A) + S(\rho_B) - S(\rho)\,. \end{aligned}$$

We obtain the sub-additivity of the von Neumann entropy $S(\rho_A)+S(\rho_B) \geq S(\rho)$. From this expression, we see that strange things happen for a pure entangled state. In the previous chapter, we have seen that for a maximally entangled pure state $S = S(\rho_A) = S(\rho_B)$ and $S(\rho) = 0$. We then obtain

$$I^{A:B}(\rho) = 2S\,,$$

which is two times more than the classical limit. We also see that the quantity $S(\rho)-S(\rho_A)$ or $S(\rho)-S(\rho_B)$, which would correspond classically to the conditional entropy, is negative. This negativity of the conditional entropy is actually a sufficient criterion to detect entanglement.

3.3.3 Monotonicity of the quantum mutual information

Monotonicity of the quantum relative entropy immediately implies the monotonicity of the mutual quantum information. We consider the action of a quantum map $\Lambda\otimes\mathbb{1}$ on ρ. We have that

$$\begin{aligned} \mathrm{Tr}_B[(\Lambda\otimes\mathbb{1})[\rho]] &= \Lambda[\rho_A] \\ \mathrm{Tr}_A[(\Lambda\otimes\mathbb{1})[\rho]] &= \rho_B\,. \end{aligned}$$

We can then write the mutual information after evolution as

$$\begin{aligned} I^{A:B}((\Lambda\otimes\mathbb{1})[\rho]) &= D((\Lambda\otimes\mathbb{1})[\rho]||\Lambda[\rho_A]\otimes\rho_B) \\ &= D((\Lambda\otimes\mathbb{1})[\rho]||(\Lambda\otimes\mathbb{1})[\rho_A\otimes\rho_B]) \\ &\leq D(\rho||\rho_A\otimes\rho_B)\,. \end{aligned}$$

We thus obtain

$$I^{A:B}((\Lambda \otimes \mathbb{1})[\rho]) \leq I^{A:B}(\rho) \tag{3.10}$$

3.3.4 Strong sub-additivity

We consider a density matrix ρ over a tripartite system A, B, C and the map $\mathbb{1}_A \otimes \mathrm{Tr}_C$, where Tr_C is the partial trace map acting on the BC subsystem. One can show that

$$I^{A:B}(\rho_{AB}) \leq I^{A:BC}(\rho)\,. \tag{3.11}$$

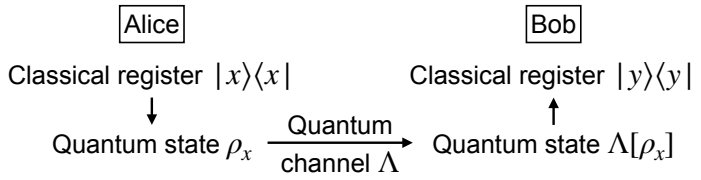
On the right hand side, the mutual information is defined as the relative entropy between ρ and $\rho_A \otimes \rho_{BC}$. On the left hand side, the reduced density matrix is $\rho_{AB} = \mathrm{Tr}_C[\rho]$. We can rewrite this inequality using von Neumann entropies and obtain the strong sub-additivity of the von Neumann entropy

$$S(\rho_A) + S(\rho_B) - S(\rho_{AB}) \leq S(\rho_A) + S(\rho_{BC}) - S(\rho)\,. \tag{3.12}$$

The proof of this result is a bit technical and is similar to the one of the monotonicity that we already admitted. We again refer the reader to [3] for the proof. The result is however not very surprising. It would be strange if tracing over the subsytem C could increase the mutual information between A and B.

3.4 Holevo bound

As before, we suppose that Alice sends the message x with probability $p(x)$. She now encodes her message as the quantum state ρ_x and sends it to Bob over a quantum channel Λ. Bob receives $\Lambda[\rho_x]$ and performs a POVM to obtain the outcome y with probability $p(y)$. The classical message x of Alice may be represented by the pure state $|x\rangle\langle x|$. Bob decodes the message and obtains a classical register $|y\rangle\langle y|$ as shown below:



The different states $|x\rangle$ and $|y\rangle$ are orthogonal, therefore fully distinguishable, and have the same properties than classical messages. The final density matrix between the two classical registers may be written as $\sum_{xy} p(xy)\,|x\rangle\langle x| \otimes |y\rangle\langle y|$. From the second Shannon theorem, we know that the amount of information shared by Alice and Bob is the mutual information $I(X:Y)$ associated to the probability distribution $p(xy)$. The maximum mutual information that Bob and Alice can obtain after the communication can in general not be computed. But Holevo derived a bound to the mutual information, which is now known as the Holevo bound [4]. We will now derive its expression.

Prior to the communication, the density matrix on Alice's side is

$$\rho = \sum_x p(x)\,|x\rangle\,\langle x| \otimes \rho_x\,.$$

Bob performs a POVM measurement with Kraus operator M_y and stores the result in his classical register. The density matrix becomes

$$\rho' = \sum_{xy} p(x) \underbrace{|x\rangle \langle x|}_{A} \otimes \underbrace{M_y \Lambda[\rho_x] M_y^\dagger}_{B} \otimes \underbrace{|y\rangle \langle y|}_{C} .$$

The classical mutual information shared by Alice and Bob is the quantum mutual information between the A and C subsystems. Using the strong sub-additivity and the monotonicity properties, we can bound this mutual information and obtain

$$I^{A:C}(\rho') \leq I^{A:B}(\rho) = H(X) + S\left(\sum_x p(x)\rho_x\right) - S(\rho) .$$

If we rewrite the von Neumann entropy $S(\rho)$ as

$$\begin{aligned} S(\rho) &= \mathrm{Tr}_{AB}[(\sum_x p(x) |x\rangle \langle x| \otimes \rho_x)(\sum_x \log p_x |x\rangle \langle x| \otimes \mathbb{1} + \sum_x |x\rangle \langle x| \otimes \log \rho_x)] \\ &= \mathrm{Tr}_A[\sum_x p(x) \log p_x |x\rangle \langle x|] + \mathrm{Tr}_B[\sum_x p_x \rho_x \log \rho_x] \\ &= H(X) + \sum_x p(x) S(\rho_x) , \end{aligned}$$

we obtain

$$I^{A:C}(\rho') \leq \chi(\{p(x); \rho_x\}) = S\left(\sum_x p(x)\rho_x\right) - \sum_x p(x) S(\rho_x) . \tag{3.13}$$

The quantity χ is the Holevo bound, which is also called the Holevo-χ. If the ρ_x are pure states $|\phi_x\rangle \langle \phi_x|$, it reduces to the first term, which is the von Neumann entropy of the density matrix $\sum_x p(x) |\phi_x\rangle \langle \phi_x|$. This entropy is maximal when the states are orthogonal and the maximum of χ is $H(X)$. This means than no more than n bits of classical information can be stored into n qubits.

3.5 Classical capacity of a quantum channel

The previous paragraph gives a limit to the amount of classical information but does not constitute a proper generalization of the second Shannon theorem to the quantum case. Finding such a generalization is still an active domain of research and is a difficult problem. For example, in the general case where the message may be a multipartite entangled states and if the channel is used to send each part of the entangle state, the channel capacity per letter cannot be properly defined, because the mutual information that Alice and Bob can obtain after n uses of the channel may not be linear with n. Another big difference with the classical case is that the amount of information may depend on some other resources than the channel itself. For example, if Alice and Bob share some entangled states that they obtained prior to the communication, they can use them to improve the amount of information that they obtain from the use of the channel.

In order to remain in a framework that resembles the classical one, we first restrict ourselves to the case where the messages that are sent over the channel are not entangled. We also suppose that this is the only resource that is used by Alice and Bob. In the next section, we will give two examples of entanglement assisted protocols.

3.5.1 Saturating the Holevo bound: the HSW theorem

We now suppose that Alice prepares a message of n letters $\rho(\vec{x}) = \rho_{x_1} \otimes \rho_{x_2} \otimes \ldots \otimes \rho_{x_n}$ and sends it to Bob, who must find the optimal POVM that maximizes the mutual information. The Holevo-Schumacher-Westmoreland theorem states that there exists an optimal POVM, which is called the Pretty Good Measurement, that allows Alice and Bob to saturate the Holevo bound when n is large [5–7]. We will not prove this result but rather describe the optimal PGM measurement and show on a case study that this measurement indeed performs better than a more naive decoding scheme.

Letter by letter decoding: the Helstrom bound

We start by considering a naive decoding scheme, where Bob decodes the message letter by letter. We suppose that Alice uses only two letters ρ_0 and ρ_1 to compose her message. In this case, the optimal way to distinguish between the two letters is known and the associated error is known as the Helstrom bound. In order to decode each letter, Bob uses a POVM consisting of two operators E_0 and E_1. Bob obtains a correct result if he obtains the outcome E_0 for the state ρ_0 and the outcome E_1 for ρ_1. Otherwise the letter is incorrectly decoded. The average probability of error is

$$P_{\text{err}} = p_0 \operatorname{Tr}[\rho_0 E_1] + p_1 \operatorname{Tr}[\rho_1 E_0] \,,$$

where p_0 and p_1 are the probabilities that Alice sends the letter ρ_0 or ρ_1. Using $p_0 + p_1 = 1$ and $E_0 + E_1 = \mathbb{1}$, the error probability can be rewritten as

$$P_{\text{err}} = 1 + \operatorname{Tr}[(p_0\rho_0 - p_1\rho_1)E1] = 1 - \operatorname{Tr}[(p_0\rho_0 - p_1\rho_1)E0]$$

We define the operator $K = (p_0\rho_0 - p_1\rho_1)$. It is hermitian and can diagonalized as

$$K = \sum_{k_- \leq 0} k_- \left|k_-\right\rangle\!\left\langle k_-\right| + \sum_{k_+ > 0} k_+ \left|k_+\right\rangle\!\left\langle k_+\right|$$

where we separate the negative eigenvalues k_- from the positive ones k_+. In order to minimize the error, one sees that E_0 should be the projector on positive eigenvalues, $E_0 = \sum_{k_+>0} |k_+\rangle\langle k_+|$, and E_1 the projector on negative eigenvalues, $E_1 = \sum_{k_->0} |k_-\rangle\langle k_-|$. The two operators E_0 and E_1 are complementary orthogonal projectors and define a projective measurement. The probability of error is given by

$$P_{\text{err}} = 1 + \sum_{k_- \leq 0} k_- = 1 - \sum_{k_+ > 0} k_+$$

from which we obtain

$$P_{\text{err}} = \frac{1}{2} + \frac{1}{2} \sum_k |k| = \frac{1}{2} + \frac{1}{2} \operatorname{Tr}[|p_0\rho_0 - p_1\rho_1|] \,.$$

In the case where ρ_0 and ρ_1 are pure states $|\phi_0\rangle\langle\phi_0|$ and $|\phi_1\rangle\langle\phi_1|$, the eigenvalues of K can be explicitly computed, and we obtain

$$P_{\text{err}} = \frac{1}{2} + \frac{1}{2}\sqrt{1 - 4p_0p_1 \left|\left\langle\phi_0|\phi_1\right\rangle\right|^2} \,. \tag{3.14}$$

This formula is known as the Helstrom bound and gives the minimal error that can be achieved when distinguishing two pure states. As expected, the probability of error is zero iff the two states are orthogonal.

As an example, we consider that Alice uses two letters that are pure states defined as

$$|0\rangle = \cos\frac{\theta}{2}\,|\uparrow\rangle + \sin\frac{\theta}{2}\,|\downarrow\rangle$$
$$|1\rangle = \cos\frac{\theta}{2}\,|\uparrow\rangle - \sin\frac{\theta}{2}\,|\downarrow\rangle\ ,$$

where the angle θ sets the degree of distinguishability between the two states. If she uses the two letters with equal probability, we have proven that the optimal measurement is a set of orthogonal projectors, which are obtained from the eigenstates of $|0\rangle\langle 0| - |1\rangle\langle 1|$. Here, it corresponds to a measurement in the x basis. The probability that Bob mistakes one state for another is given by the Helstrom bound

$$\epsilon = \frac{1}{2}(1 - \sqrt{1 - |\,\langle 0|1\rangle\,|^2}) = \frac{1}{2}(1 - \sin\theta)$$

It corresponds to the probability $p(1|0) = p(0|1) = \epsilon$, the probability to obtain the correct result is $p(0|0) = p(1|1) = 1 - \epsilon$. The mutual information per letter is therefore

$$\begin{aligned} I_1 &= H(Y) - H(Y|X) \\ &= 1 + (1-\epsilon)\log(1-\epsilon) + \epsilon\log\epsilon\,. \end{aligned}$$

For a message of n letters, Alice and Bob obtain a mutual information nI_1. We can compare it to the Holevo bound of the message. Because we consider product states of pure states, the Holevo bound is given by n times the von Neumann entropy per letter. The Holevo χ per letter is thus

$$\chi = S\left(\frac{|0\rangle\langle 0| + |1\rangle\langle 1|}{2}\right) = -\cos^2\frac{\theta}{2}\log\cos^2\frac{\theta}{2} - \sin^2\frac{\theta}{2}\log\sin^2\frac{\theta}{2}$$

The left panel of figure 3.2 compares I_1 and χ as a function of the mixing angle θ. We observe that I_1 is always below χ showing that this naive decoding scheme falls short of the Holevo bound.

Pretty good measurement

In order to saturate the Holevo bound, Alice and Bob must adopt a more clever approach. In particular, Bob must decode the messages not letter by letter but as a whole. He must thus distinguish between many not orthogonal states. In this case, the optimal measurement is in general not known. But a solution, which is known as the "Pretty Good Measurement" (PGM), is often nearly optimal [8]. Given a set of states $\rho_i = |\phi_i\rangle\langle\phi_i|$, the PGM is constructed as follows. One introduces the operator[1] $\Pi = \sum_i \rho_i$ and the set of operators $E_i = \Pi^{-1/2}\rho_i\Pi^{-1/2}$. It is easy to check that $\sum_i E_i = \mathbb{1}$ and that $E_i \geq 0$ and therefore the E_i define a POVM. If $\rho_i = |\phi_i\rangle\langle\phi_i|$ and the $|\phi_i\rangle$ are orthogonal, then $E_i = |\phi_i\rangle\langle\phi_i|$ and we recover the optimal set of orthogonal projectors.

We now suppose that Bob has received a n letter message and decodes it at once using the PGM. If Alice and Bob use the set of 2^n possible messages, this strategy is strictly equivalent to the letter by letter strategy. But, if Alice and Bob restrict the set of messages to the ones that are the most distinguishable, they can reduce the detection error and obtain

[1] In this paragraph, the operator definitions are restricted to the support of Π.

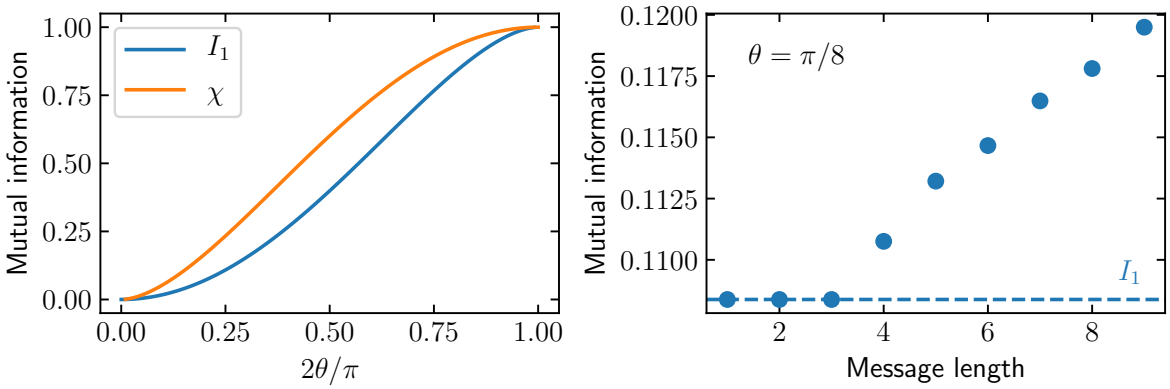


Figure 3.2: Left: Comparison of the Holevo bound χ with the mutual information per letter I_1 obtained from a letter by letter decoding of the message as a function of the mixing angle θ. Right: Alice and Bob can improve their mutual information per letter when the message length increases. Alice must choose a reduced set of messages and Bob can use a "pretty good measurement" to distinguish between the different messages.

more information. This is not an obvious result, because reducing the number of messages also reduces $H(Y)$. But the reduction of the measurement error reduces $H(Y|X)$ even more, resulting in an increase of the mutual information $I(X:Y)$.

For example, when $n = 4$, limiting the code space to the eight messages $|0000\rangle$, $|1111\rangle$, $|0101\rangle$, $|1010\rangle$, $|1001\rangle$, $|0110\rangle$, $|1100\rangle$ and $|0011\rangle$ leads to a conditional entropy

$$H(Y|X) = \epsilon_4 \log(\epsilon_4) + 6\epsilon_2 \log(\epsilon_2) + (1 - \epsilon_4 - 6\epsilon_2) \log(1 - \epsilon_4 - 6\epsilon_2) \tag{3.15}$$

where ϵ_2 is the probability to mistake one state with one of the six others that differ by two bits and ϵ_4 is the probability to mistake one state with the one that differ by four bits. Considering the PGM associated to these eight states, these probabilities can be computed numerically and one finds that the mutual information is slightly larger than $4I_1$ when, for example, $\theta = \pi/8$.

In the right panel of figure 3.2, we perform a naive implementation of this global decoding method as follows. For each n, we look for the subset of $m < 2^n$ code words with largest entropy and compute the mutual information per letter I_m obtained with the PGM over this subset. Increasing m, we observe that I_m increases, passes by a maximum and then decreases to I_1. In the figure, we plot the optimal value that we obtain. This algorithm is not optimal but still demonstrates that a global measurement performs better than decoding the message letter by letter.

HSW theorem

In a series of papers [5–7], Holevo, Schumacher, Westmoreland and coauthors have demonstrated that this "pruning" technique followed by PGM allows Alice and Bob to reach the Holevo bound when n is large. The construction of the PGM is obtained as follows (see [9] for more details). We now consider that Alice prepares her message with letters taken from the alphabet ρ_j with probability p_j. The density matrix for one letter of the message is $\rho = \sum_j p_j \rho_j$. We introduce its eigen decomposition as $\rho = \sum_l q_l \, |l\rangle\langle l|$. The density matrix of the message is $\rho^{\otimes n} = \sum_{\vec{l}} q_{\vec{l}} \left|\vec{l}\right\rangle\!\left\langle\vec{l}\right|$, where $q_{\vec{l}} = q_{l_1} \dots q_{l_n}$ and $\left|\vec{l}\right\rangle = |l_1\rangle \otimes \dots \otimes |l_n\rangle$. The

typical states $\left|\vec{l}\right\rangle$, which are those with a probability $q_{\vec{l}} \approx 2^{-nS[\rho]}$, span the subspace $\mathscr{H}_{\text{typ}}$. As in the demonstration of the Schumacher quantum source coding theorem, one can show that this subspace contains with high probability the message sent by Alice. We define the projector $\Pi_{\text{typ}} = \sum_{\text{typ}} \left|\vec{l}\right\rangle\!\left\langle\vec{l}\right|$, where the sum goes over the typical states. Then Π_{typ} projects onto $\mathscr{H}_{\text{typ}}$ and the probability $\text{Tr}[\rho^{\otimes n}\Pi_{\text{typ}}]$ is close to one.

We now decompose each letter of the alphabet as $\rho_j = \sum_k \lambda_k^{(j)} \left|e_k^{(j)}\right\rangle\!\left\langle e_k^{(j)}\right|$. A given message $\rho(\vec{x})$ writes $\rho(\vec{x}) = \sum_{\vec{k}} \lambda_{\vec{k}}^{(\vec{x})} \left|e_{\vec{k}}^{(\vec{x})}\right\rangle\!\left\langle e_{\vec{k}}^{(\vec{x})}\right|$, where $\lambda_{\vec{k}}^{(\vec{x})} = \lambda_{k_1}^{(x_1)} \dots \lambda_{k_n}^{(x_n)}$ and $\left|e_{\vec{k}}^{(\vec{x})}\right\rangle = \left|e_{k_1}^{(x_1)}\right\rangle \otimes \dots \otimes \left|e_{k_n}^{(x_n)}\right\rangle$. When n is large, the sum is dominated by the typical states such that $\lambda_{\vec{k}}^{(\vec{x})}$ is close to $2^{-n(S[\rho]-\chi)}$, where χ is the Holevo bound. The message can be identified by the projector $\Pi_{\vec{x}} = \sum_{\text{typ}} \left|e_{\vec{k}}^{(\vec{x})}\right\rangle\!\left\langle e_{\vec{k}}^{(\vec{x})}\right|$. These projectors are in general not orthogonal. The PGM used by Bob is thus obtained by considering the POVM set composed of the operators

$$\left(\sum_{\vec{x}} \Pi_{\text{typ}}\Pi_{\vec{x}}\Pi_{\text{typ}}\right)^{-1/2} \Pi_{\text{typ}}\Pi_{\vec{x}}\Pi_{\text{typ}} \left(\sum_{\vec{x}} \Pi_{\text{typ}}\Pi_{\vec{x}}\Pi_{\text{typ}}\right)^{-1/2}$$

With this POVM, one can show that the averaged probability of error, where the average is taken over all the possible codes chosen by Alice, tends to zero when n becomes large.

This allows us to define the classical capacity C of a quantum channel. So far, we have supposed that Alice directly gives her state to Bob, but we could consider that she sends each letter through a quantum channel Λ and that Bob receives $\Lambda[\rho_{x_1}] \otimes \Lambda[\rho_{x_2}] \otimes \dots \otimes \Lambda[\rho_{x_n}]$. The capacity of the channel is the maximum of χ when optimizing over ρ_x and $p(x)$.

$$C = \max_{\{p(x);\rho_x\}} \chi\left(\{p(x); \Lambda[\rho_x]\}\right) . \tag{3.16}$$

The strong limitation of this formula is that we do not allow for entanglement between the letters that are received by Bob: he always receives a string of separable states. Allowing the use of entangled states can actually increase the channel capacity, but so far a full theory of channel capacity including the use of entanglement is still to be discovered. More information can be found in [10], which is a nice introduction written by Shor to the problem of determining the capacity of quantum channels.

3.6 Relations with thermodynamics

In this section, we briefly comment on the thermodynamic interpretation of the properties of the quantum relative entropy that we assumed above. We consider a quantum system in contact with a reservoir at temperature T. The free energy of the system is $F = U - k_B T S$, where U is the mean energy and S the von Neumann entropy. If the state of the system is described by a density matrix ρ, we obtain

$$F(\rho) = \text{Tr}[\rho H + k_B T \rho \log \rho] ,$$

where H is the Hamiltonian of the system. At equilibrium, the density matrix is

$$\rho_{\text{eq}} = e^{-\beta H} / \text{Tr}\left[e^{-\beta H}\right] .$$

We obtain that the free energy of the equilibrium density matrix is

$$\begin{aligned} F(\rho_{\text{eq}}) &= \text{Tr}\left[e^{-\beta H} H + k_B T e^{-\beta H}(-\beta H - \log \text{Tr}\left[e^{-\beta H}\right])\right] / \text{Tr}\left[e^{-\beta H}\right] \\ &= -k_B T \log \text{Tr}\left[e^{-\beta H}\right] . \end{aligned}$$

The relative entropy between a generic state ρ and the equilibrium distribution is

$$\begin{aligned} D(\rho||\rho_{\text{eq}}) &= \text{Tr}[\rho \log \rho - \rho \log \rho_{\text{eq}}] \\ &= \beta \,\text{Tr}\Big[k_B T \rho \log \rho + \rho H + k_B T \rho \log \text{Tr}\Big[e^{-\beta H}\Big]\Big] \\ &= \beta[F(\rho) - F(\rho_{\text{eq}})]\,. \end{aligned}$$

Using the positivity of the quantum relative entropy, we obtain that $F(\rho) \geq F(\rho_{\text{eq}})$. This proves that the equilibrium distribution is the one that minimizes the free energy of the system.

We can also consider the evolution towards equilibrium through the action of a quantum map that leaves the equilibrium distribution unchanged, $\Lambda[\rho_{\text{eq}}] = \rho_{\text{eq}}$. From the monotonicity of the quantum relative entropy, we obtain

$$D(\Lambda[\rho]||\rho_{\text{eq}}) = D(\Lambda[\rho]||\Lambda[\rho_{\text{eq}}]) \leq D(\rho||\rho_{\text{eq}})\,,$$

which gives $\beta[F(\Lambda[\rho]) - F(\rho_{\text{eq}}) \leq \beta[F(\rho) - F(\rho_{\text{eq}})]$. The free energy can thus only decrease during the evolution $F(\Lambda[\rho]) \leq F(\rho)$. To conclude, we see that the properties 3.7 and 3.8 transpose to the quantum case some important properties of classical distributions obtained in statistical physics, in agreement with the second principle of thermodynamics. In other words, the relation (3.8), which we did not prove, has to be true, otherwise we could find situations where the second law of thermodynamics is violated. More on this topic can be found in [11] and [12].

3.7 Entanglement assisted protocols

In this section, we suppose that Alice and Bob met in the past and took the opportunity to prepare together some entangled states that they since have been sharing in the form of Bell states[2]. They can now use this resource to perform some entanglement assisted communication. We will give two examples of such protocols. In the dense coding protocol, Alice will be able to communicate two bits of information to Bob by transmitting a single qubit. In the quantum teleportation protocol, Alice will transfer the state of one qubit to Bob through the communication of two classical bits.

3.7.1 Dense coding

Dense coding relies on the fact that unitary evolution on only one half of a Bell state can prepare any other Bell state. If we suppose that Alice and Bob share $|\Phi_+\rangle$, it is easy to check that Alice can prepare any other Bell state using

$$\begin{aligned} |\Phi_+\rangle &\xrightarrow{\mathbb{1}} |\Phi_+\rangle & |\Phi_+\rangle &\xrightarrow{X} |\Psi_+\rangle \\ |\Phi_+\rangle &\xrightarrow{Z} |\Phi_-\rangle & |\Phi_+\rangle &\xrightarrow{ZX} |\Psi_-\rangle\ . \end{aligned}$$

[2]The four Bell states are defined as (see chapter 2):

$$\begin{aligned} |\Phi_+\rangle &= (|00\rangle + |11\rangle)/\sqrt{2} & |\Psi_+\rangle &= (|01\rangle + |10\rangle)/\sqrt{2} \\ |\Phi_-\rangle &= (|00\rangle - |11\rangle)/\sqrt{2} & |\Psi_-\rangle &= (|01\rangle - |10\rangle)/\sqrt{2}\,. \end{aligned}$$

where X, Y and Z are the three Pauli matrices acting on Alice's qubit. After applying one of the four transformations, she sends her qubit to Bob, who projects the total state in the Bell state basis. Depending on Alice's transformation, he obtains one of the four outcome with certainty. Alice has thus transferred two bits of classical information to Bob via the transfer of a single qubit and the consumption of one ebit.

3.7.2 Quantum teleportation

Teleportation is the dual process, where one qubit is transferred from Alice to Bob by using two classical bits and one ebit. We suppose that Alice wants to teleport $|\phi\rangle = \alpha|0\rangle + \beta|1\rangle$ and that Alice and Bob share the Bell state $|\Phi_+\rangle$. The total state can we written

$$
\begin{aligned}
&|\phi\rangle \otimes |\Phi_+\rangle = (\alpha|0\rangle + \beta|1\rangle) \otimes \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \\
&= \frac{1}{\sqrt{2}}(\alpha|000\rangle + \beta|100\rangle + \alpha|011\rangle + \beta|111\rangle) \\
&= \frac{1}{2}\left[\alpha(|\Phi_+\rangle + |\Phi_-\rangle)|0\rangle + \beta(|\Psi_+\rangle - |\Psi_-\rangle)|0\rangle + \alpha(|\Psi_+\rangle + |\Psi_-\rangle)|1\rangle + \beta(|\Phi_+\rangle - |\Phi_-\rangle)|1\rangle\right] \\
&= \frac{1}{2}\left[|\Phi_+\rangle(\alpha|0\rangle + \beta|1\rangle) + |\Phi_-\rangle(\alpha|0\rangle - \beta|1\rangle) + |\Psi_+\rangle(\alpha|1\rangle + \beta|0\rangle) + |\Psi_-\rangle(\alpha|1\rangle + \beta|0\rangle)\right] .
\end{aligned}
$$

When Alice projects her two qubits in the Bell state basis, Bob's qubit is projected into one of the four states shown in the last equation. Alice sends the result of her measurement using two classical bits of information to Bob, who uses this information to apply a unitary transform on his qubit and to retrieve the original state prepared by Alice.

3.8 Summary

- The Shannon entropy $H(X) = -\sum_x p(x)\log p(x)$ quantifies our ignorance about a message. A message of n letters can be optimally encoded with $nH(X)$ bits. There exists codes with this compression rate such that the decoding error will tend to zero when n tends to infinity.

- The von Neumann entropy $S(\rho) = -\operatorname{Tr}[\rho\log\rho]$ does the same for a quantum state. A message of n quantum states can be optimally encoded with $nS(\rho)$ qubits. There exists compression protocols with this compression rate such that the fidelity of the decoded state with the initial state will tend to one when n tends to infinity.

- A classical communication channel can be described by the conditional probability distribution $p(y|x)$ to obtain the output y given the input x. The mutual information $I(X:Y) = H(Y) - H(Y|X) = H(X) - H(X|Y)$ gives the transmission rate of the communication, that is the number of bit of information per transmitted letter. The capacity of the channel is the optimal rate when optimizing over the input probability distribution $p(x)$.

- If a quantum communication channel is used to output messages that are tensor products of n separable letters, an optimal communication protocol will reach a classical information rate given by the Holevo-χ or Holevo bound. This allows one to define the capacity of the channel as the optimum of χ over the possible input states.

In terms of applications, we have arrived at the conclusion that quantum communication cannot increase the rate of classical communication, which may come as a surprise. Indeed, a naive argument is to say that the Hilbert space is huge and thus could store a lot of information. For example, in order to represent a pure quantum state of n qubits in a classical computer, one needs a vector of 2^n complex elements. Assuming that each complex number is coded with m bits, the vector occupies $m2^n$ bits of memory, which is exponentially large with the number of qubits. The classical description of an arbitrary quantum state represents a very large amount of information. But, this does not mean that a n qubit register may be used to store and retrieve this amount of information. As we have argued through this chapter, a more correct reasoning is the following. When elaborating a code to store messages in the register, Alice decides to attribute the first message to a state $|\phi_1\rangle$, then the second message to a state $|\phi_2\rangle$, which she chooses to be orthogonal to $|\phi_1\rangle$ in order to be sure to be able to distinguish it from $|\phi_1\rangle$, and so on. She ends up with a set of 2^n orthogonal states, each coding for one message. Adding more messages results in states which are not orthogonal. The probability of decoding error increases and reduces the mutual information more than the gain due to the increase of the number of messages. As a result, the n qubit register optimally encodes the same number of messages as a classical n bit register.

Of course, one of the great advantage of quantum communication, which we did not consider at all in this chapter, is that the privacy of the communication can be guaranteed by the laws of quantum mechanics. Quantum cryptography would deserve a chapter on its own, and we apologize for blindly ignoring the topic.

References

[1]M. Tribus and E. McIrvine, "Energy and information", Scientific American **225**, 179–188 (1971).

[2]B. Schumacher, "Quantum coding", Physical Review A **51**, 2738–2747 (1995).

[3]J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, 2018).

[4]A. S. Holevo, "Bounds for the Quantity of Information Transmitted by a Quantum Communication Channel", Probl. Peredachi Inf., 10.18287/0134-2452-2015-39-4-459-461. (1973).

[5]P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters, "Classical information capacity of a quantum channel", Physical Review A **54**, Publisher: American Physical Society, 1869–1876 (1996).

[6]B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels", Physical Review A **56**, Publisher: American Physical Society, 131–138 (1997).

[7]A. Holevo, "The capacity of the quantum channel with general signal states", IEEE Transactions on Information Theory **44**, Conference Name: IEEE Transactions on Information Theory, 269–273 (1998).

[8]A. Peres and W. K. Wootters, "Optimal detection of quantum information", Physical Review Letters **66**, 1119–1122 (1991).

[9]V. Giovannetti, S. Lloyd, and L. Maccone, "Achieving the Holevo bound via sequential measurements", Physical Review A **85**, Publisher: American Physical Society, 012302 (2012).

[10]P. W. Shor, "Capacities of quantum channels and how to find them", Mathematical Programming **97**, 311–335 (2003).

[11]B. Schumacher and M. D. Westmoreland, "Relative entropy in quantum information theory", 10.1090/conm/305/05225 (2000).

[12]T. Sagawa, "Second Law-Like Inequalities with Quantum Relative Entropy: An Introduction", 10.1142/9789814425193_0003 (2012).