Quantum Information and Computation

draft of notes 2013

Nicolas Macris

Contents

Preface

Draft of course notes

Part I

Quantum Mechanics and Quantum Bits

1 Experiments with light

At the beginning of the 20th century a major change of paradigm occured in the laws of physics. This revolution was triggered by a host of experimental discoveries which lead to a major revision of our concepts of particle, wave and measurements of observables such as for example position, velocity, magnetic moment. The quantum theory that emerged is today the best tested theory of physical phenomena. The classical laws of physics are seen as a limiting case of quantum laws, that are valid when quantum effects can be neglected. This is the case for a wide range of phenomena which roughly speaking are macroscopic phenomena for which Newton's (or perhaps relativistic) laws of motion and Maxwell equations are adequate. Quantum effects cannot be neglected when we want to describe microscopic phenomena. But note that macroscopic quantum phenomena also exist and the borderline between classical and quantum behaviors is a deep, subtle and not totally solved problem. In any case, quantum theory explains the chemical bond, is thus at the basis of chemistry, it explains the stucture of the atom and the periodic table of elements, and is the basis for nuclear, particle and high energy physics. Quantum mechanics is also necessary to explain many properties of condensed matter for example metals, semi-conductors, magnets, superconductors, superfluids. Quantum mechanics is necessary to explain the interaction of matter and light.

Quantum mechanics was largely discovered by studying the interaction of matter with light. The early experiments of the 20th century, and some of the late 19th century, forced physicist to revise completely their views on the intimate nature of light and matter. It was gradually realized that light has both particle-like and wave-like behaviors. Similarly particles (e.g. the electron) have both particle-like and wave-like behaviors. Today we view these constituents of matter as entities called "quantum fields". Wave and particle behaviors are manifestations of the quantum fields.

The laws of physics are expressed in mathematical language. It is thus not so surprising that these conceptual revolutions were couched in a mathematical formalism that departs quite radically from the one of classical physics. For example Heisenberg was bold enough to represent observable and measurable quantities, such as position and velocity of an electron orbiting an atom, by "matrices" or "linear operators". From our modern perpective it is hard to appreciate why this was so bold. Let us just point out here that matrices were not part of the

curriculum of physics students in the 1920's and that Heisenberg constructed the rule of matrix multiplication and other linear algebra facts by means of guessing at the laws of physics. He was guided by experiment, was a genius, and guessed right! The mathematical formalism of quantum mechanics has posed new interesting problems in functional analysis, geometry, group theory (today quantum information theory also offers new mathematical challenges).

The development of quantum mechanics in its modern form spans a period of at least 25-30 years between 1900 and 1930's. It is the achievement of many experimental and theoretical physicists. This was a golden age of discovery in physics full of surprising developments. It will not be possible to go through and understand the historical development of quantum mechanics in this course. Starting with chapter 2, the modern formalism of quantum mechanics is presented. This mathematical formalisation of the physical laws discovered by the founding fathers, was first clearly spelled out by Dirac and von Neumann around 1930-1932 in two influencial books, and has remained for the main part unchanged since then.

Before proceeding directly to the mathematical formalism it is nevertheless good to motivate it thanks to simple experiments that can be performed with light. The experiments are presented here as "thought experiments", but they can be performed in a real lab. We will gradually introduce some of the basic ideas of quantum mechanics through the discussion of these experiments. This is the goal of this chapter.

1.1 Electromagnetic waves

According to Maxwell (1862) and Hertz (1886), light is an electromagnetic wave of electric $\mathbf{E}(\mathbf{x},t)$ and magnetic $\mathbf{B}(\mathbf{x},t)$ fields freely oscillating in vacuum. The solutions of Maxwell equations in empty space are superpositions of monochromatic modes of frequency ω. A mode, or plane wave, propagating along the z axis, is given by

$$\mathbf{E}(\mathbf{x},t) = Re\,\mathbf{E}_0\, e^{i(kz-\omega t)}, \qquad \mathbf{B}(\mathbf{x},t) = \frac{1}{c}\hat{\mathbf{z}}\times\mathbf{E}(\mathbf{x},t), \qquad \omega = ck \tag{1.1}$$

The amplitude vector $\mathbf{E}_0$ (thus $\mathbf{E}$ and $\mathbf{B}$ also) always belongs to the $(x,y)\perp z$ plane,

$$\mathbf{E}_0 = E_0 \begin{bmatrix} \cos\theta e^{i\delta_x} \\ \sin\theta e^{i\delta_y} \\ 0 \end{bmatrix} \tag{1.2}$$

The energy per unit time per unit surface that would be imparted to a material object by the wave, is given by the norm of the Poynting vector

$$\mathbf{S} = \epsilon_0 c^2 \mathbf{E}\times\mathbf{B} \tag{1.3}$$

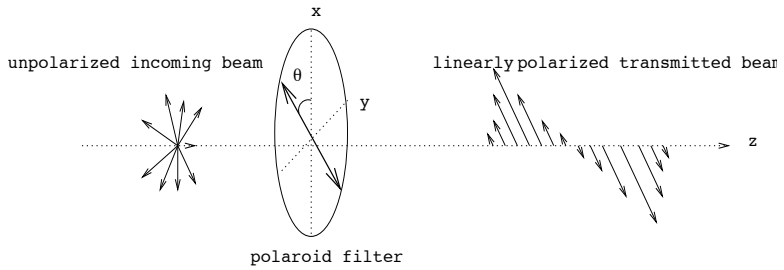


Figure 1.1 Preparation of beam polarized along θ

A convenient measure of the intensity I of the wave is given by the average of its norm, over a period $T = \frac{2\pi}{\omega}$,

$$I = \frac{1}{2}\epsilon_0 c|\mathbf{E}_0|^2 = \frac{1}{2}\epsilon_0 c E_0^2 \tag{1.4}$$

From (1.1), (1.2) it follows that the tip of the electric (and hence also magnetic) field vector describes, as a function of time, an ellipse in the (x, y) plane. There are two degenerate cases of special importance. *Linear polarization* corresponds to $\delta_x - \delta_y = m\pi$ (m integer) and the tip of the field oscillates in the (x, y) plane on a line making the angle θ with x (m even/odd). For $\theta = \frac{\pi}{4}$ (so that $\cos\theta = \sin\theta = \frac{1}{\sqrt{2}}$) and $\delta_x - \delta_y = m\frac{\pi}{2}$ (m odd integer) the polarization is left/right *circular* which means that the tip of the field rotates along a circle of radius E_0.

A light beam can be easily prepared in a state of linear polarization with the help of a filter which transmits only the component of the electric field along θ. All our subsequent discussion does not rely on a detailed explanation of the phenomenon and we do not need to know more about it[1]. Such a device is called a *polarizer* with axis θ (figure 1.1).

Analyzer-detector apparatus. Assume that a source of light has been prepared in a state of linear polarization along θ as in figure 1.1.

$$\mathbf{E}_{\text{in}}(\mathbf{x}, t) = E_0 \begin{bmatrix} \cos\theta \\ \sin\theta \\ 0 \end{bmatrix} Re\, e^{i(kz-\omega t)} \tag{1.5}$$

The intensity of the prepared beam (1.5) is proportional to E_0^2. Suppose now that this ray is transmitted through a second polarizer at an angle α. This second polarizer is called the *analyzer*. The light is then collected by a detector[2] and its

[1] In fact so-called absorptive polarizers are made of sheets of anisotropic crystals allowing electron motion preferentially in the $\theta_\perp$ direction. The $\theta_\perp$ component of the electric field sets electrons into a state of oscillation which produces the emission of an emitted anti-phase electromagnetic wave polarized along $\theta_\perp$. The later cancels the progressive $\theta_\perp$ component of the wave so that the net effect is to leave out a θ transmitted component and a $\theta_\perp$ reflected component.

[2] This can be a photoelectric cell which transforms the electromagnetic energy into a current.

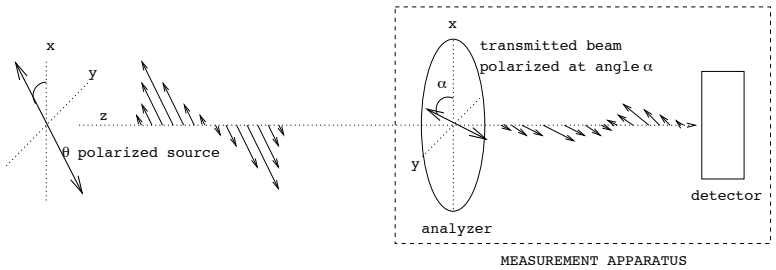


Figure 1.2 analyzer-detector measurement apparatus

intensity measured (see figure 1.2). The electric field of the final beam is obtained by projecting the incoming electric field on the analyzer axis $\mathbf{e}_\alpha$

$$\mathbf{E}_{\text{out}} = (\mathbf{E}_{\text{in}} \cdot \mathbf{e}_\alpha)\, \mathbf{e}_\alpha = E_0 \cos(\theta - \alpha) \begin{bmatrix} \cos\alpha \\ \sin\alpha \\ 0 \end{bmatrix} Re\, e^{i(kz-\omega t)} \tag{1.6}$$

and the intensity received in D is proportional to $E_0^2 \cos^2(\theta - \alpha)$. To summarize, when a beam polarized along θ is transmitted through an analyzer at an angle α, the outgoing beam is polarized along α and the fraction of intensity collected by the detector (average power per unit surface) is[3]

$$\frac{I_{\text{out}}}{I_{\text{in}}} = \cos^2(\theta - \alpha) \tag{1.7}$$

In particular if $\alpha - \theta = 0, \pi$ all the light passes through the analyzer, while if $\alpha - \theta = \pm\frac{\pi}{2}$ none of it is transmitted. The analyzer-detector system can be used as a measurement apparatus to determine the polarization of a wave (assuming we know a priori that it is linear) by adjusting the angle α such that the collected intensity varies from 0 to its maximal value. Let us now describe two simple experiments with electromagnetic waves.

Polarizing beam-splitter experiment. There exist prisms[4] that have the property of splitting a beam in two linearly polarized ones, one is polarized perpendicular to the incidence plane while the other is polarized parallel to that plane. In figure 1.3 the incidence plane is (x, z) so one ray has y polarization while the other one has x polarization. Two detectors D_x and D_y measure the outgoing intensities of each beam. Note that the polarization degree of freedom is coupled to the orbital (path of ray) degree of freedom. Before the polarizing beam-splitter the electric filed is given by (1.5) and has intensity proportional

[3] Malus law.

[4] These are made of quartz or calcite crystals whose refraction index are different for polarization perpendicular to, versus into, the incidence plane. Such crystals are called birefringent, one ray is called ordinary because the direction of refraction obeys the usual Snell law, while the other ray is called extraordinary.

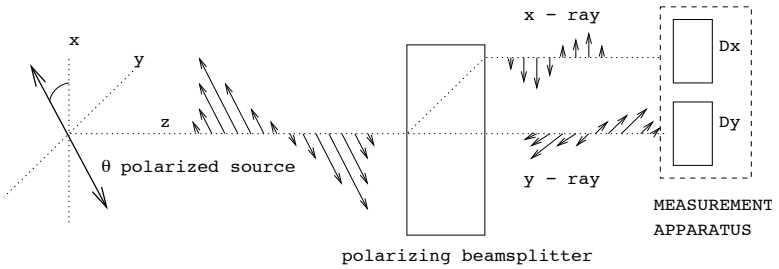


Figure 1.3 polarizing beam-splitter experiment

to E_0^2. After the beam-splitter the x-polarized ray has an electric field

$$\mathbf{E}_x = E_0 \begin{bmatrix} \cos\theta \\ 0 \\ 0 \end{bmatrix} Re\, e^{i(kz-\omega t)} \tag{1.8}$$

and the intensity detected at D_x is proportional to $E_0^2 \cos^2\theta$, while the y-polarized ray has a field

$$\mathbf{E}_y = E_0 \begin{bmatrix} 0 \\ \sin\theta \\ 0 \end{bmatrix} Re\, e^{i(kz-\omega t)} \tag{1.9}$$

and its intensity measured by D_y is proportional to $E_0^2 \sin^2\theta$. *Both detectors collect a fraction of the intensity*,

$$\frac{I_{\text{out},x}}{I_{\text{in}}} = \cos^2\theta, \qquad \frac{I_{\text{out},y}}{I_{\text{in}}} = \sin^2\theta \tag{1.10}$$

In this experiment absorption and reflection by the prism are negligible so that the sum of the these two fractions equals 1.

Decomposition-recombination experiment. Once we have decomposed light with a polarizing beam-splitter, we can recombine it with a symmetric prism. We analyze the recombined beam with an analyzer-detector apparatus (see figure 1.4). Let us carefully review the situation. Before the first beam-splitter we have *one ray* with electric field given by (1.5). The first beam-splitter splits the ray in *two parts* with electric fields given by (1.8) and (1.9). After the second beam-splitter the two rays *interfere* and the electric field of the recombined beam is the sum of (1.8) and (1.9), which equals (1.5). *The fraction of intensity collected by the analyzer-detector system is*

$$\cos^2(\theta - \alpha) \tag{1.11}$$

a fact consistent with the first experiment.

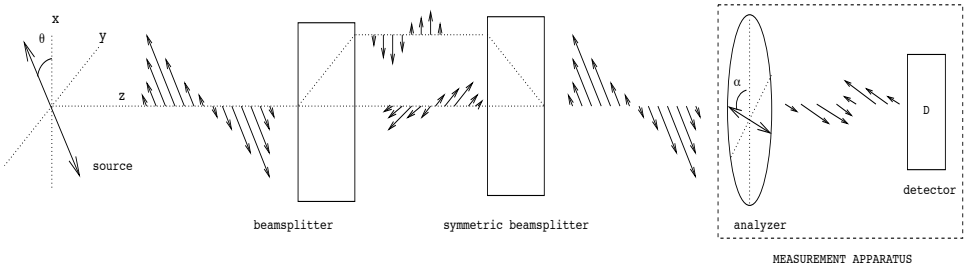


Figure 1.4 decomposition-recombination experiment

1.2 Photons

The works of Planck (1900) on the spectrum of black-body radiation, of Einstein (1905) on the photoelectric effect and Bohr (1913) on the atomic structure (and spectral lines), taught us that the interaction of light with matter occurs through discrete quanta (quantities) of energy and momentum that are absorbed and emitted. These quanta are called photons, and each photon carries an energy $\hbar\omega$ and momentum $\hbar k$ (where $\omega = ck$ still holds). If we think of the beam as a collection of independent photons, its intensity is $\hbar\omega c\frac{N}{V}$ where $\frac{N}{V}$ is the number of photons per unit volume[5]. Identifying this quantity with (1.4) we find a relation between the electric field and the number of photons associated to the electromagnetic wave.

If we diminish sufficiently the intensity of the source we arrive at a situation where in principle photons are emitted one by one. We will repeat the experiments with such a *single photon source*, that prepares them in a state of polarization θ.

Analyzer-detector apparatus. Let us first discuss how the analyzer-detector measurement apparatus works. We repeat the experiment of figure 1.2 and collect photons at the detector D. When a photon hits the detector the later clicks (an electric pulse is triggered) - we record this event as a 1, otherwise we record 0. This experiment produces a sequence

$$100111100010101001110 1... \tag{1.12}$$

that looks random and where the empirical fraction of 1's is $\cos^2(\theta - \alpha)$. From this experiment we infer

$$\text{probability of detecting a photon} = \cos^2(\theta - \alpha) \tag{1.13}$$

In particular if $\alpha - \theta = 0, \pi$ all photons are detected while if $\alpha - \theta = \pm\frac{\pi}{2}$ no photon is detected.

This experiment suggests that *photons behave as particles which carry a polarization degree of freedom.* Indeed if they would behave as waves, then a part

[5] $c\frac{N}{V}$ is the number of photons per unit time per unit surface that hit a detector.

of the wave would be transmitted through the analyzer and some energy would always be measured in the detector. However *the event is discrete, the detector clicks or does not click.* Moreover *it seems impossible to predict the precise polarization outcome for each individual photon: clicks are random.* Note that the statistics of the outcomes seems to satisfy a definite formula (1.13); and this formula is the one found in the theory of electromagnetic waves (!).

The randomness of the outcome is a fundamental feature of the measurement process for quantum systems and that it is not at all obvious to reconcile this fact with our classical intuitions. One could attempt a classical interpretation[6] by saying that the photon is a particle-like object that undergoes complicated but otherwise deterministic collision processes within the analyzer, which result in a probability $\cos^2(\theta - \alpha)$ of being transmitted. Such attempts do not resist the tests of other experiments.

Let us now repeat the two previous experiments with photons that are sent one by one.

Polarizing beam-splitter experiment. Each single photon (polarized at an angle θ) goes through the prism. We observe that either D_x clicks (the upper detector register a 1 and the lower a 0) or D_y clicks (the upper detector registers a 0 and the upper a 1); but they never click simultaneously. We record two random complementary sequences with respective fractions of 1 equal to $\cos^2\theta$ and $\sin^2\theta$. Empirically,

$$\text{prob detect photon at } \mathrm{D_y} = \sin^2\theta, \qquad \text{prob detect photon at } \mathrm{D_x} = \cos^2\theta \quad (1.14)$$

The sum is equal to one which means that the photon has certainly passed through the beam-splitter.

The fact that the detectors never click simultaneously suggest as above that the *photons behave as particles.* Indeed, would they behave as waves, both detectors would collect some energy at the same time.

One may attempt the same (wrong) classical interpretation as above. A photon is a *particle*, which due to complicated but otherwise deterministic collisions with the crystal, is deflected towards the lower path with probability $\sin^2\theta$ *or* through the upper path with probability $\cos^2\theta$. This turns out to be incompatible with the next experiment.

Decomposition-recombination experiment. let us consider again the setting of figure 1.4. When photons are sent one by one we again record a sequence of random clicks, and we infer from this sequence

$$\text{prob detect photon at D} = \cos^2(\theta - \alpha) \quad (1.15)$$

This should comes as a great surprise to the reader. Indeed this result *is not consistent with the particle-like picture of a photon, but rather with a wave-like picture*, as we now show.

[6] in the spirit of statistical mechanics, say

Theoretical prediction of the particle picture. If a photon takes the lower path in figure 1.4 its polarization is horizontal before the second beam-splitter and comes out of it in a horizontal state. Therefore the probability of transmission of such a lower-path photon through the analyzer is $\cos^2(\frac{\pi}{2} - \alpha) = \sin^2 \alpha$. Therefore

$$\text{prob(D clicks} \mid \text{lower path)} = \sin^2 \alpha \tag{1.16}$$

If the photon takes the upper path its polarization is vertical just before the second beam-splitter and comes out in a state of vertical polarization. Therefore the probability of transmission of such an upper-path photon is $\cos^2(0 - \alpha)$ and

$$\text{prob(D clicks} \mid \text{upper path)} = \cos^2 \alpha \tag{1.17}$$

Now, we have

$$\begin{aligned} \text{prob(D clicks)} =& \text{prob(D clicks} \mid \text{lower path)prob(lower path)} \\ &+ \text{prob(D clicks} \mid \text{upper path)prob(upper path)} \end{aligned} \tag{1.18}$$

Thus because of (1.14), (1.16), (1.17)

$$\text{prob detect photon at D} = \sin^2 \theta \sin^2 \alpha + \cos^2 \theta \cos^2 \alpha \tag{1.19}$$

This contradicts the experimental result (1.15) *and is therefore plain wrong !*

The term that is missing is precisely

$$2 \cos\theta \cos\alpha \sin\theta \sin\alpha \tag{1.20}$$

which, in wave theory, appears because of the *interference* between the x and y components of the electric field. This suggests that a single photon follows both paths, just as a wave would do, and interferes with itself just as a wave would do

Let us summarize. We face the following situation: the decomposition experiment suggests that photons behaves in a *particle-like* manner, while he recombination experiment (1.4) suggests that photons behave in a *wave-like* fashion. As for most dilemmas, the resolution offered by *quantum theory teaches us that both pictures are two faces of a more subtle reality that goes beyond this dichotomy.* One sometimes refers to this dual behavior of light, and all known forms of matter, as the "particle-wave duality" or the "complementarity principle".

1.3 The quantum setting: first encounter

In fact all known forms of matter[7] display this particle/wave duality. As we will now see quantum mechanics offers us a picture which accommodates both behaviors and superseedes the classical pictures of wave and particle[8].

[7] For example photons, electrons, nuclei and their constituents ...

[8] According to modern physics, matter is described by relativistic quantum fields. There are underlying quantum fields (e.g. the quantum electromagnetic field, the quantum electronic

We will illustrate how the rules of quantum theory consistently explain the three experiments. The situation will be modeled in the simplest possible way which retains the basic essence of quantum mechanics.

The state of a photon is described by two degrees of freedom, an *orbital* degree of freedom and a *polarization* degree of freedom. Let us first concentrate on polarization. The state of polarization is described by a unit vector $\mathbf{e}$ perpendicular to the direction of motion. Following Dirac we call these state vectors *kets* and denote them as $|\mathbf{e}\rangle$. Since the polarization vector lies in the x, y plane it can be described in a orthonormal basis $|\updownarrow\rangle$, $|\leftrightarrow\rangle$, corresponding to the two linear states of polarization along x and y

$$|\mathbf{e}\rangle = \lambda|\updownarrow\rangle + \mu|\leftrightarrow\rangle, \qquad |\lambda|^{\mathbf{2}} + |\mu|^{\mathbf{2}} = \mathbf{1} \tag{1.21}$$

Here λ and μ are complex numbers. thus a general polarization state is a normalized two component vector belonging to $\mathbf{C}^2$. The space $\mathbf{C}^2$ is our first and simplest example of a space of quantum states.

A state of linear polarization along θ corresponds to $\lambda = \cos\theta$ and $\mu = \sin\theta$, so that (1.21) becomes

$$|\theta\rangle = \cos\theta|\updownarrow\rangle + \sin\theta|\leftrightarrow\rangle \tag{1.22}$$

On the other hand for circular polarization the x and y components of the polarization vector have a $\frac{\pi}{2}$- phase difference. Two basis states with circular polarization are,

$$|L/R\rangle = \frac{1}{\sqrt{2}}(|\updownarrow\rangle \pm i|\leftrightarrow\rangle) \tag{1.23}$$

Given a state vector $|\Phi\rangle$ its *adjoint* (also called hermitian conjugate) is obtained by taking the complex conjugate and transposing $\overline{|\Phi\rangle}^T$. This is denoted as a *bra*

$$\langle\Phi| = \overline{|\Phi\rangle}^T \tag{1.24}$$

The usual inner product (defined over a complex vector space) is called the *bracket*

$$\langle\Psi|\Phi\rangle = (\overline{|\Psi\rangle}^T) \cdot (|\Phi\rangle) \tag{1.25}$$

As an example consider the inner product between two polarization state vectors. First the conjugate of a linearly polarized state is

$$\langle\alpha| = \langle\updownarrow| \cos\alpha + \langle\leftrightarrow| \sin\alpha \tag{1.26}$$

The inner product with $|\theta\rangle$ then is

$$\begin{aligned}\langle\alpha \mid \theta\rangle &= (\langle\updownarrow| \cos\alpha + \langle\leftrightarrow| \sin\alpha) \cdot (\cos\theta|\updownarrow\rangle + \sin\theta|\leftrightarrow\rangle) \\ &= \cos\alpha\cos\theta + \sin\alpha\sin\theta \\ &= \cos(\theta - \alpha)\end{aligned} \tag{1.27}$$

field, the quark field etc...) which may manifest themselves in a wave-like or particle-like fashion depending on the situation. We will not introduce field theoretical notions in this course.

To obtain the second equality one expands the braces into four terms, uses linearity of the bracket and the orthonormality condition,

$$\langle p \mid p' \rangle = \delta_{pp'} \tag{1.28}$$

This trivial calculation has been done in the linear polarization orthonormal basis $\{|\leftrightarrow\rangle, |\updownarrow\rangle\}$. It is instructive to check that the circularly polarized states $\{|L\rangle, |R\rangle\}$ form another orthonormal basis of the two dimensional complex vector space.

Let us now introduce the orbital degree of freedom in the picture. For a freely moving photon, i.e a photon that does not interact with a material object, the orbital state is entirely described once we know its *momentum* $\mathbf{k}$, which has a *direction* $\mathbf{k}/k$ and a *norm* $k = \frac{\omega}{c}$. The state vector is now denoted as $|\mathbf{k}, \mathbf{e}\rangle$. This state freely evolves with time and for a photon of frequency ω the time evolution simply amounts to a multiplicative phase factor, which does not change the momentum and the polarization. The photon state at time t is

$$|\Psi_{\mathbf{k},\mathbf{e}}(t)\rangle = e^{-i\omega t}|\mathbf{k}, \mathbf{e}\rangle \tag{1.29}$$

An explanation is in order here about the kets indexed by two degrees of freedom. We will see in the next chapter that the mathematical rule to combine degrees of freedom is the tensor product; this means that $|\mathbf{k}, \mathbf{e}\rangle = |\mathbf{k}\rangle \otimes |\mathbf{e}\rangle$ and that the inner product is

$$\langle \mathbf{k}', \mathbf{e}' \mid \mathbf{k}, \mathbf{e}\rangle = \langle \mathbf{k}' \mid \mathbf{k}\rangle \cdot \langle \mathbf{e}' \mid \mathbf{e}\rangle \tag{1.30}$$

Finally the momentum vectors themselves form an orthonormal basis $\langle \mathbf{k}' \mid \mathbf{k}\rangle = \delta_{\mathbf{k}',\mathbf{k}}$.

As we will see in the next chapter, in general, the time-evolution of isolated systems is given by a unitary transformation. In (1.29) the unitary transformation is simply the multiplication by the phase factor. When the photon interacts with matter (for example with the analyzer, the beam-splitter) one has in principle to describe the unitary evolution of the total system (photon + analyzer or photon + beam-splitter), which is then more complicated. Here we do not have to discuss such issues as we consider only the in-going and out-going states which are those of freely moving photons.

When we make a measurement on a system, the system that is observed cannot be considered as isolated and the state is modified in a non-unitary way. Explaining the measurement process is a subject that has been (and sometimes is still) much debated since the early days of quantum mechanics. An operational rule, to determine the outcome of a measurement is given by the so-called *measurement postulate* (Born, Heisenberg, Bohr 1924-1927) in the form advocated

Figure 1.5 measurement with initial state $|\Psi\rangle$ and outcome $|\Phi\rangle$.

by what has been named the Copenhagen School[9] (figure 1.5). Here we give it in a rough form, and will be more precise in the next chapter.

If a system is initially prepared in the state $|\Psi\rangle$ and the outcome of the measurement is a state $|\Phi\rangle$, the probability of the transition $|\Psi\rangle \to |\Phi\rangle$ is

$$\text{Prob}(|\Psi\rangle \to |\Phi\rangle) = |\langle\Phi \mid \Psi\rangle|^2 \tag{1.31}$$

One cannot predict the outcome of the transition but only its frequency of occurrence during repeated identical experiments with identical initial states.

The transition between the initial and final state is also called "reduction" or "collapse" of the state. In a more precise formulation of the measurement postulate, in the next chapter, we will see that the transition probabilities of all possible outcomes sum to one.

The re-interpretation of the experiments in the next section should make this rather abstract postulate a bit more "natural".

1.4 Quantum interpretation of experiments.

Analyzer-detector apparatus. We assume that the source prepares single photons in the linearly polarized, freely moving state

$$|\Psi_{\mathbf{k},\theta}(t)\rangle = e^{-i\omega t}|\mathbf{k},\theta\rangle \tag{1.32}$$

If the measurement apparatus is the analyzer-detector system of figure 1.2, the measurement postulate tells us that the probability to find the photon in state $|\mathbf{k},\alpha\rangle$ is

$$|\langle\mathbf{k},\alpha \mid \Psi_{\mathbf{k},\theta}(t)\rangle|^2 = |\langle\alpha \mid \theta\rangle|^2 = \cos^2(\theta-\alpha) \tag{1.33}$$

This is consistent with the experimentally measured frequency of clicks in D.

Polarizing beam-splitter experiment. Before the beam-splitter the photon

[9] Einstein never agreed that this rule is the final story. In his words "I, at any rate, am convinced that He (God) does not throw dice". Bohr replied "Einstein, don't tell God what to do". In any case, this rule has not been challenged by experiment so far, and there is hardly any more satisfying theoretical framework to date. In this course we stick to this rule !

state is (1.32), which is equal to

$$e^{-i\omega t}(\cos\theta|\mathbf{k},\updownarrow\rangle + \sin\theta|\mathbf{k},\leftrightarrow\rangle) \tag{1.34}$$

After the beam-splitter it becomes

$$e^{-i\omega t}(\cos\theta|\mathbf{k}_u,\updownarrow\rangle + \sin\theta|\mathbf{k}_l,\leftrightarrow\rangle) \tag{1.35}$$

where $\mathbf{k}_u$ and $\mathbf{k}_l$ label the upper and lower paths[10]. Notice that contrary to (1.34), in (1.35) we cannot separate the orbital and polarization degrees of freedom into a tensor product: it can be shown that for (1.35) this is an intrinsic property that does not depend on the basis. We say that the orbital and polarization degrees of freedom have been *entangled* by the beam-splitter. Entangled states depart fundamentally from the classical picture and retain quantum correlations that are missing in the classical interpretation. As we will see in this course they play a very important role in quantum information and computation because they may offer resources that are non-classical.

Now we consider the two detectors as our measurement apparatus. The measurement postulate tells us that the probability to observe the photon in state $|\mathbf{k_u},\updownarrow\rangle$ is

$$|\langle\mathbf{k_u},\updownarrow|\ e^{-i\omega t}(\cos\theta|\mathbf{k}_u,\updownarrow\rangle + \sin\theta|\mathbf{k}_l,\leftrightarrow\rangle)|^2 = \cos^2\theta \tag{1.36}$$

Similarly the probability to observe it in the state $|\mathbf{k}_l,\leftrightarrow\rangle$ is

$$|\langle\mathbf{k}_l,\leftrightarrow|\ e^{-i\omega t}(\cos\theta|\mathbf{k}_u,\updownarrow\rangle + \sin\theta|\mathbf{k}_l,\leftrightarrow\rangle)|^2 = \sin^2\theta \tag{1.37}$$

This is consistent with the experimental fractions of clicks at D_x and D_y.

Recombination experiment. The second polarizing beam-splitter transforms the entangled state (1.35) back to (1.32). The later state enters the measurement apparatus constituted by the analyzer-detector system. Therefore the probability of observing $|\mathbf{k},\alpha\rangle$ is simply given by (1.33). This is the experimental frequency of clicks at D ! The quantum interpretation does not loose track of the interference term (1.20).

1.5 Notion of quantum bit

There exist many quantum systems in nature that can be described by state vectors which belong to the vector space $\mathbf{C}^2$, the two dimensional complex vector space. If we call $|0\rangle$ and $|1\rangle$ two orthonormal basis states a general state vector takes the form

$$|\psi\rangle = \lambda|0\rangle + \mu|1\rangle, \qquad |\lambda|^2 + |\mu|^2 = 1 \tag{1.38}$$

[10] Here we may imagine that the paths are not quite in the same direction so that these two labels are different. In principle one should make a more complete description of the orbital part of the state that takes into account the finite width of the beams.

It will often be convenient to identify

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{1.39}$$

and

$$|1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{1.40}$$

and in quantum information theory it is customary to call this canonical basis the *computational basis*. Of course one can represent the quantum bit $|\psi\rangle$ in any other basis, and one that we will often use one that is obtained by a standard 45 degree real rotation

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \qquad |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \tag{1.41}$$

This basis will be called the Hadamard basis. Since the vector space is complex we can make more general unitary transformations. For example

$$|L\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle), \qquad |R\rangle = \frac{1}{\sqrt{2}}(|0\rangle - i|1\rangle) \tag{1.42}$$

We have already seen a physical realization of a quantum bit, namely the photon polarization. If we identify the computational basis with horizontal/vertical polarized photon states, then the Hadamard basis corresponds to polarized states at 45 degree angle, and the last basis obtained by a unitary transformation is physically realized by circularly left/right polarized photons. A physically meaningful parametrization of general polarization state is

$$|\psi\rangle = e^{i\delta_x}\cos\theta|\updownarrow\rangle + e^{i\delta_y}\sin\theta|\leftrightarrow\rangle \tag{1.43}$$

If we rotate our reference frame (around z) by angle β, then the state vector is obtained from the above expression by $\theta \to \theta - \beta$. In particular if the reference frame is rotated by 2π we recover the same state vector. These states form ratgher trivial representations of the group of two-dimensional rotations (about the z-axis say).

Another very common but physically different quantum bit is the *spin* $\frac{1}{2}$. The most famous elementary particle (of obvious importance in our everyday life since it transports electricity, interacts with sunlight ...) that has spin $\frac{1}{2}$ is the electron[11]. There exist also many composite systems, such as nuclei or atoms that carry a total spin of $\frac{1}{2}$. A very rough intuitive way of thinking about spin is to view the particle (the electron say) as having intrinsic spinning motion. If the particle spins about the z axis, its spin is (pointing) $|\uparrow\rangle$ or $|\downarrow\rangle$ according to its direction of rotation. These two states form a basis and the most general spin state is

$$|\psi\rangle = \lambda|\uparrow\rangle + \mu|\downarrow\rangle, \qquad |\lambda|^2 + |\mu|^2 = 1 \tag{1.44}$$

[11] Constituents of nuclei, protons and neutrons also have spin $\frac{1}{2}$. In particular the interaction of the nuclear spins with magnetic fields is at the basis of Nuclear Magnetic Resonance, used for example in medical imaging.

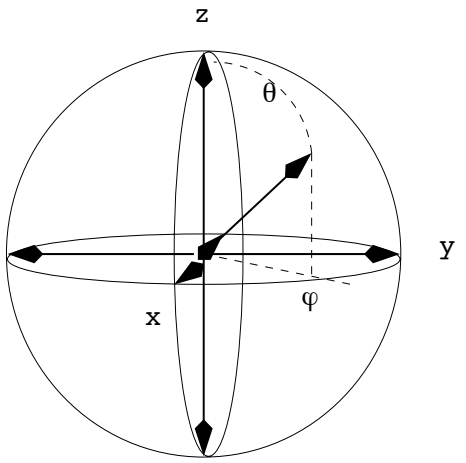


Figure 1.6 Bloch sphere. Computational (z), Hadamard (x), circular (y) basis states

Spin $\frac{1}{2}$ states are two dimensional (complex) representations of the group of rotations in three dimensions. A meaningful parametrization of the states is

$$|\psi\rangle = e^{i\frac{\phi}{2}} \cos\frac{\theta}{2}|\uparrow\rangle + e^{-i\frac{\phi}{2}} \sin\frac{\theta}{2}|\downarrow\rangle \tag{1.45}$$

These states can be represented by the tip of a vector on the *Bloch sphere* (figure 1.6) with the usual spherical coordinates (θ, ϕ). We have the following correspondence (up to phase factors):

$$\theta = 0,\, \pi \qquad |\uparrow\rangle, |\downarrow\rangle, \qquad \text{particle spin along z} \tag{1.46}$$

$$\theta = \frac{\pi}{2}, \phi = 0,\, \pi \qquad |\uparrow\rangle \pm |\downarrow\rangle, \qquad \text{particle spin along x} \tag{1.47}$$

$$\theta = \frac{\pi}{2}, \phi = \pm\frac{\pi}{2} \qquad |\uparrow\rangle \pm i|\downarrow\rangle, \qquad \text{particle spin along y} \tag{1.48}$$

The polarization and spin $\frac{1}{2}$ quantum bits are different representations of the rotation group in quantum mechanics (ultimately coming from the representations of the Lorentz group of relativity).

There exist also other realizations of the quantum bit that have nothing to do with the representations of the rotation group. An example is given by the benzene molecule C_6H_6 that can be in the two states that differ in the arrangement of single and double electronic bonds (figure 1.7). But the molecule can also be found in a *resonating state* such as

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|1\rangle + |2\rangle) \tag{1.49}$$

What is the difference between a classical bit and a quantum bit ? A classical bit is an abstraction of a physical quantity that can be reasonably well described

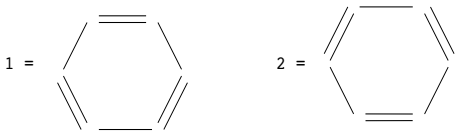


Figure 1.7 possible arrangements of chemical bonds

by a two valued quantity. Examples are the charge in a capacitor, a voltage difference, or the magnetization of a Weiss domain. Classical information theory is sufficiently universal so that it does not have to account for the detailed physical properties of the classical bits. The only underlying assumption is that these exist in two definite values 0 or 1 (let us pretend that noise is absent). Suppose a classical bit is given to you and that you have no information whatsoever about its value. To gain information about its value you can observe it (measure the charge, the voltage difference) and its value is then discovered. By *discovered* we mean that it already had the observed value before the measurement, and that the measurement has not destroyed it.

A quantum bit is also an abstraction of physical as the above examples have shown. It is well described by a two dimensional complex vector. In the same spirit than in the classical case, quantum information theory is sufficiently universal so that many of its aspects are independent of the concrete physical realization. However the important point is that it takes into account the general underlying laws of quantum mechanics. This means in particular that extracting information from quantum bits is quite different than in the classical case. Suppose that a quantum bit is given to you in some state $|\psi\rangle$ on which you do not have any information whatsoever. In order to determine $|\psi\rangle$, we have to observe it (agree ?). To perform a measurement we have to select an apparatus, in other words an orthonormal basis $\{|b_1\rangle, |b_2\rangle\}$. The measurement process then reduces the quantum bit to $|b_1\rangle$ or to $|b_2\rangle$. So we have lost the original state (forever) and have not gained any knowledge (of the initial state) because the final state depends on *our own* choice of basis. Note however that if we are given many copies of $|\psi\rangle$ we can measure all of them in the same basis and get a hold of the probabilities $|\langle b_1 \mid \psi\rangle|^2$, $|\langle b_2 \mid \psi\rangle|^2$.

1.6 A random number generator

At this point the reader may well wonder if quantum laws offer any useful resource in order to process information. In this course we will see that this is so. Here we illustrate this with a very simplified model for a random number generator.

A source sends a beam of photons on a semi-transparent mirror (figure 1.8). The later splits the beam in two parts, the transmitted and reflected beams. If the source is classical we observe that the two detectors each collect a fraction of

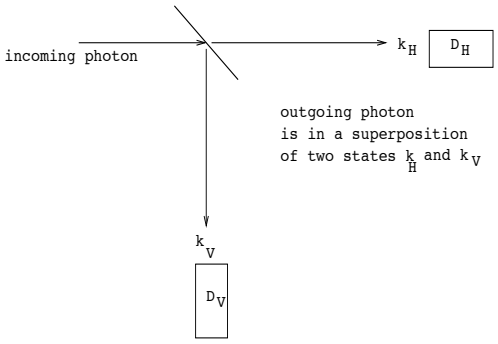


Figure 1.8 semi-transparent mirror

the incoming intensity of the beam. Assuming that the semi-transparent mirror is perfect each detector collects half of the intensity.

When the intensity of our source is lowered sufficiently so that it becomes a single photon source. Photons go through the mirror one at a time, we observe that *either* D_H *or* D_V clicks, never the two at the same time. We obtain a sequence of clicks 0100011101010100011011110 0 that looks Bernoulli with parameter $p = \frac{1}{2}$.

The interpretation of this experimental setup, in the framework of quantum mechanics, is as follows. We drop the polarization index as it plays no role here. A single photon is incoming in the semi-transparent mirror and the state of the photon after the mirror is,

$$e^{i\omega t}\frac{1}{\sqrt{2}}(|\mathbf{k}_H\rangle + |\mathbf{k}_V\rangle) \tag{1.50}$$

This state is a superposition. The outcome of the measurement by the detectors cannot be predicted. The probability that the photon is observed in state $|\mathbf{k}_H\rangle$ is

$$|\langle \mathbf{k}_H \mid e^{i\omega t}\frac{1}{\sqrt{2}}(|\mathbf{k}_H\rangle + |\mathbf{k}_V\rangle)|^2 = \frac{1}{2} \tag{1.51}$$

and similarly the probability that it is observed in state $\mathbf{k}_V$ is

$$|\langle \mathbf{k}_V \mid e^{i\omega t}\frac{1}{\sqrt{2}}(|\mathbf{k}_H\rangle + |\mathbf{k}_V\rangle)|^2 = \frac{1}{2} \tag{1.52}$$

So the measurement process produces a perfectly random sequence.

What do we mean by "perfectly random sequence"? Of course, the sequence is perfectly random only in principle, because in the real experiment there are imperfections, for example, the source is only approximately a single photon source and the semi-transparent mirror has a small bias etc.... But the point here is that, according to the standard interpretation of quantum mechanics, the measurement process produces "true randomness" and not "pseudo-randomness": the clicks are not the result of some underlying deterministic process. This point

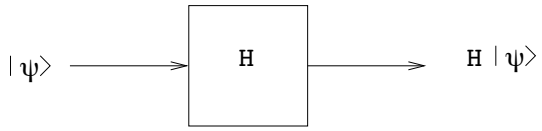


Figure 1.9 Hadamard gate as a model for a semi-transparent mirror

has been much debated by the founding fathers of 20-th century physics and notably by Einstein and Bohr. According to Einstein "God does not play dice", a view that Bohr dismissed. Until today, no other theoretical framework has, successfully described as many phenomena as quantum theory does, and we have so far no experiment that forces us to abandon the standard quantum framework. It is in this sense that we declare the sequence perfectly random.

A slightly more abstract representation in quantum information theory language of this experiment is depicted on figure 1.9. We prepare and measure states in the computational basis $|0\rangle, |1\rangle$. The initial state $|0\rangle$ goes through a Hadamard gate

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{1.53}$$

which produces the state

$$H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \tag{1.54}$$

When we perform a measurement on this state the outcome is $|0\rangle$ with probability

$$|\langle 0|H|0\rangle|^2 = \frac{1}{2} \tag{1.55}$$

or $|1\rangle$ with probability

$$|\langle 1|H|0\rangle|^2 = \frac{1}{2} \tag{1.56}$$

We note that quantum random number generators based on these principles have been realized and are even commercialized. See for example http://www.idquantique.com/true-random-number-generator/products-overview.html

2 Mathematical formalism of quantum mechanics

Quantum mechanics is the best theory that we have to explain the physical phenomena (if we exclude gravity). The elaboration of the theory has been guided by experimental discoveries, as well as thought experiments and conceptual ideas of a great generation of physicist. Milestones of the development of quantum theory are from 1900 to 1930 are: Planck on black body spectrum (1900), Einstein on the photon (1905), Bohr on the atom (1913), De Broglie on the wave function (1924), Schroedinger on the wave function evolution (1926), Born on the interpretation of the wave function (1926), Heisenberg on matrix mechanics (1925), Dirac on relativistic QM (1930). Some never completely accepted their own ideas, although these still form the best theory that we have today. The mathematical form of the theory that we find in textbooks has been put forward by Dirac and von Neumann in the 30's Since then the quantum laws of physics have been used unchanged[1] to successfully describe an impressive range of phenomena ranging from macroscopic solid state, molecular to atomic, nuclear, sub-nuclear and particle physics scales.

The arena of QM is Hilbert space so we begin with some mathematical reminders on linear algebra in such spaces. Our goal is also to carefully introduce the reader to Dirac's bra and ket notation. Then we introduce 5 basic principles that define QM. We also discuss two genuine quantum notions, namely, entangled states and the no-cloning theorem.

2.1 Linear algebra in Dirac notation

A *Hilbert space* $\mathcal{H}$ is a vector space over the field of complex numbers $\mathbf{C}$, with an inner product. For a finite dimensional Hilbert space that is all. For an infinite dimensional Hilbert space we require that it is complete and separable[2]. In quantum information theory we will almost always deal with Hilbert spaces of quantum bits which are discrete by nature, hence our Hilbert spaces are finite dimensional and we do not have to worry about completeness and separability.

The vectors will be denoted $|\psi\rangle$ (pronounced ket psi). The hermitian conjugate

[1] Combined with special relativity when needed

[2] Complete means that all Cauchy sequences converge in the norm induced by the inner product and separable that there is a contable orthonormal basis.

(transpose and complex conjugate) is denoted by $\langle\psi|$ (pronounced bra psi). The inner product is denoted $\langle\phi|\psi\rangle$. This is the inner product of the vectors $|\phi\rangle$ and $|\psi\rangle$ and is called a *bracket* (for bra-ket). The inner product must satisfy:

1. *Positivity*: $\langle\phi|\phi\rangle \geq 0$ with equality if and only if $|\phi\rangle = 0$.
2. *Linearity*: $\langle\phi|(\alpha|\psi_1\rangle + \beta|\psi_2\rangle) = \alpha\langle\phi|\psi_1\rangle + \beta\langle\phi|\psi_2\rangle$, $\alpha, \beta \in \mathbf{C}$
3. *Skew symmetry*: $\langle\phi|\psi\rangle = \overline{\langle\psi|\phi\rangle}$ where the bar denotes complex conjugation.

A ray is an equivalence class of vectors of the form $\lambda|\psi\rangle$ where $\lambda \in \mathbf{C}$ and $|\psi\rangle$ is a specified vector. This specified vector is a representative of the ray.

Example 1: Qbit or two level system. $\mathcal{H} = \mathbf{C}^2 = \left\{ \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \text{ with } \alpha, \beta \in \mathbf{C} \right\}$.

The inner product is $(\overline{\gamma}, \overline{\delta}) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \overline{\gamma}\alpha + \overline{\delta}\beta$. In Dirac notation we have

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \alpha|0\rangle + \beta|1\rangle$$

where $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Moreover

$$(\bar{\gamma}, \bar{\delta}) = \bar{\gamma}\langle 0| + \bar{\delta}\langle 1|$$

and

$$(\bar{\gamma}\langle 0| + \bar{\delta}\langle 1|)(\alpha|0\rangle + \beta|1\rangle) = \overline{\gamma}\alpha\langle 0|0\rangle + \overline{\gamma}\beta\langle 0|1\rangle + \overline{\delta}\alpha\langle 1|0\rangle + \overline{\delta}\beta\langle 1|1\rangle = \overline{\gamma}\alpha + \overline{\delta}\beta$$

Example 2: particle in three dimensional space. $\mathcal{H} = L^2(\mathbf{R}^3) = \{f : \mathbf{R}^3 \to \mathbf{C}, \int d^3x|f(x)|^2 < \infty\}$. The inner product is $\langle f|g\rangle = \int d^3x \overline{f(x)} g(x)$ and the induced norm $||f||_2 = \langle f|f\rangle^{1/2} = \int d^3x|f(x)|^2$. This space plays a fundamental role in quantum mechanics but we will not need it in this course, since we deal only with discrete degrees of freedom.

We will need the notion of *tensor product*. Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two Hilbert spaces with two finite basis. Let the basis of the first space be $|i\rangle_1$, $i = 1, ..., n_1$, $\dim \mathcal{H}_1 = n_1$ and that of the second space $|j\rangle_2$, $j = 1, ..., n_2$, $\dim \mathcal{H}_2 = n_2$. We can form the tensor product space

$$\mathcal{H}_1 \otimes \mathcal{H}_2$$

which is simply the new Hilbert space spanned by the basis vectors

$$|i\rangle_1 \otimes |j\rangle_2$$

(also denoted $|i, j\rangle$ or $|i\rangle_1|j\rangle_2$). There are $n_1 n_2$ such vectors so

$$\dim \mathcal{H}_1 \otimes \mathcal{H}_2 = n_1 n_2$$

A general element of the tensor product space is of the form

$$|\psi\rangle = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} c_{ij}|i,j\rangle = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} c_{ij}|i\rangle_1 \otimes |j\rangle_2$$

Lastly we have to say what is the inner product in the product space:

$$\langle i',j'|i,j\rangle = (\langle i'|_1 \otimes \langle j'|_2)(|i\rangle_1 \otimes |j\rangle_2) = \langle i'|i\rangle_1\langle j'|j\rangle_2$$

Example 3. For one Qbit the Hilbert space is $\mathbf{C}^2$. We will see that the Hilbert space of two Qbits is $\mathbf{C}^2 \otimes \mathbf{C}^2$. The basis vectors of $\mathbf{C}^2 \otimes \mathbf{C}^2$ are $\{|0\rangle \otimes |0\rangle$, $|0\rangle \otimes |1\rangle$, $|1\rangle \otimes |0\rangle$, $|1\rangle \otimes |1\rangle\}$ or $\{|0,0\rangle$, $|0,1\rangle$, $|1,0\rangle$, $|1,1\rangle\}$. A general state is

$$|\psi\rangle = \alpha_{00}|0,0\rangle + \alpha_{01}|0,0\rangle + \alpha_{10}|0,0\rangle + \alpha_{11}|1,1\rangle$$

We have $\dim \mathbf{C}^2 \otimes \mathbf{C}^2 = 4$ and of course $\mathbf{C}^2 \otimes \mathbf{C}^2$ is isomorphic to $\mathbf{C}^4$: However it is important to stress that in QM the meaning of the first representation is really that *two Qbits are involved* : in general it is too difficult (meaningless in some sense) to do physics in a bad representation. Here a few inner products are $\langle 0,0|0,0\rangle = \langle 0|0\rangle\langle 0|0\rangle = 1$, $\langle 0,1|0,1\rangle = \langle 0|0\rangle\langle 1|1\rangle = 1$, $\langle 0,1|1,1\rangle = \langle 0|1\rangle\langle 1|1\rangle = 0$ etc... From these one can compute the inner product of $|\psi\rangle$ and $|\phi\rangle = \beta_{00}|0,0\rangle + \beta_{01}|0,0\rangle + \beta_{10}|0,0\rangle + \beta_{11}|1,1\rangle$. We find the natural product of $\mathbf{C}^4$, $\langle\phi|\psi\rangle = \overline{\beta}_{00}\alpha_{00} + \overline{\beta}_{01}\alpha_{01} + \overline{\beta}_{10}\alpha_{10} + \overline{\beta}_{11}\alpha_{11}$. It is often useful to work in the canonical basis of $\mathbf{C}^4$

$$\begin{pmatrix}1\\0\\0\\0\end{pmatrix} = |0,0\rangle \quad \begin{pmatrix}0\\1\\0\\0\end{pmatrix} = |0,1\rangle \quad \begin{pmatrix}0\\0\\1\\0\end{pmatrix} = |1,0\rangle \quad \begin{pmatrix}0\\0\\0\\1\end{pmatrix} = |1,1\rangle$$

Once this (conventional) correspondence is fixed we can infer the rules for tensoring vectors in their coordinate representation

$$\begin{pmatrix}1\\0\end{pmatrix} \otimes \begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \qquad \begin{pmatrix}1\\0\end{pmatrix} \otimes \begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}0\\1\\0\\0\end{pmatrix}$$

$$\begin{pmatrix}0\\1\end{pmatrix} \otimes \begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}0\\0\\1\\0\end{pmatrix}, \qquad \begin{pmatrix}0\\1\end{pmatrix} \otimes \begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}0\\0\\0\\1\end{pmatrix}$$

You can see that in this course the convention is that you multiply the first set of coordinates by the second vector . All these rules generalize to $\mathbf{C}^2 \otimes \mathbf{C}^2 \otimes \mathbf{C}^2$ etc...

Cauchy-Schwarz inequality. As usual:

$$|\langle\phi|\psi\rangle| \leq \langle\phi|\phi\rangle^{1/2}\langle\psi|\psi\rangle^{1/2}$$

Closure relation. Let $|i\rangle, i = 1, ..., n$ be an orthonormal basis of the n-dimensional Hilbert space. Any vector $|\phi\rangle$ can be expanded as

$$|\phi\rangle = \sum_{i=1}^{n} c_i|i\rangle, \qquad c_i = \langle i|\phi\rangle$$

where the components c_i are obtained by projecting $|\phi\rangle$ over the basis vectors. The above expansion can be rewritten as

$$|\phi\rangle = \sum_{i=1}^{n} |i\rangle\langle i|\phi\rangle$$

Note that $|i\rangle\langle i|$ is the projection operator on vector $|i\rangle$. We can view $\sum_{i=1}^{n} |i\rangle\langle i|$ as the identity operator acting on $|\phi\rangle$, thus we have the *closure relation*

$$\sum_{i=1}^{n} |i\rangle\langle i| = I$$

This turns out to be a very useful identity for doing practical calculations in Dirac notation. Note that this identity is simply the spectral decomposition of the identity.

Observables. In QM observable quantities are represented by linear operators (matrices) on $\mathcal{H}$. Let us briefly review a few important facts. The map $A : \mathcal{H} \to \mathcal{H}$, $|\psi\rangle \to A|\psi\rangle$ is linear if

$$A(\alpha|\phi_1\rangle + \beta|\phi_2\rangle) = \alpha(A|\phi_1\rangle) + \beta(A|\phi_2\rangle)$$

The matrix elements of A in a basis $\{|i\rangle, i = 1, ..., n\}$ of $\mathcal{H}$ are denoted by $\langle i|A|j\rangle$ or A_{ij}. Given A, the *adjoint* of A is denoted $A^\dagger$ and defined by

$$\langle\phi|A^\dagger|\psi\rangle = \overline{\langle\psi|A|\phi\rangle}$$

So the adjoint (or hermitian conjugate) is the operator which has transposed and conjugate matrix elements. We say that A is self-adjoint (or hermitian) if $A = A^\dagger$. The later type of operators play a very central role in QM because observable quantities are represented by self-adjoint operators: the reader can guess that this must be so because any physical measurement is expressed by a real number (why ?) and self-adjoint operators have real eigenvalues. The reader can check that $(A + B)^\dagger = A^\dagger + B^\dagger$ and $(AB)^\dagger = B^\dagger A^\dagger$.

We will also need the following notations for the *commutator*

$$[A, B] = AB - BA$$

and the anticommutator

$$\{A, B\} = AB + BA$$

Projectors in Dirac notation. The linear operator $|i\rangle\langle i| = P_i$ is the projector on the basis vector $|i\rangle$. To check that P_i is a projector we need to verify that $P_i^\dagger = P_i$ and $P_i^2 = P_i$. Here is how one does it in Dirac notation

$$P_i^\dagger = (|i\rangle\langle i|)^\dagger = (\langle i|)^\dagger(|i\rangle)^\dagger = |i\rangle\langle i| = P_i$$

$$P_i^2 = (|i\rangle\langle i|)(|i\rangle\langle i|) = |i\rangle\langle i|i\rangle\langle i| = |i\rangle\langle i| = P_i$$

Since $|i\rangle$ and $|j\rangle$ are orthogonal for $i \neq j$ we have $P_iP_j = P_jP_i = 0$. Indeed

$$P_iP_j(|i\rangle\langle i|)(|j\rangle)(\langle j|) = |i\rangle\langle i|j\rangle\langle j|) = 0$$

$$P_jP_i(|j\rangle\langle j|)(|i\rangle\langle i|) = |j\rangle\langle j|i\rangle\langle i|) = 0$$

Note that if $|\phi\rangle$ is any vector of the Hilbert space, then $P_\phi = |\phi\rangle\langle\phi|$ is the projector on $|\phi\rangle$.

Spectral decomposition. Hermitian operators (matrices) on a Hilbert space have a *spectral decomposition* or *spectral representation*,

$$A = \sum_n a_n P_n$$

where $a_n \in \mathbf{R}$ are the eigenvalues and P_n the eigenprojectors of A. The eigenspaces of A are spanned by the orthonormal eigenvectors $|\phi_{nj}\rangle$ associated to the eigenvalue a_n:

$$A|\phi_{nj}\rangle = a_n|\phi_{nj}\rangle, \;\; P_n = \sum_j |\phi_{nj}\rangle\langle\phi_{nj}|$$

The index j takes into account the possible degeneracy of a_n. From the orthonormality of the eigenvectors one sees that $P_nP_m = P_mP_n = 0$ for $n \neq m$. Note that for given n one always has the liberty to rotate the basis $\{|\phi_{nj}\rangle\}$ in the subspace of P_n. Moreover we have the closure relation

$$I = \sum_n P_n = \sum_{n,j} |\phi_{n,j}\rangle\langle\phi_{nj}|$$

We will often write the spectral decomposition as

$$A = \sum_{n,j} a_n|\phi_{nj}\rangle\langle\phi_{nj}|$$

In the non-degenerate case this becomes simply $A = \sum_n a_n|\phi_n\rangle\langle\phi_n|$.

2.2 Principles of quantum mechanics

In this paragraph we explain the 5 basic principles of QM. In a nutshell:

- isolated systems are described by *states of a Hilbert space*,
- they *evolve unitarily with time*,
- *observable* quantities are described by *hermitian matrices*,

- measurement is a *distinct process* from time evolution: it is a *random projection*,
- systems can be brought together and *composed*: their Hilbert space is a tensor product space.

Their meaning, interpretation and soundness has been debated over the first half of the 20-th century by the founding fathers of QM and by their followers, specially the measurement postulate.

Principle 1: states. The state of a quantum system - that is isolated from the rest of the universe - is *completely* described by a ray in a Hilbert space. We require that the representative vector $|\psi\rangle \in \mathcal{H}$ is normalized to one, $\langle\psi|\psi\rangle = 1$.

Example 4.

- To describe the polarization of the photon we take $\mathcal{H} = \mathbf{C}^2$. States are vectors in $\mathbf{C}^2$, $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, $|\alpha|^2 + |\beta|^2 = 1$. For a linearly polarized state $|\theta\rangle = \cos\theta|0\rangle + \sin\theta|1\rangle$, for a circularly polarized state $|\tilde{\theta}\rangle = \cos\theta|0\rangle + i\sin\theta|1\rangle$, and for elliptic polarization $\cos\theta|0\rangle + e^{i\delta}\sin\theta|1\rangle$.
- The spin $\frac{1}{2}$ of an electron (say) is described by the same Hilbert space.
- For a Benzene molecule the Hilbert space is again the same and is spanned by the two valence bond states (see chapter 1):

$$|\psi\rangle = \alpha|1\rangle + \beta|2\rangle$$

- For a particle in $\mathbf{R}^3$ we have $\mathcal{H} = L^2(\mathbf{R}^3)$ as explained before. These are called wave functions and are normalized $\int d^3x|\psi(x)|^2 = 1$.

Remark. If $|\psi\rangle$ is a description of a system then $e^{i\lambda}|\psi\rangle$ is an equally good description. The *global* phase $\lambda \in \mathbf{R}$ is not an observable quantity and can be fixed arbitrarily. This is why QM states should really be defined as rays. However the *relative* phase of states is observable through interference effects. You might also wonder what is the difference between spin one-half and photon polarization. In fact photon polarization states and spin one-half states behave very differently under spatial rotations of the coordinate system (or the lab). Under a rotation of the reference frame the state of polarization of a photon behaves like a vector. In particular under a 2π rotation we recover the same state. On the other hand for spin one-half behaves as a spinor (sometimes called half-vector) under a rotation of the reference frame. In particular, under a 2π rotation we recover the opposite state. In QM the representations of the rotation group (and any other group) on the Hilbert space does not have to satisfy $\mathbf{R}(2\pi) = 1$, precisely because states are rays. Therefore a phase is allowed for $\mathbf{R}(2\pi)|\psi\rangle = e^{i\lambda}|\psi\rangle$. All these aspects of QM will not mater too much in this course so we omit more explanations on what "spin" and "photon polarization" really are. A more profound discussion of these aspects would require to explain the representation theory of the Lorentz group of special relativity.

Principle 2: time evolution. An isolated quantum system evolves with time in a unitary fashion. This means that if $|\psi\rangle$ is the state at time 0, the state at time t is of the form $U_t|\psi\rangle$ where U_t is a unitary operator from $\mathcal{H} \to \mathcal{H}$. Here unitary means that $U_t^\dagger U_t = U_t U_t^\dagger = 1$ or equivalently $U_t^{-1} = U_t^\dagger$.

Unitary time evolution forms a group (it is a representation of translations along the time axis) in the sense that

$$U_{t=0} = I, \qquad U_{t_1} U_{t_2} = U_{t_1+t_2}$$

QM tells us how to compute U_t for a given system: one has to solve the *Schroedinger equation* or the *Heisenberg equations of motion*. These are equivalent in fact. The first one is the quantum mechanical version of the Hamilton-Jacobi equation of classical mechanics while the second is the quantum version of the Hamilton equations of motion. In quantum computation (at least in theory) we do not bother too much about these equations : we optimistically assume that if we need a specified U_t then somebody (a physicist, an engineer) will be able to construct a device (an electronic or optical device for example) which realizes the time evolution U. For us a specified time evolution is a *gate* that will ultimately be part of a quantum circuit.

It is very important to realize that *time evolution is linear*: this is quite surprising because in the classical regime one should get back the classical equations of motion which are generally non-linear[3].

Example 5. A semi-transparent mirror decomposes an incident ray into a reflected and a transmitted part (see chapter 1). Let $\mathcal{H} = \mathbf{C}^2$ the Hilbert space with basis $|T\rangle$, $|R\rangle$. The semi-transparent mirror acts in a unitary way

$$|T\rangle \to \boxed{\text{H}} \to H|T\rangle = \frac{1}{\sqrt{2}}(|T\rangle + |R\rangle)$$

$$|R\rangle \to \boxed{\text{H}} \to H|R\rangle = \frac{1}{\sqrt{2}}(|T\rangle - |R\rangle)$$

The unitary matrix H is called a Hadamard gate

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

One checks that $HH^\dagger = H^\dagger H = 1$. If we put two semi-transparent mirrors in series (see exercises)

$$|\psi\rangle \to \boxed{\text{H}} \to \boxed{\text{H}} \to H^2|\psi\rangle = |\psi\rangle$$

the output is equal to the input because $H^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. In other words if the input state is $|T\rangle$ then the output is also $|T\rangle$. If we wish to take more seriously

[3] The study of this sort of reduction has led to a whole discipline called quantum chaos. Let us also point out that non-linear versions of the Schroedinger equation may arise when some degrees of freedom are integrated out, in other words for non-isolated systems.

into account the effect of the perfect mirrors in-between the semi-transparent mirrors, we insert between the two Hadamard matrices the gate $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

$$|\psi\rangle = \alpha|T\rangle + \beta|R\rangle \rightarrow \boxed{\text{H}} \rightarrow \boxed{\text{X}} \rightarrow \boxed{\text{H}} \rightarrow HXH|\psi\rangle = \alpha|T\rangle - \beta|R\rangle$$

Principle 3: observable quantities. In quantum mechanics an observable quantity (energy, magnetic moment, position, momentum,...) is represented by a linear self-adjoint operator[4] on $\mathcal{H}$. For us this just means a hermitian matrix.

Examples 6.

- Position x, momentum $p = \frac{\hbar}{i}\frac{\partial}{\partial x}$, energy or Hamiltonian $\frac{p^2}{2m} + V(x)$. We will not need these.
- However we will need things like the polarization of a photon. Suppose we send a photon in a polarized beam-splitter (see chapter 1). If D_y clicks we record a -1 while if D_x clicks we record a $+1$. Our observations can be described by the observable

  $$\mathcal{P} = (+1)|x\rangle\langle x| + (-1)|y\rangle\langle y|$$

  This is the self-adjoint matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ (in the $|x\rangle$, $|y\rangle$ basis).
- General observables in $\mathcal{H} = \mathbf{C}^2$ can always be represented by 2×2 hermitian matrices

  $$A = \begin{pmatrix} \alpha & \beta \\ \overline{\beta} & \gamma \end{pmatrix}$$

  or in Dirac notation

  $$A = \alpha|0\rangle\langle 0| + \beta|0\rangle\langle 1| + \overline{\beta}|1\rangle\langle 0| + \gamma|1\rangle\langle 1|$$

  All such matrices can be written as linear combinations of

  $$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

  The observables (hermitian matrices !) X, Y, Z are called Pauli matrices. One of their uses is the description of the spin observable for spin $\frac{1}{2}$ particles: this is a "vector" with 3 components $\Sigma = (X, Y, Z)$. In the physics literature the notation is $\Sigma = (\sigma_x, \sigma_y, \sigma_z)$. Important properties of these matrices are

  $$X^2 = Y^2 = Z^2 = I,\ XY = -YX,\ XZ = -ZX,\ YZ = -ZY$$

  and

  $$[X, Y] = 2iZ,\ [Y, Z] = 2iX,\ [Z, X] = 2iY$$

[4] There is a "correspondence principle" which is a rule of thumb on how to construct the appropriate self-adjoint operator from the classical one; in fact this procedure may sometimes be a bit ambiguous due to non-commutativity of operators

This algebra is a special example of spin or Clifford algebras which play an important role in QM.

Principle 4: measurement postulate. This is the most disturbing postulate: it requires a rather big leap of intuition (or stroke of genius) which goes back to Max Born (one also speaks of the Born interpretation of the wave function). Let a system be prepared in a state $|\psi\rangle$. The system is to be measured with an apparatus. The apparatus is modeled by a set of orthonormal projectors $\{P_n\}$ satisfying $\sum_n P_n = I$. A single measurement *reduces*[5] the state ψ of the system to

$$|\phi_n\rangle = \frac{P_n|\psi\rangle}{||P_n|\psi\rangle||} = \frac{P_n|\psi\rangle}{\langle\psi|P_n|\psi\rangle^{1/2}}$$

For a single measurement *there is no way to predict* what will be the specific outcome n: it is random. If the experiment is repeated many times (assuming this is a reproducible experiment) one finds that the probability (in a frequentist interpretation of the term) of the outcome n is

$$\text{Prob(outcome } n) = |\langle\phi_n|\psi\rangle|^2 = \langle\psi|P_n|\psi\rangle$$

Remark 1. Since $\sum_j P_j = I$ and $|\psi\rangle$ are normalized we have $\sum_j \text{Prob(outcome } j) = 1$.

Remark 2. When the eigenprojectors are not degenerate these formulas are slightly simpler. If $P_j = |j\rangle\langle j|$ the probability of the outcome j is

$$\text{Prob(outcome } j) = \langle\psi|P_j|\psi\rangle = |\langle j|\psi\rangle|^2$$

and the state just after the measurement is $|j\rangle$.

Consequences for the measurement of observables. This is a very important point because ultimately one really measures physical quantities. The above measurement apparatus $\{P_n\}$ gives the value of any observable of the form $A = \sum_j a_j P_j$. The measurement makes $|\psi\rangle \to |\phi_n\rangle$ for some n. Since $A|\phi_n\rangle = a_n|\phi_n\rangle$ the value of A given by the measurement is precisely a_n when the outcome is n. In particular we can know simultaneously the value of many observables, by measuring them with the same appratus, as long as they have the same eigenspaces. Such observables commute and are sometimes said to be compatible.

The average value that the measurement, on the state $|\psi\rangle$, will yield can be calculated from the probability distribution above. One finds

$$\sum_j a_j \langle\psi|P_j|\psi\rangle = \langle\psi|A|\psi\rangle$$

[5] physicist are used to say that "the wave function collapses"

and the variance is

$$\sum_j a_j^2 \langle\psi|P_j|\psi\rangle - (\sum_j a_j \langle\psi|P_j|\psi\rangle)^2 = \langle\psi|A^2|\psi\rangle - \langle\psi|A|\psi\rangle^2$$

In practice one uses the right hand side of these two formulas. That is basically all that a theorist can predict.

After a measurement the state vector is reduced $|\psi\rangle \to |\phi_n\rangle$, for some n, and thus the expectation value in the new state (i.e $|\phi_n\rangle$) becomes a_n and the variance 0. This means that if we repeat the same measurement on the same state we will get precisely the value a_n again and again.

We will return to this point when we will consider the Heisenberg uncertainty principle.

Example 7: measurement of photon polarization. Suppose we want to measure the observable $\mathcal{P} = |x\rangle\langle x| - |y\rangle\langle y|$ For this we use the apparatus constituted of an analyzer oriented along x and a detector. This apparatus is the physical realization of the measurement basis. If a photon is detected the state just after the measurement is $|x\rangle$ and if a photon is not detected (it has been absorbed by the analyzer) the state just after the measurement is $|y\rangle$. The probabilities of these outcomes are

$$\text{Prob(outcome } +1) = |\langle x|\psi\rangle|^2, \ \text{Prob(outcome } -1) = |\langle y|\psi\rangle|^2$$

If the initial preparation of the beam is $|\psi\rangle = \cos\theta|x\rangle + \sin\theta|y\rangle$ these probabilities are simply $\cos^2\theta$ and $\sin^2\theta$. Suppose that now we rotate the analyzer by an angle γ. This means that we wish to measure the observable $\mathcal{P} = |\gamma\rangle\langle\gamma| - |\gamma_\perp\rangle\langle\gamma_\perp|$. then we can compute again the probabilities of the outcomes

$$\text{Prob(outcome } +1) = |\langle\gamma|\psi\rangle|^2 = \cos^2(\theta - \gamma)$$

$$\text{Prob(outcome } -1) = |\langle\gamma_\perp|\psi\rangle|^2 = \sin^2(\theta - \gamma)$$

Finally let us note that in the first case the measured observable in matrix form is

$$\mathcal{P} = Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

and in the second

$$\mathcal{P} = \begin{pmatrix} \cos 2\gamma & \sin 2\gamma \\ \sin 2\gamma & -\cos 2\gamma \end{pmatrix} = (\cos 2\gamma)Z + (\sin 2\gamma)X$$

Uncertainty principle Suppose that we have a system in a state ψ and we consider two observables A and B. We assume that these have spectral representations

$$A = \sum_j a_j P_j, \quad B = \sum_j b_j Q_j$$

As discussed previously in a general state $|\psi\rangle$ each of these is not fixed but has an average value $\langle\psi|A|\psi\rangle$, $\langle\psi|B|\psi\rangle$ and a standard deviation $\Delta A = \sqrt{\langle\psi|A^2|\psi\rangle - \langle\psi|A|\psi\rangle^2}$, $\Delta B = \sqrt{\langle\psi|B^2|\psi\rangle - \langle\psi|B|\psi\rangle^2}$. The Heisenberg uncertainty relation states that

$$\Delta A \cdot \Delta B \geq \frac{1}{2}\langle\psi|[A,B]|\psi\rangle$$

The interpretation of this inequality as first discussed by Heisenberg is that when $[A,B] \neq 0$ it is not possible to measure A and B simultaneously with infinite precision. If we manage to make $\Delta A = 0$ then we will have $\Delta B = \infty$. The prototypical and most striking example is $A = x$ (position) and $B = p = \frac{\hbar}{i}\frac{\partial}{\partial x}$ (momentum). In this case $\Delta x \Delta p \geq \frac{h}{4\pi}$ and we cannot measure simultaneously with infinite precision the position and the momentum of a particle: this is not a technological limitation but ultimately a "God given" limitation.

Note that if $[A,B] = 0$ then there exist a common basis of the Hilbert space in which A and B are both diagonal. Then by measuring in this basis, the measurement postulate tells us that both observables can be determined with infinite precision. There is no clash with the uncertainty relation because the right hand side of the inequality vanishes.

There is a related principle called the "entropic uncertainty principle" which we now state. Suppose A and B have non degenerate eigenvalues

$$A = \sum_{n_a} a_{n_a} |n_a\rangle\langle n_a|$$

$$B = \sum_{m_b} b_{m_b} |m_b\rangle\langle m_b|$$

Set

$$H(A) = -\sum_{n_a} p(n_a)\ln p(n_a), \qquad H(B) = -\sum_{m_b} p(m_b)\ln p(m_b)$$

where

$$p(n_a) = |\langle n_a|\psi\rangle|^2, \qquad p(m_b) = |\langle m_b|\psi\rangle|^2$$

We have

$$H(A) + H(B) \geq -2\ln\Big(\frac{1 + \max_{n_a,m_b}|\langle n_a|m_b\rangle|}{2}\Big)$$

Principle 5: composite quantum systems. Suppose we have two systems $\mathcal{A}$ and $\mathcal{B}$ with Hilbert spaces $\mathcal{H}_\mathcal{A}$ and $\mathcal{H}_\mathcal{B}$. The Hilbert space of the composite system $\mathcal{AB}$ is given by the tensor product space

$$\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$$

The states of $\mathcal{AB}$ are vectors $|\psi\rangle \in \mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$. The previous postulates apply to the composite system.

This is also a highly non trivial postulate as will be seen from its consequences throughout the course. In a famous paper Einstein, Podolsky, Rosen were the first to make a sharp analysis of its consequences. This has ultimately led to Bell inequalities and to important primitive protocols of quantum information such as teleportation and dense coding.

Example 8. Two photons with polarization degrees of freedom have Hilbert space $\mathbf{C}^2 \otimes \mathbf{C}^2$. Examples of states are $|x\rangle_{\mathcal{A}} \otimes |y\rangle_{\mathcal{B}}$ or $|x\rangle_{\mathcal{A}} \otimes |y\rangle_{\mathcal{B}} + |\theta\rangle_{\mathcal{A}} \otimes |\theta\rangle_{\mathcal{B}}$. N Qbits live in the space

$$\underbrace{\mathbf{C}^2 \otimes \mathbf{C}^2 \otimes \mathbf{C}^2 \otimes ... \otimes \mathbf{C}^2}_{N \text{ copies}}$$

If $|0\rangle, |1\rangle$ is a canonical basis for $\mathbf{C}^2$, a basis for the composite system is given by

$$|b_1\rangle \otimes |b_2\rangle ... \otimes |b_N\rangle = |b_1, ..., b_N\rangle$$

where $b_i = \{0, 1\}$. There are 2^N such states and they are in one to one correspondence with the 2^N classical bit strings of length N. A general N Qbit state is a linear superposition of the basis states:

$$|\psi\rangle = \sum_{b_1,...,b_N} c_{b_1,...,b_N} |b_1, ..., b_N\rangle$$

where the coefficients $c_{b_1 ... b_N}$ satisfy

$$\sum_{b_1,...,b_N} |c_{b_1,...,b_N}|^2$$

2.3 Tensor product versus entangled states

States of a composite system $\mathcal{AB}$ lie in $\mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{B}}$. We say that a state is a *tensor product state* (or is *not entangled*) if it can be written as

$$|\psi\rangle = |\phi\rangle_{\mathcal{A}} \otimes |\chi\rangle_{\mathcal{B}}$$

An entangled state $|\psi\rangle \in \mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{B}}$ is one for which it is impossible to find $|\phi\rangle_{\mathcal{A}} \in \mathcal{H}_{\mathcal{A}}$ and $|\chi\rangle_{\mathcal{B}} \in \mathcal{H}_{\mathcal{B}}$ such that ψ is of the tensor product form.

Entangled states have very special correlations between their parts $\mathcal{A}$ and $\mathcal{B}$. These are genuine quantum correlations with no classical counterpart and as we will see later in the course they play a very important role (for example in teleportation). These definitions generalize to multipartite systems.

example 9. Two Qbit system with $\mathcal{A} \otimes \mathcal{B} = \mathbf{C}^2 \otimes \mathbf{C}^2$. Some product states are : $|0\rangle_{\mathcal{A}} \otimes |0\rangle_{\mathcal{B}} = |0, 0\rangle$, $|0\rangle_{\mathcal{A}} \otimes |1\rangle_{\mathcal{B}} = |0, 1\rangle$, $|1\rangle_{\mathcal{A}} \otimes |0\rangle_{\mathcal{B}} = |1, 0\rangle$, $|1\rangle_{\mathcal{A}} \otimes |1\rangle_{\mathcal{B}} = |1, 1\rangle$. Two less trivial ones are

$$\frac{1}{\sqrt{2}}(|0\rangle_{\mathcal{A}} + |1\rangle_{\mathcal{B}}) \otimes |0\rangle_{\mathcal{B}} = \frac{1}{2}(|0, 0\rangle + |1, 0\rangle)$$

and

$$\frac{1}{\sqrt{2}}(|0\rangle_{\mathcal{A}} + |1\rangle_{\mathcal{B}}) \otimes \frac{1}{\sqrt{2}}(|0\rangle_{\mathcal{B}} - |1\rangle_{\mathcal{B}}) = \frac{1}{2}(|0,0\rangle - |0,1\rangle + |1,0\rangle - |1,1\rangle)$$

In the same space there are also entangled states that simply *cannot* be written as a tensor product form. For example,

$$\frac{1}{\sqrt{2}}(|0\rangle_{\mathcal{A}} \otimes |0\rangle_{\mathcal{B}} + |1\rangle_{\mathcal{A}} \otimes |1\rangle_{\mathcal{B}}) = \frac{1}{\sqrt{2}}(|0,0\rangle + |1,1\rangle)$$

$$\frac{1}{\sqrt{2}}(|0\rangle_{\mathcal{A}} \otimes |0\rangle_{\mathcal{B}} - |1\rangle_{\mathcal{A}} \otimes |1\rangle_{\mathcal{B}}) = \frac{1}{\sqrt{2}}(|0,0\rangle - |1,1\rangle)$$

$$\frac{1}{\sqrt{2}}(|1\rangle_{\mathcal{A}} \otimes |0\rangle_{\mathcal{B}} + |0\rangle_{\mathcal{A}} \otimes |1\rangle_{\mathcal{B}}) = \frac{1}{\sqrt{2}}(|1,0\rangle + |0,1\rangle)$$

$$\frac{1}{\sqrt{2}}(|0\rangle_{\mathcal{A}} \otimes |0\rangle_{\mathcal{B}} - |1\rangle_{\mathcal{A}} \otimes |0\rangle_{\mathcal{B}}) = \frac{1}{\sqrt{2}}(|0,1\rangle + |1,0\rangle)$$

As we will see these four particular states play a special role and are called Bell states. The reader can check that they form a basis of the 2 Qbit space.

Production of entangled states. Suppose we have a composite system in an initial tensor product state $|\phi\rangle_{\mathcal{A}} \otimes |\chi\rangle_{\mathcal{B}}$. These could for example be two electrons in the spin state $|\uparrow\rangle \otimes |\downarrow\rangle$. If we let them evolve separately and without interaction, the unitary operator for the time evolution is of the form $U_{\mathcal{A}} \otimes U_{\mathcal{B}}$ and

$$U_{\mathcal{A}} \otimes U_{\mathcal{B}}(|\uparrow\rangle \otimes |\downarrow\rangle) = U_{\mathcal{A}}|\uparrow\rangle \otimes U_{\mathcal{B}}|\downarrow\rangle$$

so that the system remains in a tensor product state.

Thus to produce entangled states systems $\mathcal{A}$ and $\mathcal{B}$ must interact at some point in time in order to have an evolution $U_{\mathcal{AB}} \neq U_{\mathcal{A}} \otimes U_{\mathcal{B}}$. With an appropriate interaction we might be able to achieve

$$U_{\mathcal{AB}}(|\uparrow\rangle \otimes |\downarrow\rangle)$$

All known physical interactions are local: this means that in order to interact (in a non-negligible way) two systems must be "close in space-time". In particular if we are presented with an entangled state we know that the two parties have interacted in the past, i.e they have been "sufficiently close in the past".

2.4 No cloning theorem

Classical bits can be copied. For example any latex file can be duplicated or any text can be copied with a (universal) Xerox machine.

Suppose we have a set of quantum states $|\psi\rangle \in \mathcal{H}$ and we want to build a (universal) "quantum Xerox machine" to copy $|\psi\rangle$. This machine should be able to copy any state of $\mathcal{H}$. A quantum Xerox machine should be described by some

unitary operator U (this is true for any physical process except measurement). The Hilbert space is composite $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$ where $\mathcal{A}$ is the quantum file to be copied and $\mathcal{B}$ the duplicated file. We start from the state

$$|\psi\rangle \otimes |\text{blank}\rangle$$

and we feed it in the Xerox machine

$$|\psi\rangle \otimes |\text{blank}\rangle \rightarrow \boxed{\text{U}} \rightarrow |\psi\rangle \otimes |\psi\rangle$$

In mathematical terms the question is: can one find a unitary operator such that for a reasonably large set of ψ

$$U(|\psi\rangle \otimes |\text{blank}\rangle) = |\psi\rangle \otimes |\psi\rangle$$

The answer is NO and this is sometimes called the "no cloning theorem". However it is possible to copy a set of orthogonal states with an appropriate U depending on the set.

Proof of no-cloning theorem. Suppose there exists U such that $U^\dagger U = UU^\dagger = 1$ with

$$U(|\phi_1\rangle \otimes |\text{blank}\rangle) = |\phi_1\rangle \otimes |\phi_1\rangle$$

$$U(|\phi_2\rangle \otimes |\text{blank}\rangle) = |\phi_2\rangle \otimes |\phi_2\rangle$$

conjugating the second equation

$$(\langle\phi_2| \otimes \langle\text{blank}|)U^\dagger = \langle\phi_2| \otimes \langle\phi_2|$$

Taking the inner product with the first equation

$$\langle\phi_2| \otimes \langle\text{blank}|U^\dagger U|\phi_1\rangle \otimes |\text{blank}\rangle = (\phi_2| \otimes \langle\phi_2|)(|\phi_1\rangle \otimes |\phi_1\rangle)$$

which implies

$$\langle\phi_2|\phi_1\rangle\langle\text{blank}|\text{blank}\rangle = \langle\phi_2|\phi_1\rangle^2$$

so

$$\langle\phi_2|\phi_1\rangle = 0 \text{ or } \langle\phi_2|\phi_1\rangle = 1$$

We conclude that we cannot copy states $|\phi_1\rangle$ and $|\phi_2\rangle$ that are not identical or orthogonal, with the same U. In fact it is possible to copy a given orthogonal basis . To see this the reader has to construct a unitary operation that does the job.

Non orthogonal states cannot be perfectly distinguished. There are many variants and refinements of the no-cloning theorem. let us just show one such variant. Suppose we have two states $|\psi\rangle$ and $|\phi\rangle$ and we want to build a (unitary) machine to distinguish them. We seek a U such that

$$U|\psi\rangle \otimes |a\rangle = |\psi\rangle \otimes |v\rangle$$

$$U|\phi\rangle \otimes |a\rangle = |\phi\rangle \otimes |v'\rangle$$

where the outputs $|v\rangle$ and $|v'\rangle$ give some information about $|\psi\rangle$ and $|\phi\rangle$. Taking the inner product of these two equations yields

$$\langle\phi| \otimes \langle a|U^\dagger U|\psi\rangle \otimes |a\rangle = (\langle\phi| \otimes \langle v'|)(|\psi\rangle) \otimes |v\rangle)$$

This implies

$$\langle\phi|\psi\rangle\langle a|a\rangle = \langle\phi|\psi\rangle\langle v'|v\rangle$$

If $|\phi\rangle$ is not orthogonal to $|\psi\rangle$ we have $\langle\phi|\psi\rangle \neq 0$ thus

$$\langle v'|v\rangle = \langle a|a\rangle = 1$$

Thus $|v\rangle = |v'\rangle$ so there is no information in $|v\rangle$ and $|v'\rangle$ distinguishing $|\psi\rangle$ and $|\phi\rangle$.

3 Quantum key distribution

One of the first applications of quantum mechanics to the field of information theory has been the 1984 proposal of Bennett and Brassard for a secure protocol to distribute a secret key that is common to two distant parties. Since then, there have been a few other similar protocols and a new field has emerged, called "quantum cryptography". In this chapter we limit ourself to the original protocol - now called BB84 - and to a simpler one found by Bennet in 1992. In a later chapter we will also give another protocol proposed by Ekert in 1991, and based on entangled Einstein-Podolsky-Rosen pairs of particles.

The general idea of BB84 is as follows. Alice sends a string of classical bits - the secret key - to Bob by using intermediate quantum mechanical Qbits (in practice these are photons transmitted in optic fibers). Any attempt by Eve to capture some information about the key amounts to *observe* the Qbits, but according to the postulates of QM *this observation will perturb the quantum system.* Alice and Bob are then able to detect this perturbation, thus the presence of Eve, and abort communication.

The subject is in fact more complicated because in reality the channel (the optic fiber) is noisy and it is non-trivial to distinguish Eve from noise. Besides the operations performed by Alice and bob are not perfect. The proof of security (see [**?**]) for BB84 is therefore dependent on precise assumptions on the physical set-up. It involves a combination of non-trivial methods from classical and quantum information theory and is beyond the scope of this course. Here we will analyze only two basic attacks from Eve, assuming the channel is not noisy and the operations of Alice and Bob are perfect.

Quantum cryptography is not only a theoretical idea. It is also a truly experimental subject since the protocols have been implemented and shown to work in the laboratory (first at IBM in 1989 over a distance of 32 cm) and later outside the lab on distances of few tens to hundreds of kilometers (Geneva, Los Alamos ...). See [**?**] for a general review. Nowadays there exist companies proposing commercial systems[1]. Recent implementations allow the exchange of secret keys over a distance of $100km$ (resp. $250km$) at a rate of 6000 (resp. 15) bits per second [**?**]. require extensive knowledge of optics and will not be discussed here. Recently the commercial systems have been challenged by a hacking procedure exploiting the physical limitations of photo-detectors on Bob's side [**?**].

[1] Idquantique, MagiQ

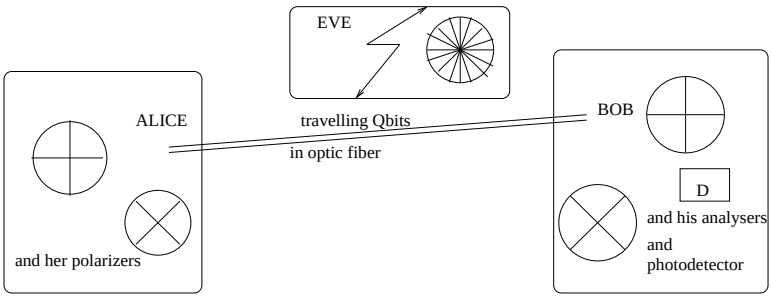


Figure 3.1 Alice and Bob exchange a private key over an optic fiber

Figure 3.2 orientations of polarizer for preparation of photons in Z basis

3.1 Key generation according to BB84

There are four essential phases: the encoding procedure of Alice, the decoding procedure of Bob, a public discussion between the two parties, and finally the common secret key generation. Figure 3.1 illustrates the general set-up described below.

Encoding procedure of Alice. She generates a classical random binary string $x_1, ..., x_N$, $x_i \in \{0, 1\}$ that she keeps secret. The common key will be a subset of these bits. She also generates a second classical random binary string $e_1, ..., e_N$, $e_i \in \{0, 1\}$ that she keeps secret *for the moment*. Alice then *encodes* the classical bits x_i into Qbits as follows:

- For $e_i = 0$ she generates a Qbit in the state $|x_i\rangle$. Concretely this can be done by sending a beam through a polarizer in the Z basis (figure 3.2)

  $$\big\{|0\rangle, |1\rangle\big\}$$

  For $x_i = 0$ (resp. $x_i = 1$) the polarizer is oriented horizontally (resp. vertically) and so photons are prepared in polarization state $|0\rangle$ (resp. $|1\rangle$). A single photon is then selected from the outgoing beam (this of course is an idealization)
- For $e_i = 1$ she generates a Qbit in the state[*2] $H|x_i\rangle$. Concretely this can be

[2] We remind the reader that H is the Hadamard matrix $\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$

Figure 3.3 orientations of polarizer for preparation of photons in X basis

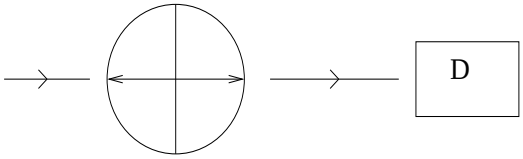


Figure 3.4 analyzer-detector set-up for the measurement of polarization in Z basis

done by sending photons through a polarizer in the X basis (figure 3.3

$$\{\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\}$$

For $x_i = 0$ (resp,. $x_i = 1$) the polarizer is rotated to the right (resp. left) and photons are prepared in polarization state $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ (resp. $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$).

Summarizing, Alice sends a string of Qbits $|A_{e_i,x_i}\rangle = H^{e_i}|x_i\rangle$, $i = 1, ..., N$ through a channel (in practice the channel is an optical fiber).

Decoding procedure of Bob. Bob generates a random classical binary string $d_1, ..., d_N$, $d_i \in \{0, 1\}$ that he keeps secret *for the moment*. He decodes the received Qbits of Alice as follows:

- If $d_i = 0$ he performs a measurement of the received Qbits $|A_{e_i,x_i}\rangle$ in the Z basis

  $$\{|0\rangle, |1\rangle\}.$$

  The photon state after the measurement

  $$|y_i\rangle \in \{|0\rangle, |1\rangle\}.$$

  is recorded in the bit y_i. To do this concretely he uses the analyzer-detector apparatus described in the first chapter: the analyzer is placed horizontally (figure 3.4); if the detector clicks this means the photons state has *collapsed* in the $|0\rangle$ state and if the detector does not click, it means that the photon state has collapsed to $|1\rangle$. We stress that, according to the measurement postulate, these outcomes are *truly random*. Only Bob knows about them.

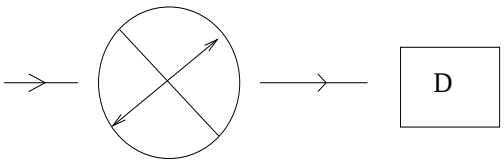


Figure 3.5 analyzer-detector set-up for the measurement of polarization in X basis

- If $d_i = 1$ he performs a measurement of the received Qbits $|A_{e_i,x_i}\rangle$ in the X basis

$$\{\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle, \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle\}.$$

  The photon sate after the measurement is in

$$H|y_i\rangle \in \{\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\}$$

  When the output is $H|y_i\rangle$, and he records the bit y_i.

  To do this concretely he uses the analyzer-detector apparatus described in the first chapter: the analyzer is rotated to the right (figure 3.5) at 45 degrees; if the detector clicks this means the photons state has *collapsed* in the $H|0\rangle$ state while if the detector does not click it means that the photon state has collapsed to $H|1\rangle$. We stress again that, according to the measurement postulate, these outcomes are *truly random* and that only Bob knows about them.

In summary Bob has decoded the Qbits sent by Alice to a classical binary string $y_1, ..., y_N$. This string is the outcome of measurements of Bob and cannot be predicted (God does play with dice ... the statistics of the outcomes can however be calculated according to the measurement postulate).

Public discussion. Alice has at her disposal two binary strings: $e_1, ..., e_N$ used to choose the encoding basis, and $x_1, ..., x_N$ that was mapped to Qbits. Bob also has two binary strings: $d_1, ..., d_N$ used to choose a measurement basis and $y_1, ..., y_N$ that are his measurement outcomes.

Alice and Bob compare $e_1, ..., e_N$ and $d_1, ..., d_N$ over a public channel, but keep their two other strings $x_1, ..., x_N$ and $y_1, ..., y_N$ secret. *It important that the public discussion starts only after Bob has finished his measurements.* They can deduce the following information (and anybody else hearing the public discussion also can):

- If $d_i = e_i$, i.e. if they used the same basis, then it must be the case that $y_i = x_i$ (the reader should convince himself of that by going through some examples with polarizer, analyzer pairs - basically if Bob and Alice used the same basis it is as if they lived in a classical world).
- If $d_i \neq e_i$, i.e. if they did not use the same basis, then genuine quantum effects

came into play when Bob did the measurement. According to the measurement postulate: $y_i \neq x_i$ with probability $\frac{1}{2}$ and $y_i = x_i$ with probability $\frac{1}{2}$. Let us formally prove this. Bob receives the Qbit

$$|A_{e_i,x_i}\rangle = H^{e_i}|x_i\rangle$$

and measures in the basis

$$\{H^{d_i}|0\rangle, H^{d_i}|1\rangle\}.$$

The outcome will be one of two basis states

$$H^{d_i}|0\rangle, \qquad \text{with prob } |\langle 0|H^{d_i}H^{e_i}|x_i\rangle|^2$$

or

$$H^{d_i}|1\rangle, \qquad \text{with prob } |\langle 1|H^{d_i}H^{e_i}|x_i\rangle|^2.$$

The reader can check that for $e_i \neq d_i$ both probabilities are equal to $\frac{1}{2}$ (and that for $e_i = d_i$ they are 0 and 1).

Key generation. Bob and Alice erase all bits x_i and y_i corresponding to i such that $e_i \neq d_i$. They keep the remaining sub-strings of $x_1, ..., x_n$ and $y_1, ..., y_n$ such that $e_i = d_i$. They are assured that these two sub-strings are identical, so this can potentially constitute the common secret key. The length of this sub-string is close to $\frac{N}{2}$ since $\text{prob}(e_i \neq d_i) = \frac{1}{2}$. Finally Alice and Bob perform a security test: according to quantum mechanics for this perfect setting (without noise or Eve) one must have

$$\text{prob}(x_i = y_i | e_i = d_i) = 1$$

Alice and Bob test this by exchanging a small fraction of the common sub-string over the public channel. If the test succeeds they keep the rest of the common sub-string secret: they have succeeded in generating a common secret key.

3.2 Attacks from Eve

We assume that Alice has a perfect single-photon source, state preparation is perfect, there is no channel noise, Bobs analyzer-detector apparatus makes no detection errors. In summary when Eve is absent communication is error-free, and any error discovered in the security test would come from Eve. Furthermore we suppose that Eve may attack by performing operations on one Qbit at a time on captured photons along the optic fiber and that she has no access to the Alice and Bob's labs. We also suppose that Eve has perfect knowledge of the set-up in Alice and Bob's labs: she knows that they use X and Z basis (but not the successive random basis choices), she knows what is their common vertical and horizontal directions, and the timing of the photons.

We consider two possible attacks : "the measurement" and "unitary" attacks.

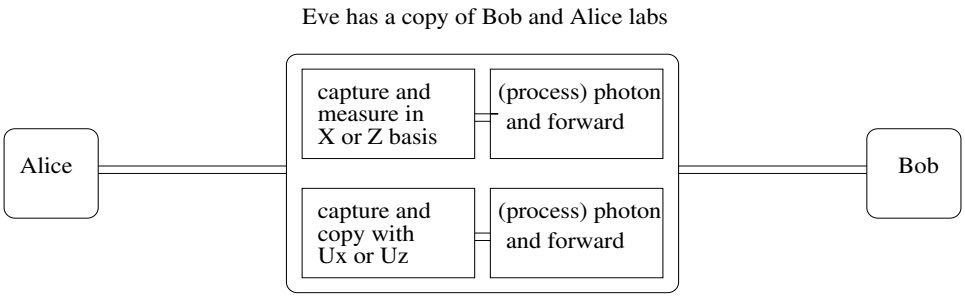


Figure 3.6 Set up of Eve's lab along the optic fiber

The two attacks consist of two steps. First Eve captures a photon, and second she forwards the photon to Bob (see figure 3.6). For each attack we will see that the basic postulates of QM imply that Bob and Alice discover the presence of Eve. when this is the case they abort the protocol.

Measurement attack. Suppose Eve captures a single photon in the optic fiber. The captured photon is in one of the the states

$$|A_{e_i,x_i}\rangle \in \big\{|0\rangle, |1\rangle, H|0\rangle, H|1\rangle\big\}$$

and she tries to measure it. If Eve uses the Z basis her outcome is in $\{|0\rangle, |1\rangle\}$ and according to it she records a bit $y_i^E \in \{0,1\}$. If she uses the X basis her outcome is in $\{H|0\rangle, H|1\rangle\}$ and she records a corresponding bit $y_i^E \in \{0,1\}$. Once she has finished the measurement she sends the photon to Bob in the state left over by the measurement[3]. Two possibilities may occur:

- Eve has used the same basis than Alice: then her outcome is $y_i^E = x_i$ and the photon state received by Bob is the "correct one".
- Eve uses a different basis than Alice: then her outcome $y_i^E = x_i$ only half of the time, so she sends the "correct" photon state to Bob only half of the time.

Let us see what Alice and Bob find when they perform the security test. Denote by EA the event "Eve uses the same basis than Alice".

$$\begin{aligned}\text{prob}(x_i = y_i | e_i = d_i) &= \text{prob}(x_i = y_i | e_i = d_i, EA)\text{prob}(EA) \\ &\quad + \text{prob}(x_i = y_i | e_i = d_i, \text{not } EA)\text{prob}(\text{not } EA) \\ &= 1 \cdot \text{prob}(EA) + \frac{1}{2} \cdot (1 - \text{prob}(EA)) \\ &= \frac{1}{2}(1 + \text{prob}(EA))\end{aligned}$$

[3] She could also further process this state by a unitary transformation but this will not improve her performance

where we used

$$\text{prob}(x_i = y_i | e_i = d_i, EA) = 1, \qquad \text{prob}(x_i = y_i | e_i = d_i, \text{not EA}) = \frac{1}{2} \quad (3.1)$$

Assuming that Eve has no information about the basis choices of Alice we take $\text{prob}(EA) = \frac{1}{2}$. Then

$$\text{prob}(x_i = y_i | e_i = d_i) = \frac{3}{4}$$

so that Alice and Bob notice that when they used the same basis about a fourth of their bits do not agree. They conclude that an eavesdropper is at work and abort the communication.

Unitary attack. The problem of Eve is that when she makes a measurement she has no information about the basis that Alice chose. One possible solution would be to copy the traveling Qbits $|A_{e_i,x_i}\rangle$, then let the original state go to Bob, and keep the copy. When Alice and Bob enter in the public discussion phase she learns about the basis of Bob in which to measure the Qbit and thus for i such that $e_i = d_i$ she gets the same outcome as Bob $y_i^E = y_i = x_i$.

However the *no-cloning theorem* (which is a consequence of the unitary evolution postulate) guarantees that there does not exist a unitary "machine" such that

$$U(|A_{e_i,x_i}\rangle \otimes |\text{blank}\rangle) = |A_{e_i,x_i}\rangle \otimes |A_{e_i,x_i}\rangle$$

The point here is that $|A_{e_i,x_i}\rangle$ is one of

$$\{|0\rangle,\ |1\rangle,\ \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle),\ \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\}$$

which is a set of non-orthogonal states.

Eve could try to use two copy machines: one for copying the two states of the Z basis and another for copying the two states of the X basis. But this time she has no way of knowing which machine to use. She will use the wrong machine half of the time and again Alice and Bob will find that

$$\text{prob}(x_i = y_i | e_i = d_i) = \frac{3}{4}$$

Discussion of security issues. In the above error-free set-up it is relatively easy to generalize the proof of security in order to take into account any local operation of Eve on single photons. In a more realistic context one has to take into account the fact that the system is noisy. For example the optic fiber is not perfect and the photo-detectors may give false counts. Therefore the string sequences of Alice and Bob do not match perfectly even when $e_i = d_i$. For this reason one adds to the protocol two classical post-processing steps: information reconciliation and privacy amplification. Both steps are carried on the public classical channel. The first step is an error correcting phase while the second allows to reduce the information that Eve might have gained about the key during

the correction phase. The detailed analysis is non-trivial and the interested reader may consult the literature [**?**].

There are various problems that may arise due to physical limitations that do not quite enter into the framework of the security proofs. Recently a successful attack was implemented [**?**] by exploiting the fact that after a photo-detector click, the detectors enter in a mode were they operate classically. By shining light on them Eve is able to maintain them in a classical mode and in effect the Eve-Bob part of the transmission line is in effect classical. In this Eve can achieve complete control of the key.

3.3 The Bennett 1992 scheme

The analysis of BB84 has shown that the security ultimately relies on the fact that Alice encodes Qbits in non-orthogonal states. The B92 scheme retains this very fact and is even simpler than BB84. Below we just sketch the main idea. There are again four main phases:

Alice encodes. Alice prepares a random binary string $e_1, ..., e_N$. She sends to Bob $|A_{e_i}\rangle = |0\rangle$ if $e_i = 0$ and $|A_{e_i}\rangle = H|0\rangle = \frac{1}{2}(|0\rangle + |1\rangle)$ if $e_i = 1$. The encoding is thus $H^{e_i}|0\rangle$.

Bob decodes. Bob generates a random binary string $d_1, ..., d_N$ and measures the received Qbit according to the value of d_i in the Z or X basis and obtains an outcome in $\{|0\rangle, |1\rangle\}$ or in $\{H|0\rangle, H|1\rangle\}$. He decodes the bit as $y_i = 0$ if the outcome is $|0\rangle$ or $H|0\rangle$ and $y_i = 1$ if the outcome is $|1\rangle$ or $H|1\rangle$.

Public discussion. Bob announces over the public channel the bits y_i. Note that when $e_i = d_i$ we have $y_i = 0$ with probability 1. On the other hand when $e_i \neq d_i$ we have $y_i = 0$ with probability $\frac{1}{2}$ and $y_i = 1$ with probability $\frac{1}{2}$. Therefore from the public discussion Alice and Bob deduce that, given $y_i = 1$, surely $d_i = 1 - e_i$.

Key generation. Alice and Bob keep the secret bits $(e_i, d_i = 1 - e_i)$ for i such that $y_i = 1$ and discard the rest. The length of this sub-string is about $\frac{N}{2}$. they perform a security test on a fraction of the sub-string on the public channel by checking that

$$\text{prob}(d_i = 1 - e_i | y_i = 1) = 1$$

Again it is not hard to check that this security condition is violated under a measurement or a unitary attack of Eve. If that is the case Alice and Bob abort communication.

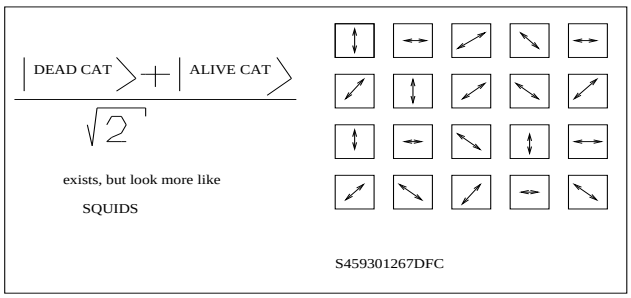


Figure 3.7 unforgeable bank note: it buys one Schroedinger cat

3.4 Conjugate coding

In the encoding method of Alice above the two basis that are used correspond to the basis diagonalizing the two Pauli matrices

$$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{3.2}$$

These two observables do not commute and are called conjugate observables by analogy with position and momentum; therefore the two basis are sometimes called *conjugate* and the corresponding scheme called *conjugate coding*.

In fact this scheme was first introduced in 1969 by Wiesner then a graduate student. Wiesner, basing himself on the principles of QM, indicated how to "fabricate unforgeable bank notes". Unfortunately nobody took him seriously, except for Bennett then also a graduate student, and his paper didn't get published till*[4] 1983. Bennett was one of the few persons who kept thinking about such problems and, with Gilles Brassard a computer scientist, had the idea to reconsider conjugate coding in the context of cryptography.

Let us briefly explain the original idea of Wiesner. One generates a random binary string $e_1, ..., e_{20}$, and prepares 20 photons in $|0\rangle, |1\rangle$ or $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, polarization states using Z or X polarizers. Then one traps the 20 photons in 20 small cavities inside the bank note. The bank note also contains a readable serial number which corresponds to the binary string $e_1, ..., e_{20}$. Only the bank knows what is the mapping between the serial number and the binary string (see figure 3.7).

Suppose somebody attempts to copy the bank note. Because of the no-cloning theorem there is no single machine U which copies simultaneously vertical and diagonal photon polarizations. If one uses two different machines one will make mistakes (with prob $1 - 2^{-20}$) because one doesn't know when to use a U_Z or a U_X. Moreover the bank can check if a bank note has been forged or not. Indeed from the serial number it deduces the binary string $e_1, ..., e_{20}$ and therefore

[4] around 1982 quantum computation came into fashion because of an equally pioneering work of Feynman

knows the basis sequence used to prepare the photons. A measurement in the correct basis (for each little cavity) is done to observe if the photons have the correct polarization. Note that if the bank note has *not* been forged it will *not* be destroyed by such a procedure. To summarize, one may say that the bank knows what exact sequence of analyzers to use so that the system behaves classically for the bank. For any other person that does not possess this information the system behaves quantum mechanically.

4 Quantum entanglement

In this chapter we study the nature of a special type of correlation displayed by the entangled states. These correlations have no classical counterpart, in other words, they cannot be described by classical probability distributions. They are genuine quantum mechanical correlations built up in the states of composite quantum systems.

We first take a close look at the so-called Bell states which violate the famous Bell inequalities[1]. These states display the essence of entanglement and the CHSH inequality provides an experimentally testable signature of it. We then describe three applications: a quantum key distribution protocol (Ekert 1991), quantum teleportation and dense coding. We stress here that all three of them have been experimentally realized, and form important primitive protocols for quantum communication.

In quantum information processing one tries to use entanglement as a quantifiable resource, much like energy or information, and it would be very convenient to be able to measure the degree or quantity of entanglement. Finding such a measure is however non-trivial. We will come back to this point in later chapters.

4.1 Bell states

Production of Bell states. We have seen in chapter 2 that in order to produce entangled states the Qbits must "interact", at some point in time. The prototypical example of entangled states are the Bell states which form a basis of $C^2 \otimes C^2$. These can be produced from the unitary gate

$$U = (CNOT)(H \otimes I) \tag{4.1}$$

This is a 4×4 matrix equal to the usual matrix product of the two 4×4 matrices $CNOT$ and $H \otimes I$. The *Control Not* gate provides the interaction between the two bits. It is defined as the NOT gate acting on the second bit provided the first one[2] equals 1

$$CNOT|x, y\rangle = |x, y \oplus x\rangle$$

[1] There is a class of such inequalities named after John Bell who derived the first ones. In this chapter we derive the more transparent Clauser-Horne-Shimony-Holt (CHSH) inequality.

[2] called the control bit

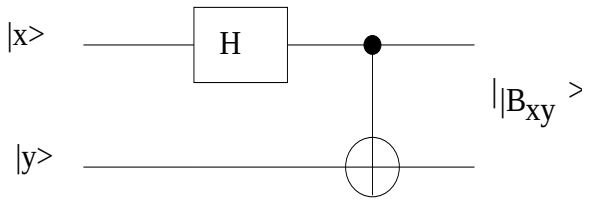


Figure 4.1 Quantum circuit producing Bell states.

The matrices H and I are the usual 2×2 Hadamard and identity matrices. The circuit representation of the unitary gate $U = (CNOT)(H \otimes I)$ is depicted in figure 1.

Let us calculate the action of this circuit on a tensor product state $|x\rangle \otimes |y\rangle = |x, y\rangle$.

$$\begin{aligned}(CNOT)(H \otimes I)|x, y\rangle &= (CNOT)\frac{1}{\sqrt{2}}\big(|0\rangle + (-1)^x|1\rangle\big) \otimes |y\rangle \\ &= \frac{1}{\sqrt{2}}CNOT|0, y\rangle + \frac{(-1)^x}{\sqrt{2}}CNOT|1, y\rangle \\ &= \frac{1}{\sqrt{2}}|0, y\rangle + \frac{(-1)^x}{\sqrt{2}}|1, y \oplus 1\rangle) \\ &= |B_{xy}\rangle\end{aligned}$$

More explicitly we have

$$|B_{00}\rangle = \frac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big) = U|00\rangle$$

$$|B_{01}\rangle = \frac{1}{\sqrt{2}}\big(|01\rangle + |10\rangle\big) = U|01\rangle$$

$$|B_{10}\rangle = \frac{1}{\sqrt{2}}\big(|00\rangle - |11\rangle\big) = U|10\rangle$$

$$|B_{11}\rangle = \frac{1}{\sqrt{2}}\big(|01\rangle - |10\rangle\big) = U|11\rangle$$

These four states are a unitary "rotation" of the four canonical basis states of $C^2 \otimes C^2$ and thus also form a basis, called the Bell basis.

Here the interaction is effected by the $CNOT$ gate: building such a gate in a laboratory requires bringing two particles supporting the Qbits $|x\rangle$ and $|y\rangle$ close enough in space and time (interactions are local). Photons do not interact directly with one another (Maxwell equations are linear) but they can interact indirectly through their direct interaction with matter (one speaks of non-linear optics). Localized sources producing pairs of entangled photons are excited atoms or nuclei, emitting photons when they fall in their ground state. Electron spin can also be entangled because the combination of the Coulomb interaction with

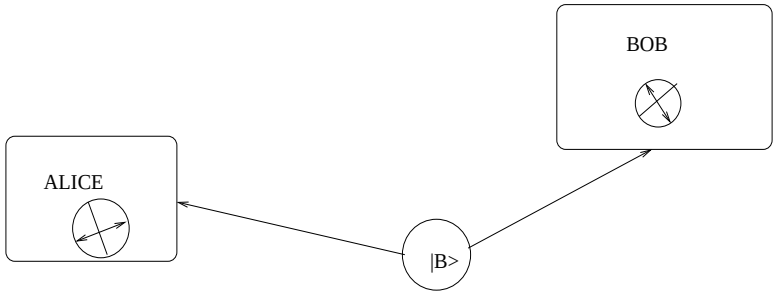


Figure 4.2 Alice and Bob share an entangled pair.

the Pauli principle can produce special magnetic correlations. In fact this kind of entanglement is very common place: in a hydrogen molecule the spin part of the chemical valence bond*[3] between two hydrogen atoms is the state

$$|B_{11}\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$$

The reader should check that

$$|B_{00}\rangle = \frac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big) = \frac{1}{\sqrt{2}}\big(|\gamma\gamma\rangle + |\gamma_\perp\gamma_\perp\rangle\big) \tag{4.2}$$

where $|\gamma\rangle$ is any state of $\mathbf{C}^2$. This has remarkable consequences as the following discussion will show. For the sake of the argument we suppose that Alice has captured one photon in her lab and Bob has captured the other photon in his lab (figure 2). Irrespective how remote the two labs are, it is always true that the two photons have come from a common localized source. Now we look at the outcome of several simple measurements that Alice and Bob might do each in their own lab. *We are assuming that they cannot communicate the outcomes of these measurements.* We will consider the three specific situations where: Alice measures first/Bob measures after; Bob measures first/Alice measures after; Alice and Bob measure simultaneously[4].

- *Alice measures first and Bob after.* The "measurement apparatus" of Alice is formed by the projectors $\{|\alpha\rangle\langle\alpha|\otimes I, |\alpha_\perp\rangle\langle\alpha_\perp|\otimes I\}$ so that, according to the measurement postulate, the Bell state collapses on one of the projections (remember we have to normalize after projecting)

$$|\alpha\rangle\langle\alpha|\otimes I|B_{00}\rangle = \frac{1}{\sqrt{2}}|\alpha\alpha\rangle \to |\alpha\rangle\otimes|\alpha\rangle, \qquad \text{with prob } \frac{1}{2}$$

$$|\alpha_\perp\rangle\langle\alpha_\perp|\otimes I|B_{00}\rangle = \frac{1}{\sqrt{2}}|\alpha_\perp\alpha_\perp\rangle \to |\alpha_\perp\rangle\otimes\alpha_\perp\rangle, \qquad \text{with prob } \frac{1}{2}$$

[3] the anti-symmetry of the spin part allows the orbital part to be in the symmetric energetically favorable state (Heitler-London theory)

[4] For definiteness we have in mind a Galilean picture of space-time. However the discussion is essentially the same for a relativistic picture for space-time.

Therefore Alice observes *her photon* in the collapsed state $|\alpha\rangle$ or $|\alpha_\perp\rangle$. Bob, on his side, does not know anything, and doesn't even know that Alice has performed measurements ! In order for him to learn something he can try to perform a measurement on his photon. But he has to choose a basis $\{|\beta\rangle, |\beta_\perp\rangle\}$. Given that his photon is in the state $|\alpha\rangle$, his photon collapses to $|\beta\rangle$ with prob $\cos^2(\alpha - \beta)$ or to $|\beta_\perp\rangle$ with prob $\sin^2(\alpha - \beta)$. Similarly, given that his photon is in the state $\alpha_\perp$ we get the same result with $\cos^2$ and $\sin^2$ interchanged. The fact that Bob does not know the initial state of his photon or that he does not even known what Alice has done should not bother you: the point is that he does a specific experiment (measurement in the $\beta, \beta_\perp$ basis) and finds a net outcome. The net outcome in Bob's lab is that the photon is in the state $|\beta\rangle$ with prob $\frac{1}{2}$ or $|\beta_\perp\rangle$ with prob $\frac{1}{2}$.

- *Bob measures first and Alice after.* The same discussion shows that, if Bob performs measurements first (in the $\beta, \beta_\perp$ basis) while Alice sleeps and Alice measures after (in the $(\alpha, \alpha_\perp$ basis) the net outcome of each party is the same.
- *Bob and Alice measure simultaneously.* You might think (?) that if both parties perform *simultaneous* local measurements the whole scenario is different. Let us try. Suppose Alice and Bob perform simultaneous measurements in the basis

$$\{|\alpha, \beta\rangle, |\alpha, \beta_\perp\rangle, |\alpha_\perp, \beta\rangle, |\alpha_\perp, \beta_\perp\rangle\}$$

The Bell state

$$|B_{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = \frac{1}{\sqrt{2}}(|\gamma\gamma\rangle + |\gamma_\perp\gamma_\perp\rangle)$$

will collapse to one of the four basis states. So Alice will be in possession of a photon in state $|\alpha\rangle$ or $|\alpha_\perp\rangle$ and Bob in possession of a photon in the state $|\beta\rangle$ or $|\beta_\perp\rangle$. The situation is exactly the same than in the previous situations ! It is very instructive to compute the probabilities of the respective collapsed states (which are nothing else than the basis states). One finds that these are[5]

$$\frac{1}{2}\cos^2(\alpha - \beta),\ \frac{1}{2}\sin^2(\alpha - \beta),\ \frac{1}{2}\sin^2(\alpha - \beta),\ \frac{1}{2}\cos^2(\alpha - \beta)$$

Alice finds that the probability of her outcomes $|\alpha\rangle$ (resp $|\alpha_\perp\rangle$)

$$\frac{1}{2}\cos^2(\alpha - \beta) + \frac{1}{2}\sin^2(\alpha - \beta) = \frac{1}{2}$$

(for both cases) as in the previous scenarios; and the same holds true for Bob. Therefore the conclusions that Alice and Bob infer from their simultaneous local measurements are the same than in the non-simultaneous cases above.

[5] fortunately independent of γ

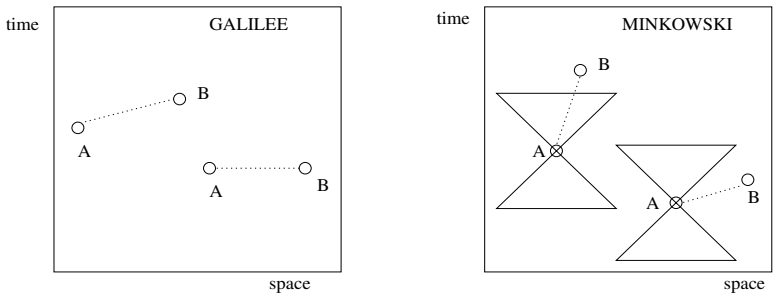


Figure 4.3 Galilean space-time: B is in the future of A (left); B and A are simultaneous (right). Minkowski space-time: the slopes of the light-cones equal $1/c$ (in natural relativistic units the speed of light is $c = 1$ as on the picture). B is in the future light cone of A (left); B and A are space-like separated and cannot be causally related. Note that for $c \to +\infty$ the slope of the light cone vanishes and one recovers the Galilean picture.

To summarize the situation, we see that when Alice and/or Bob perform successive or simultaneous local measurements on their photons, whatever is their choice of basis they find the photon in one of the two chosen basis states with probability $\frac{1}{2}$. In other words the entropy of the probability distribution of their local outcomes is maximal (it equals $\ln 2$ bits) and they may infer that their photon is in a "maximally disordered state". In fact if they don't know that the source produced an entangled pair or if nobody tells them that the two photons are entangled they have no way of even noticing that the pair is entangled. It seems that we have no way of knowing if we are entangled to some distant parts in the universe, just by performing local experiments in our part of the universe. We will see in the next section that Alice and Bob can assert that their photons are entangled if they are allowed to communicate. Here by communicate we mean the perfect or approximate transmission of a message.

Let us also point out that here we have discussed the situation having in mind a Galilean picture of space-time. In other words the meaning of the words "before", "simultaneous" and "after" is the "usual" one. However this is only an approximation and one might question if a proper account of Minkowskian space-time (see figure 4.3) would change our conclusions. According to the theory of special relativity these words are relative to each observer's frame of reference. What has an absolute meaning is the space-time interval which may be space-like, time-like (or zero). If the local measurement events (events are points in space-time) of Alice and Bob are separated by a space-like vector there cannot possibly be a causal connection between the events, and in particular it is guaranteed that Alice and Bob cannot establish a classical communication link during the experiment. On the other hand if the measurement events are separated by a time-like vector it is conceivable that there is a causal connection between the events, however unless Alice and Bob set up such a communication link, there

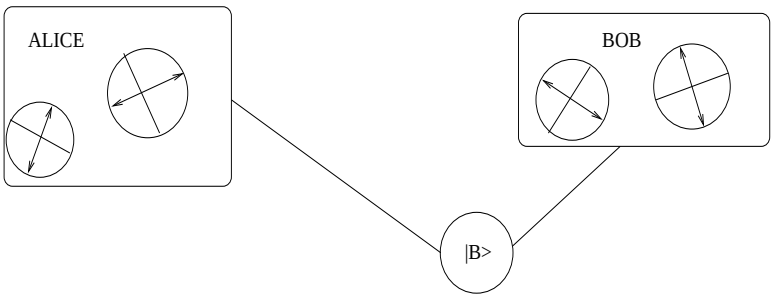


Figure 4.4 Experimental set up

is no reason to believe that there is a causal connection between the outcomes since they are exactly the same as in the case of space-like separation.

4.2 Bell inequalities and Aspect experiment

We saw in the last section that if there is no communication between Alice and Bob they can only infer that the photons are in a maximally disordered state. In this section we will see that by doing repeated measurements and by communicating the results afterwards, Alice and Bob can assert if the state produced by the source is entangled or not.

The procedure that we are going to describe was initially invented by John Bell, and motivated by a famous paper of Einstein-Podolsky-Rosen. The later claimed that the entangled states do not provide a "complete" description of the correlations present in the system, and where seeking a "classical" theory of these correlations. Bell's approach to the problem is to try to decide if the correlations in a real pair of entangled photons (produced by an excited atomic source say) *can be described* or *cannot be described* by *a classical theory.* The general idea is that if a pair of photons is described by a classical theory then appropriate correlation functions of the measurements of Alice and Bob satisfy very special constraints. These constraints are violated if the pair is described by a quantum mechanical Bell state. We will see that Bell's approach is able to discriminate between a huge set of classical theories and QM. Famous experiments of Aspect-Grangier-Roger have shown that standard QM wins!

The experimental protocol. A source S produces, at each instant of time n, a pair of photons. We do not have any prejudice as to what is the state or the description of the pair. One photon flies to Alice's lab and the other flies to Bob's lab. In each lab our two protagonists operate independently: they do not communicate and do not care what the other one does.

- At each time instant n, Alice randomly uses analyzers

$$\{|\alpha\rangle, |\alpha_\perp\rangle\} \qquad \text{or} \qquad \{|\alpha'\rangle, |\alpha'_\perp\rangle\}$$

to measure the polarization of her photon. When she records a click in the detector she sets $a_n = +1$ or $a'_n = +1$ and when the detector does not click she sets $a_n = -1$ or $a'_n = -1$. She keeps track of her choices for the analyzer at each n.

- At each time instant n, Bob randomly uses analyzers

$$\{|\beta\rangle, |\beta_\perp\rangle\} \qquad \text{or} \qquad \{|\beta'\rangle, |\beta'_\perp\rangle\}$$

to measure the polarization of his photon. When he records a click in the detector he sets $b_n = +1$ or $b'_n = 1$ and when the detector does not click he sets $b_n = -1$ or $b'_n = -1$. He keeps track of his choices of analyzers for each n.

- Now there is a classical communication phase. Alice and Bob meet and discuss all their measurements. They classify them according to the four experimental setups. At each time instant n the possible arrangements of analyzers were

$$1 = (\alpha, \beta),\ 2 = (\alpha, \beta'),\ 3 = (\alpha', \beta),\ 4 = (\alpha', \beta')$$

For each arrangement they compute the following empirical averages

$$\frac{1}{N_1}\sum_{n_1} a_{n_1} b_{n_1},\ \frac{1}{N_2}\sum_{n_2} a_{n_2} b'_{n_2},\ \frac{1}{N_3}\sum_{n_1} a'_{n_3} b_{n_3},\ \frac{1}{N_4}\sum_{n_4} a'_{n_4} b'_{n_4}$$

Then they compute the following correlation function

$$X_{exp} = \frac{1}{N_1}\sum_{n_1} a_{n_1} b_{n_1} + \frac{1}{N_2}\sum_{n_2} a_{n_2} b'_{n_2} - \frac{1}{N_3}\sum_{n_1} a'_{n_3} b_{n_3} + \frac{1}{N_4}\sum_{n_4} a'_{n_4} b'_{n_4}$$

Prediction of classical theories. We assume that the quantities that Alice and Bob measure correspond to well defined observables A, A', B, B' that have simultaneous definite values a, a', b, b' independently of any measurement. Basically, this is analogous to saying that a particle has a definite position *and* velocity (here analogous to *a and* a') even when these quantities are not observed or measured. Furthermore, we assume that the outcomes of Alice and Bob can be modeled by a joint probability distribution[6]

$$P_{class}(a, a', b, b')$$

Here by a, b, a' and b' we mean the random variables modeling the measurement outcomes. The expectation with respect to P_{class} is denoted by $\mathbf{E}_{class}$. The corresponding theoretical prediction for *each* empirical average above is

$$\mathbf{E}_{class}[ab],\ \mathbf{E}_{class}[ab'],\ \mathbf{E}_{class}[a'b],\ \mathbf{E}_{class}(a'b')$$

and using *only the linearity of expectation*

$$X_{class} = \mathbf{E}_{class}[ab + ab' - a'b + a'b']$$

[6] This assumption follows from "local realism" as explained in the next paragraph.

Notice that

$$ab + ab' - a'b + a'b' = a(b + b') + a'(b' - b)$$

and that

$$-2 \leq a(b + b') + a'(b' - b) \leq 2$$

Indeed if $b = b'$ then only the first term survives which leads to the inequality; while if $b \neq b'$ only the second term survives which again leads to the inequality. Thus we have for the expectation,

$$-2 \leq X_{class} \leq 2$$

This is one of the simplest Bell type inequalities which was derived by Clauser-Horne-Shimony-Holt and is called the CHSH inequality.

In order to derive this result we haven't assumed anything about the state of preparation of the source. We have only assumed that the experimental results can be cast into a joint probability distribution. In fact this is not a priory so obvious. There are four experimental arrangements so that when Alice and Bob meet they have four histograms that can be fitted to 4 probability distributions:

$$P_1(a, b),\ P_2(a', b),\ P_3(a, b'),\ P_4(a', b')$$

Are these the marginals of a common $P_{class}(a, a', b, b')$? It is not a priory clear that, in this experiment, nature gives us histograms that are marginals of a common joint distribution. In fact this is *not* always the case. Indeed any of us can construct four probability distributions that are not marginals of a common one, and this is an outcome of our brains (viewed as a physical systems). So why is the assumption leading to the CHSH inequality very reasonable ? We answer this question below, but do not attempt to provide the most general argument.

Let us admit that the laws of physics are "local". By this we mean that when Alice (resp. Bob) perform measurements that are space-like separated Alice's experimental outcomes (resp. Bob's) depend only on her own local choice of analyzers. As far as we know, this is an assumption that underlies all the known (i.e. experimentally verified) fundamental laws of physics.

Furthermore let us suppose, following our classical intuition, or following Einstein, that the outcomes of experiments should be well defined preexisting functions of the system state and the experimental set-up. This is sometimes called "realism".

Moreover the results of the measurement in Alice's and Bob's labs should only depend on the "local" set-up of the experiment. In particular the measurements of alice and Bob do not influence each other.

In mathematical terms "local realism" means that there should be a function, such that

$$a = f_{\mathcal{A}}(\alpha; \lambda),\ a' = f_{\mathcal{A}}(\alpha'; \lambda),\ b = f_{\mathcal{B}}(\beta; \lambda),\ b' = f_{\mathcal{B}}(\beta'; \lambda)$$

Here λ is a set of variables accounting for the state of the system and whatever is

needed to compute the experimental outcome. It has become customary to call them "hidden variables".

More generaly one can deal with random measurement results modelled by probability distributions: $p_{\mathcal{A}}(a|\alpha,\lambda)$, $p_{\mathcal{A}}(a'|\alpha',\lambda)$, $p_{\mathcal{B}}(b|\beta,\lambda)$, $p_{\mathcal{B}}(b'|\beta',\lambda)$. The previous model corresponds to have $p_{\mathcal{A}}(a|\alpha,\lambda) = \delta(a - f_{\mathcal{A}}(\alpha,\lambda)$. The conclusions are however the same (this is interesting because it shows that it is not determinism that is at stake here).

The hidden variables may be random or deterministic[7] and their set of values is described by a probability distribution $h(\lambda)$. According to "local realism" the histograms of Alice and Bob are modeled by

$$P_1(a,b) = \int d\lambda h(\lambda) p_{\mathcal{A}}(a|\alpha,\lambda) p_{\mathcal{B}}(b|\beta,\lambda)$$

$$P_2(a,b') = \int d\lambda h(\lambda) p_{\mathcal{A}}(a|\alpha,\lambda) p_{\mathcal{B}}(b'|\beta',\lambda)$$

$$P_3(a',b) = \int d\lambda h(\lambda) p_{\mathcal{A}}(a'|\alpha',\lambda) p_{\mathcal{B}}(b|\beta,\lambda)$$

$$P_4(a',b') = \int d\lambda h(\lambda) p_{\mathcal{A}}(a'|\alpha',\lambda) p_{\mathcal{B}}(b'|\beta',\lambda)$$

Evidently these are the marginals of a joint probability distribution

$$P_{class}(a,a',b,b') = \int d\lambda h(\lambda) p_{\mathcal{A}}(a|\alpha,\lambda) p_{\mathcal{A}}(a'|\alpha',\lambda) p_{\mathcal{B}}(b|\beta,\lambda) p_{\mathcal{B}}(b'|\beta',\lambda)$$

Prediction of QM for a Bell state. First of all we notice that according to the quantum formalism the measurements of Alice and Bob are measurements of the 4 observables (hermitian matrices)

$$A = (+1)|\alpha\rangle\langle\alpha| + (-1)|\alpha_\perp\rangle\langle\alpha_\perp|, \qquad A' = (+1)|\alpha'\rangle\langle\alpha'| + (-1)|\alpha'_\perp\rangle\langle\alpha'_\perp|$$

and

$$B = (+1)|\beta\rangle\langle\beta| + (-1)|\beta_\perp\rangle\langle\beta_\perp|, \qquad B' = (+1)|\beta'\rangle\langle\beta'| + (-1)|\beta'_\perp\rangle\langle\beta'_\perp|$$

At each time instant n the state of the photon pair is described by some ket $|\Psi\rangle \in \mathcal{C}^2 \otimes \mathcal{C}^2$. The quantum mechanical prediction for the four empirical averages of Alice and Bob is

$$\langle\Psi|A\otimes B|\Psi\rangle,\ \langle\Psi|A\otimes B'|\Psi\rangle,\ \langle\Psi|A'\otimes B|\Psi\rangle,\ \langle\Psi|A'\otimes B'|\Psi\rangle$$

and for the correlation function

$$X_{QM} = \langle\Psi|A\otimes B|\Psi\rangle + \langle\Psi|A\otimes B'|\Psi\rangle - \langle\Psi|A'\otimes B|\Psi\rangle + \langle\Psi|A'\otimes B'|\Psi\rangle$$

Now let us compute this quantity for the Bell state

$$|\Psi\rangle = |B_{00}\rangle$$

[7] in this case the distribution is simply a Dirac $h(\lambda) = \delta(\lambda - \lambda_0)$

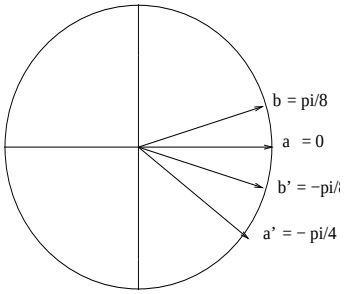


Figure 4.5 Optimal choice of analyzer orientation

The first average is best computed by expressing the Bell state as $\frac{1}{\sqrt{2}}(|\alpha\alpha\rangle + |\alpha_\perp\alpha_\perp\rangle)$.

$$\begin{aligned}
\langle B_{00}|A\otimes B|B_{00}\rangle &= \frac{1}{2}\langle\alpha\alpha|A\otimes B|\alpha\alpha\rangle + \frac{1}{2}\langle\alpha_\perp\alpha_\perp|A\otimes B|\alpha_\perp\alpha_\perp\rangle \\
&+ \frac{1}{2}\langle\alpha\alpha|A\otimes B|\alpha_\perp\alpha_\perp\rangle + \frac{1}{2}\langle\alpha_\perp\alpha_\perp|A\otimes B|\alpha\alpha\rangle \\
&= \frac{1}{2}\langle\alpha|A|\alpha\rangle\langle\alpha|B|\alpha\rangle + \frac{1}{2}\langle\alpha_\perp|A|\alpha_\perp\rangle\langle\alpha_\perp|B|\alpha_\perp\rangle \\
&= \frac{1}{2}\cdot 1\cdot\big(|\langle\alpha|\beta\rangle|^2 - |\langle\alpha|\beta_\perp\rangle|^2\big) + \frac{1}{2}\cdot(-1)\cdot\big(|\langle\alpha_\perp|\beta\rangle|^2 - |\langle\alpha_\perp|\beta_\perp\rangle|^2\big) \\
&= \frac{1}{2}\big(\cos^2(\alpha-\beta) - \sin^2(\alpha-\beta)\big) - \frac{1}{2}\big(\sin^2(\alpha-\beta) - \cos^2(\alpha-\beta)\big) \\
&= \cos^2(\alpha-\beta) - \sin^2(\alpha-\beta) = \cos 2(\alpha-\beta)
\end{aligned}$$

Performing similar calculations for the other averages we find

$$X_{QM} = \cos 2(\alpha-\beta) + \cos 2(\alpha-\beta') - \cos 2(\alpha'-\beta) + \cos 2(\alpha'-\beta')$$

This quantity is maximized for the following choice of angles (and all global rotations of this choice of course, figure 4),

$$\alpha = 0,\ \alpha' = -\frac{\pi}{4},\ \beta = \frac{\pi}{8},\ \beta' = -\frac{\pi}{8}$$

and equals

$$X_{QM} = \cos\frac{\pi}{4} + \cos\frac{\pi}{4} - \cos\frac{3\pi}{4} + \cos\frac{\pi}{4} = 2\sqrt{2}$$

We see that the CHSH inequality is violated ! For the three other Bell states on finds the same result. In the exercises you will show that this is the maximum possible violation over all quantum states of $C^2\otimes C^2$. In this sense the Bell states are maximally entangled.

QM predicts that the four histograms of Bob and Alice are

$$P_1(a,b) = \frac{1}{4}(1 + ab\cos 2(\alpha - \beta))$$
$$P_2(a,b') = \frac{1}{4}(1 + ab'\cos 2(\alpha - \beta'))$$
$$P_3(a',b) = \frac{1}{4}(1 + a'b\cos 2(\alpha' - \beta))$$
$$P_4(a',b') = \frac{1}{4}(1 + a'b'\cos 2(\alpha' - \beta'))$$

For example: $P_2(+1,-1) = |\langle \alpha, \beta'_\perp | B_{00}\rangle|^2 = \frac{1}{4}(1 - \cos 2(\alpha - \beta))$. There are special choices of the angles $\alpha, \beta, \alpha', \beta'$ for which these *are not the marginals of a common distribution* $P_{class}(a,b,a',b')$ otherwise we would have $|X| \leq 2$: this is just a mathematical fact[8]. Now, nature produces these four histograms in an experiment satisfying locality in the sense that all analyzer choices of Alice and Bob are independent. But she plays a very subtle magic trick with us: *the correlations that are built up in Bell's states are non-local* in the sense that correlations are present in the measurement outcomes even though the measurements on the photons are purely local. Alice and Bob cannot notice these non local correlations by purely local means in their own lab. They have to meet or to communicate by exchanging matter.

Experiments. In a famous set of experiments performed in the 80's Aspect-Grangier-Roger showed that experiment agrees with QM and not with classical theories. The difficulty of these experiments is that, one wants to rotate the analyzers of Alice and Bob fast enough so that the measurement events are separated by a space-like interval. Otherwise, one may always argue that some form of classical communication or interaction conspires to make up the results (on speaks of locality loophole). This is the challenge that the Aspect experiments were the first to address, as compared with other slightly earlier experiments. This locality loophole has been since then conclusively settled by more recent experiments[9]. There are other issues, that one has to address in principle, such as the efficiency of coincident detections (called detection loophole). So far there are no experiments that completly address all loopholes at the same time.

The Aspect experiments tell us that we have to abandon the "local realism". More concretly we have to abndon the assumption that there exist a joint distribution $P_{class}(a,b,a',b')$ describing the outcomes of all measurements.

[8] In some sense they are the marginals of a quantum state

[9] see the review by Anton Zeilinger "Experiment and the foundations of quantum physics", in Reviews of Modern Physics **71**, S288-S297 (1999)

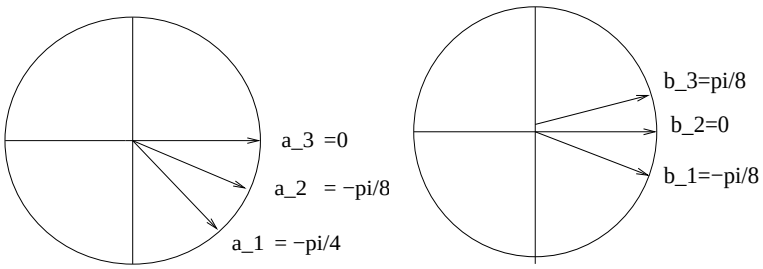


Figure 4.6 Alice and Bob's random choices of analyzers

4.3 Ekert protocol for QKD

A nice application of the CHSH inequality is a protocol for the generation of a secret key by two parties. We assume that a localized source of EPR particles delivers entangled Qbits to Alice and Bob at each time instant n in the state

$$|B_{00}\rangle = \frac{1}{2}(|00\rangle + |11\rangle) = \frac{1}{2}(|\theta\theta\rangle + |\theta_\perp\theta_\perp\rangle)$$

Moreover they have also established a noiseless communication channel.

The protocol:

- Alice has analyzers oriented in directions $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ and records the results of measurements, at each time instant, for the observables

  $$A(\mathbf{a}) = (+1)|\mathbf{a}\rangle\langle\mathbf{a}| + (-1)|\mathbf{a}_\perp\rangle\langle\mathbf{a}_\perp|$$

  where she chooses $\mathbf{a}$ randomly among $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ (figure 5).
- Bob has three analyzers oriented along $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ and records the results of measurements, at each time instant, for the observables

  $$B(\mathbf{b}) = (+1)|\mathbf{b}\rangle\langle\mathbf{b}| + (-1)|\mathbf{b}_\perp\rangle\langle\mathbf{b}_\perp|$$

  where he chooses $\mathbf{b}$ randomly among $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ (figure 5).
- Alice and Bob start a public discussion over the communication channel: they inform each other on what vectors they used at each time instant.
- They do a security check to ensure that no eavesdropper is present. Alice and Bob select the time instants when the basis choices were

  $$(\mathbf{a}_3, \mathbf{b}_3),\ (\mathbf{a}_3, \mathbf{b}_1),\ (\mathbf{a}_1, \mathbf{b}_1),\ (\mathbf{a}_1, \mathbf{b}_3)$$

  Note that these are the same four analyzer arrangements used for the Bell inequalities (figure 6). For such configurations and only for such ones they exchange their measurement results. Each party computes an empirical correlation coefficient

  $$X_{\text{exp}} = Av[a_n(\mathbf{a}_3)b_n(\mathbf{b}_3)] + Av[a_n(\mathbf{a}_3)b_n(\mathbf{b}_1)] - Av[a_n(\mathbf{a}_1)b_n(\mathbf{b}_3)] + Av[a_n(\mathbf{a}_1)b_n(\mathbf{b}_1)]$$

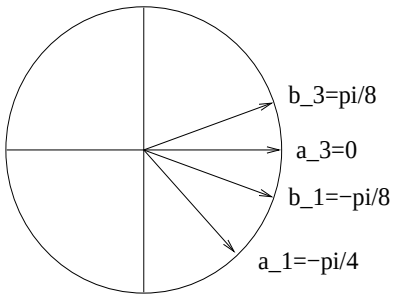


Figure 4.7 CHSH configuration

where Av is the empirical average. In a perfect world they should find $X_{exp} = 2\sqrt{2}$. We will see later that when an eavesdropper is present they will certainly find $X_{exp} \leq 2$ because the effect of the eavesdropper is to destroy the entanglement of the EPR pair and the system then behaves "classically". The security check thus consists in checking that

$$X_{exp} > 2$$

If the test passes they conclude there is no eavesdropper and generate the key, if not they stop communication.

- The key generation process is as follows. For every time n such that they used the same basis - that is $(\mathbf{a}_3, \mathbf{b}_2)$ or $(\mathbf{a}_2, \mathbf{b}_1)$ - they know for sure that

$$a_n = b_n = 1, \qquad \text{or} \qquad a_n = b_n = -1$$

(one can also check that in this case $\langle B_{00}|A \otimes B|B_{00}\rangle = \cos 2(\hat{\mathbf{a}, \mathbf{b}}) = 1$). Thus they have a common subsequence of ± 1's that they keep secret and forms their shared secret key.

Attacks from Eve. Let us consider the simplest measurement attack in which Eve captures each photon of the EPR pair and makes a measurement (figure 7). Then she sends each photon (in the resulting state) to Alice and Bob. She measures Alice's photon in the basis $\{\mathbf{e}_a, \mathbf{e}_a^\perp\}$ and Bob's photon in the basis $\{\mathbf{e}_b, \mathbf{e}_b^\perp\}$. Her strategy for the successive choices of basis at each time instant is described by a probability distribution

$$\rho(\mathbf{e}_a, \mathbf{e}_b) \geq 0, \qquad \int\int d^2\mathbf{e}_a d^2\mathbf{e}_b \, \rho(\mathbf{e}_a, \mathbf{e}_b) = 1$$

After Eve's measurement the pair of photons is left in one of the four tensor product states

$$|\mathbf{e}_a, \mathbf{e}_b\rangle, \ |\mathbf{e}_a, \mathbf{e}_b^\perp\rangle, \ |\mathbf{e}_a^\perp, \mathbf{e}_b\rangle, \ |\mathbf{e}_a^\perp, \mathbf{e}_b^\perp\rangle$$

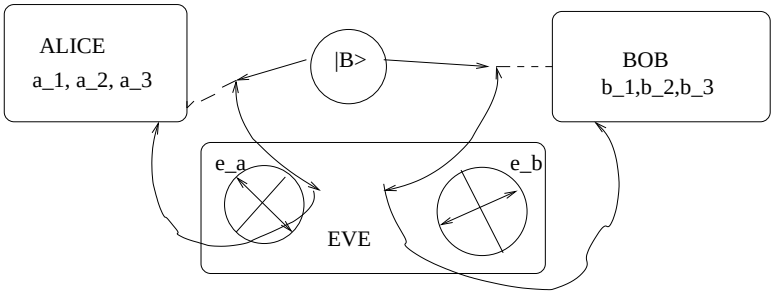


Figure 4.8 Eve collapses the pair in a tensor product state

with corresponding probabilities

$$|\langle \mathbf{e}_a, \mathbf{e}_b | B_{00}\rangle|^2 = \frac{1}{2}\cos^2(\widehat{\mathbf{e}_a, \mathbf{e}_b}), \qquad |\langle \mathbf{e}_a, \mathbf{e}_b^\perp | B_{00}\rangle|^2 = \frac{1}{2}\sin^2(\widehat{\mathbf{e}_a, \mathbf{e}_b})$$

$$|\langle \mathbf{e}_a^\perp, \mathbf{e}_b | B_{00}\rangle|^2 = \frac{1}{2}\sin^2(\widehat{\mathbf{e}_a, \mathbf{e}_b}), \qquad |\langle \mathbf{e}_a^\perp, \mathbf{e}_b^\perp | B_{00}\rangle|^2 = \frac{1}{2}\cos^2(\widehat{\mathbf{e}_a, \mathbf{e}_b})$$

Let us compute the correlation coefficient that Alice and Bob would find during the security test. Given Eve's choice $(\mathbf{e}_a, \mathbf{e}_b)$ we have

$$\begin{aligned} X(\mathbf{e}_a, \mathbf{e}_b) \; &= \tfrac{1}{2}\cos^2(\widehat{\mathbf{e}_a, \mathbf{e}_b}) S(\mathbf{e}_a, \mathbf{e}_b) + \tfrac{1}{2}\sin^2(\widehat{\mathbf{e}_a, \mathbf{e}_b^\perp}) S(\mathbf{e}_a, \mathbf{e}_b^\perp) \\ &+ \tfrac{1}{2}\sin^2(\widehat{\mathbf{e}_a^\perp, \mathbf{e}_b}) S(\mathbf{e}_a^\perp, \mathbf{e}_b) + \tfrac{1}{2}\cos^2(\widehat{\mathbf{e}_a^\perp, \mathbf{e}_b^\perp}) S(\mathbf{e}_a^\perp, \mathbf{e}_b^\perp) \end{aligned}$$

where $S(\mathbf{v}, \mathbf{w})$ is the correlation coefficient for a pair of photons in the state $|\mathbf{v}, \mathbf{w}\rangle$ resulting from Eve's measurement,

$$S(\mathbf{v}, \mathbf{w}) = \langle \mathbf{v}, \mathbf{w} | A(\mathbf{a}_3) \otimes B(\mathbf{b}_3) + A(\mathbf{a}_3) \otimes B(\mathbf{b}_1) - A(\mathbf{a}_1) \otimes B(\mathbf{b}_3) + A(\mathbf{a}_1) \otimes B(\mathbf{b}_1) | \mathbf{v}, \mathbf{w}\rangle$$

The average correlation coefficient found by Alice and Bob when Eve operates is

$$X = \int\int d^2\mathbf{e}_a d^2\mathbf{e}_b \, \rho(\mathbf{e}_a, \mathbf{e}_b) X(\mathbf{e}_a, \mathbf{e}_b)$$

We leave it as an exercise to check that $|S(\mathbf{v}, \mathbf{w})| \leq 2$. This is not too surprising since $|\mathbf{v}, \mathbf{w}\rangle$ is a tensor product state. This immediately leads to,

$$|X| \leq 2.$$

Thus Alice and Bob notice the presence of Eve. Note that Eve could manipulate (unitarily) the pair after her measurements in order to send other photon states to Alice and Bob. However if she re-entangles the photons she behaves as a new source for Alice and Bob, and she gets no information from their measurements!

Finally let us note that if Eve copies the EPR pair (this can be done with a machine that copies the four orthogonal Bell states) and waits for the public discussion before doing the measurements, she gets no information about the secret key. Indeed her measurements operate on a different pair and thus she

gets the same result than Alice and Bob only half of the time. This is equivalent to flip a coin at each time instant and cannot yield information.

Experiments. see in Review of Modern Physics **74** p 145-190 (2002) the extensive article "Quantum cryptography" by N. Gisin, G. Ribordy, W. Tittel, H. Zbinden.

4.4 Quantum teleportation

Suppose that Alice and Bob are spatially separated and that Alice possesses a Qbit state,

$$|\Phi\rangle = \alpha|0\rangle + \beta|1\rangle, \qquad |\alpha|^2 + |\beta|^2 = 1$$

The state (i.e α and β) is not necessarily known to Alice and is not known to Bob. They also share an EPR pair

$$|B_{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

and have at their disposal a classical communication channel.

We are going to explain that by sending only two classical bits of information over the classical channel, Alice can *teleport* the state to Bob. Here *teleportation* means that $|\Phi\rangle$ is destroyed in Alice's lab and is reconstructed in Bob's lab. Note that destruction of $|\Phi\rangle$ in Alice's lab is to be expected because of the no-cloning theorem. After the teleportation process, Bob knows that he possesses the state $|\Phi\rangle$ but still does not know the state itself (i.e he does not know α and β). We stress that the teleportation process involves physical transport of matter in the classical communication phase between Alice and Bob. Of course this classical communication phase cannot happen at speeds greater than that of light, so that the whole teleportation process does not violate the principles of relativity. We also note that the material support of the state (e.g. photon polarization, electron spin) $|\Phi\rangle = \alpha|0\rangle + \beta|1\rangle$ is not necessarily the same in Alice's and Bob's lab.

Teleportation can be summarized by the following "law"

$$\text{teleporting 1 Qbit} \;=\; \text{sending 2 Cbits} \;+\; \text{sharing 1 EPR pair}$$

and can be thought of, as some form of communication between Alice and Bob which share a classical channel and an "EPR like channel". The quantum state $|\Phi\rangle$ in Alice's lab is erased on her side and reproduced in Bob's lab - the information contained in α and β has not been communicated.

The protocol.

- A source produces an EPR pair of particles in the Bell state $|B_{00}\rangle_{23}$. One particle, called particle 2 is sent to Alice and one particle , called particle 3

is sent to Bob. The Hilbert space of the entangled system 23 is $\mathcal{H}_2 \otimes \mathcal{H}_3 = C^2 \otimes C^2$.

- Alice prepare a particle, called 1, in the state $|\Phi\rangle_1 = \alpha|0\rangle + |\beta\rangle$. The Hilbert space of particle 1 is $\mathcal{H}_1 = C^2$.
- The total Hilbert space of the composite system 123 is $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3 = C^2 \otimes C^2 \otimes C^2$ and the total state is

$$|\Psi\rangle = |\Phi\rangle_1 \otimes |B_{00}\rangle_{23}$$

At this point a short calculation will facilitate the subsequent discussion

$$|\Psi\rangle = \frac{\alpha}{\sqrt{2}}|000\rangle + \frac{\beta}{\sqrt{2}}|100\rangle + \frac{\alpha}{\sqrt{2}}|011\rangle + \frac{\beta}{\sqrt{2}}|111\rangle$$

- Alice makes a local measurement in her lab, i.e on partiles 12. She uses an apparatus that has measurement basis of $\mathcal{H}_1 \otimes \mathcal{H}_2$

$$\{|B_{00}\rangle_{12}, |B_{01}\rangle_{12}, |B_{10}\rangle_{12}, |B_{11}\rangle_{12}\}$$

The associated projectors for the total system are

$$P_{00} = |B_{00}\rangle\langle B_{00}|\otimes I_3, P_{01} = |B_{01}\rangle\langle B_{01}|\otimes I_3, P_{10} = |B_{10}\rangle\langle B_{10}|\otimes I_3, P_{11} = |B_{11}\rangle\langle B_{11}|\otimes I_3$$

As usual the outcome of the measurement is one of the four possible collapsed states[10] (check this calculation and also that the probability of each outcome is $\frac{1}{4}$)

$$P_{00}|\Psi\rangle = \frac{1}{2}|B_{00}\rangle_{12} \otimes (\alpha|0\rangle_3 + \beta|1\rangle_3)$$

$$P_{01}|\Psi\rangle = \frac{1}{2}|B_{01}\rangle_{12} \otimes (\beta|0\rangle_3 + \alpha|1\rangle_3)$$

$$P_{10}|\Psi\rangle = \frac{1}{2}|B_{10}\rangle_{12} \otimes (\alpha|0\rangle_3 - \beta|1\rangle_3)$$

$$P_{11}|\Psi\rangle = \frac{1}{2}|B_{11}\rangle_{12} \otimes (-\beta|0\rangle_3 - \alpha|1\rangle_3)$$

- Depending on the random outcome Bob has one of the four states

$$\begin{aligned}
\alpha|0\rangle_3 + \beta|1\rangle_3 &= |\Phi\rangle \\
\beta|0\rangle_3 + \alpha|1\rangle_3 &= X|\Phi\rangle \\
\alpha|0\rangle_3 - \beta|1\rangle_3 &= Z|\Phi\rangle \\
\beta|0\rangle_3 - \alpha|1\rangle_3 &= iY|\Phi\rangle
\end{aligned}$$

but he does not know the state he has.

[10] up to normalization

- Alice knows that the outcome of the measurement (in her lab) is one of the four Bell states. She can thus use the Bell basis to re-measure (this will not perturb Bob's particle this time) and determine her outcome. This outcome can be encoded by two classical bits

$$00, 01, 10, 11$$

that she sends to Bob over the classical communication channel. As soon as Bob receives Alice's message he knows that she has finished her operations and he has the two bits of information needed to decide which *unitary* operation he has to perform on his state in order to recover $|\Phi\rangle$,

$$\begin{aligned} I(\alpha|0\rangle_3 + \beta|1\rangle_3) &= |\Phi\rangle \\ X(\beta|0\rangle_3 + \alpha|1\rangle_3) &= |\Phi\rangle \\ Z(\alpha|0\rangle_3 - \beta|1\rangle_3) &= |\Phi\rangle \\ -iY(\beta|0\rangle_3 - \alpha|1\rangle_3) &= |\Phi\rangle \end{aligned}$$

4.5 Dense coding

Suppose Alice and Bob have established a quantum channel over which they can send Qbits (for example a optic fiber over which photons travel). We will study the capacity of such a noisy channel later in the course but for the moment let us address a simpler question. Assume that Alice and Bob share an EPR pair. How much information does one Qbit convey over the quantum channel ?

The answer is that 2 classical bits of information can be transmitted by Alice to Bob, by sending only 1 Qbit as long as they share an EPR pair. The protocol that achieves this is called *dense coding*.

We will come back to the problem of communicating classical/quantum messages over noisy quantum channels assisted/or not by entanglement in later chapters. As we will see even for simple analogs of Shannon's channel coding theorem there are various open questions.

Dense coding can be summarized as follows:

$$\text{communicating 2 Cbits} = \text{sending 1 Qbit} + \text{sharing 1 EPR pair}$$

This "law" may seem complementary to the one of teleportation. Note however that here only two particles are involved and it is the Qbit that is physicaly transported form Alice to Bob.

Protocol.

- An EPR pair in the state $|B_{00}\rangle$ is prepared by a source and each particle sent to Alice and Bob.
- Alice wants to communicate two bits of information to Bob:

  - To send 00 she leaves her particle intact (or applies the unitary gate I) and physically sends her particle to Bob. Bob receives the particle and is now in possession of the whole state
$$|B_{00}\rangle$$
  - To send 01 she applies the unitary gate X to her particle and then physically sends her particle to Bob. Bob is now in possession of the pair in the state
$$X_1 \otimes I_2|B_{00}\rangle = |B_{01}\rangle$$
  - To send 10 she applies the unitary gate Z to her particle and then physically sends her particle. Bob is now in possession of the pair in the state
$$Z_1 \otimes I_2|B_{00}\rangle = |B_{10}\rangle$$
  - To send 11 she applies the unitary gate iY to her particle and then physically sends her particle. Bob is now in possession of the pair in the state
$$(iY)_1 \otimes I_2|B_{00}\rangle = |B_{11}\rangle$$

- Bob now has the EPR pair 12 in some state $|B_{xy}\rangle$. In order to determine the two Cbits that Alice sent he must decide which Bell state he has. Since he knows that he has one of the four Bell states in his lab, he can do a local measurement in the Bell basis, and access the information xy.

Measurement in the Bell basis. One might think that measuring in the Bell basis is a theoretician's wishful thinking. In fact this has been realized experimentally, and although explaining how is beyond the scope of this course, we give here an argument that shows that, in principle, it suffices to have H and $CNOT$ gates (the simplest unitary gates) together with polarization analyzers (the simplest measurement apparatus).

We have seen at the beginning of this chapter that Bell states can be generated as $|B_{xy}\rangle = (CNOT)(H \otimes I)|xy\rangle$. The projectors on the Bell basis states are therefore related to the ones over the Z basis,

$$|B_{xy}\rangle\langle B_{xy}| = (CNOT)(H \otimes I)|xy\rangle\langle xy|(H \otimes I)(CNOT)$$

(here we have used that the Hadamard and control not matrices are hermitian). The projectors $|xy\rangle\langle xy|$ correspond to the analyzer-photo-detector apparatus for photons or to spin analyzers (Stern-Gerlach analyzer) for spins (Z basis). The circuit representation of a measurement device in the Bell basis in given on figure 8. The input is any state $|\Psi\rangle$, and the output is one of the four states

$$|B_{xy}\rangle \frac{\langle B_{xy}|\Psi\rangle}{|\langle B_{xy}|\Psi\rangle|}$$

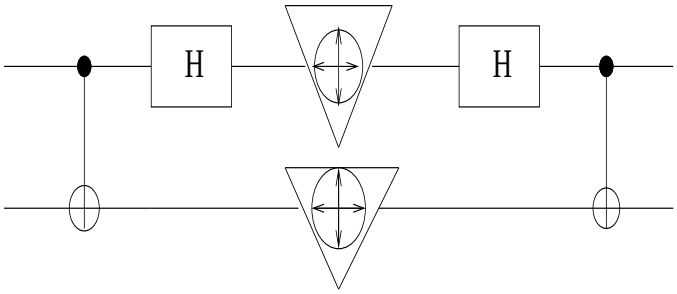


Figure 4.9 Device for Bell basis measurements

Experiments. Quantum teleportation and dense coding have been realized experimentally. A summary of the subject can be found in "Les dossiers de la recherche" no 18, février 2005, "L'étrange pouvoir de l'intrication quantique", by N. Gisin.

Part II

Quantum Information Theory

5 Density matrix formalism

In chap 2 we formulated quantum mechanics for isolated systems. In practice systems interact with their environment and we need a description that takes this feature into account. Suppose the system of interest which has Hilbert space $\mathcal{H}$ is coupled to some environment with space $\mathcal{H}_\mathcal{E}$. The total system is isolated and is described by a state vector $|\Psi\rangle \in \mathcal{H} \otimes \mathcal{H}_\mathcal{E}$. An observable for the system of interest is of the form $A \otimes I$ where A acts only in $\mathcal{H}$. We suppose that A has spectral decomposition $A = \sum_n a_n P_n$ so that

$$A \otimes I = \sum_n a_n P_n \otimes I$$

A measurement of the observable will leave the system in one of the states

$$\frac{P_n \otimes I|\Psi\rangle}{\langle\Psi|P_n \otimes I|\Psi\rangle^{1/2}}$$

with probability

$$\text{prob}(n) = \langle\Psi|P_n \otimes I|\Psi\rangle$$

and the average value of the observable is

$$\langle\Psi|A \otimes I|\Psi\rangle.$$

If we introduce the matrix[1]

$$\rho = Tr_{\mathcal{H}_\mathcal{E}}|\Psi\rangle\langle\Psi|$$

which acts on $\mathcal{H}$, we can rewrite all these formulas as follows,

$$\text{prob}(n) = TrP_n \otimes I|\Psi\rangle\langle\Psi| = Tr_\mathcal{H}Tr_\mathcal{E}P_n \otimes I|\Psi\rangle\langle\Psi| = Tr_\mathcal{H}P_n\rho$$

and

$$\langle\Psi|A \otimes I|\Psi\rangle. = TrA \otimes I|\Psi\rangle\langle\Psi| = Tr_\mathcal{H}Tr_\mathcal{E}A \otimes I|\Psi\rangle\langle\Psi| = Tr_\mathcal{H}A\rho$$

Thus we see that the system of interest is described by the matrix ρ called "density matrix". At the level of the reduced density matrix the collapse of the state vector becomes

$$\rho = Tr_\mathcal{E}|\Psi\rangle\langle\Psi| \to \rho_{\text{after}} = Tr_\mathcal{E}\frac{P_n \otimes I|\Psi\rangle\langle\Psi|P_n \otimes I}{\langle\Psi|P_n \otimes I|\Psi\rangle} = \frac{P_n\rho P_n}{TrP_n\rho}$$

[1] here a partial trace is performed. This is formally defined in a later section. Readers who are not comfortable with this paragraph can skip to the next one.

Thus a *density matrix can describe part of a system* (Landau).

There is also another kind of preparation of a quantum system for which density matrices are useful. Suppose a source emits with probability p_1 photons in state $|\Psi_1\rangle \in \mathcal{H}$ and with probability p_2 photons in state $|\Psi_2\rangle \in \mathcal{H}$ (with $p_1 + p_2 = 1$). Then the average value of an observable A acting in $\mathcal{H}$ is

$$p_1\langle\Psi_1|A|\Psi_1\rangle + p_2\langle\Psi_2|A|\Psi_2\rangle = Tr\rho A$$

where

$$\rho = p_1|\Psi_1\rangle\langle\Psi_1| + p_2|\Psi_2\rangle\langle\Psi_2|$$

This *density matrix describes a system that is prepared in an ensemble of state vectors with a definite proportion for each state vector* (von Neumann). Of course this example can be generalized to an ensemble of more than two vectors.

These two examples are sufficient motivation for introducing a slightly more general formalism, that formulates the rules of QM in terms of the density matrix. This is the subject of this chapter.

5.1 Mixed states and density matrices

Let $\mathcal{H}$ be the Hilbert space of a system of reference (isolated or not). From now on the vectors of the Hilbert space will be called *pure states*. As we remarked earlier a global phase is unobservable so that giving a pure state $|\Psi\rangle$ or its associated projector $|\Psi\rangle\langle\Psi|$ is equivalent. So a pure state can be thought of as a projector on a one dimensional subspace of $\mathcal{H}$.

A very general notion of state is as follows (von Neumann)

General definition of a state. Given a Hilbert space $\mathcal{H}$, consider $\mathcal{B}(\mathcal{H})$ the space of bounded linear self-adjoint operators from $\mathcal{H} \to \mathcal{H}$. A state is a *positive linear* functional

$$Av : \mathcal{B}(\mathcal{H}) \to C,\ A \to Av(A) \tag{5.1}$$

such that $Av(A) = 1$ (normalization condition).

A general theorem (that we do not prove here) then shows that it is always possible to represent this functional by a positive self-adjoint operator ρ with $Tr\rho = 1$. That is

$$Av(A) = Tr\rho A, \qquad \rho^\dagger = \rho, \qquad \rho \geq 0, \qquad Tr\rho = 1$$

This operator is called a density matrix.

If ρ is a one dimensional projector[2] it is said to be a *pure state*, while if it is not a projector, i.e. $\rho^2 \neq \rho$ it is said to be a *mixed state*.

[2] to check this it enough to have $\rho^2 = \rho$ because then it is a projector so its eigenvalues are 1 and 0; so if we already know that $Tr\rho = 1$ the multiplicity of 1 is one so its a one-dimensional projector

Examples.

- A pure state $\rho = |\Psi\rangle\langle\Psi|$.
- A mixture of pure states - *not necessarily orthogonal* - $\rho = \sum_n \lambda_n |\phi_n\rangle\langle\phi_n|$, $\lambda_n \geq 0$, $\sum_n \lambda_n = 1$.

There are two kind of physical interpretations of ρ that we have already given in the introduction. In fact these correspond also to two mathematical facts.

First we will see at the end of the chapter that a system that is in a mixed state can always be "purified". By this we mean that one can always construct (mathematically) a bigger Hilbert space and find a pure state $|\Psi\rangle$ such that $\rho = Tr|\Psi\rangle\langle\Psi|$. Thus we may always *interpret* ρ as describing part of a bigger system (Landau).

Second, given ρ, since it is self-adjoint, positive and its trace is normalized it always has a spectral decomposition

$$\rho = \sum_i \rho_i |i\rangle\langle i|, \qquad \rho_i \geq 0, \qquad \sum_i \rho_i = 1$$

Thus we can always *interpret* ρ as describing a mixture of pure states $|i\rangle$ each state occurring in the proportion ρ_i (von Neumann). In quantum statistical mechanics for example we have $\rho_i = \frac{e^{-\beta E_i}}{Z}$, $Z = \sum_i e^{-\beta E_i}$, β the inverse temperature. Of course there are other ways (not corresponding to the spectral decomposition) of rewriting ρ as a convex combination of one dimensional projectors so there is an ambiguity in this interpretation. In quantum information theory it is important to have in mind that, given ρ, if we do not know the state preparation of the system - that is the set $\{\lambda_n, |\phi_n\rangle\}$ - there is an ambiguity in the interpretation as a mixture. We can access part of the information about the preparation by making measurements, and as we will see in the next chapter the Holevo quantity gives a bound on the mutual information between the preparation and the measurement outcomes.

Lemma 5.1.1 The set of states of a quantum system is convex. The extremal points are pure states, in other words they are one dimensional projectors $|\Psi\rangle\langle\Psi|$. Conversely the pure states are extremal points of this set.

Proof Let ρ_1 and ρ_2 be two density matrices. Then evidently any convex combination $\rho = \lambda\rho_1 + (1-\lambda)\rho_2$ for $\lambda \in [0,1]$ satisfies $\rho^\dagger = \rho$, $\rho \geq 0$ and $Tr\rho = 1$. Hence the set of density matrices is convex.

If ρ is an extremal point then it cannot be written as a non trivial linear combination of other density matrices. But all ρ have a spectral decomposition $\rho = \sum_i \rho_i |i\rangle\langle i|$ with $0 \leq \rho_i$ and $\sum_i \rho_i = 1$. Since this is a convex combination it must be trivial so only one of the ρ_i equals 1 and the other vanish: thus $\rho = |i\rangle\langle i|$ for some i.

Now let ρ be a pure state: there exits a $|\Psi\rangle$ st $\rho = |\Psi\rangle\langle\Psi|$. We want to show that it is impossible to find $\rho_1 \neq \rho_2$ and $0 < \lambda < 1$ st $\rho = \lambda\rho_1 + (1-\lambda)\rho_2$. If $P_\perp$

is the projector on the orthogonal complement of $|\Psi\rangle$,

$$0 = TrP_\perp \rho P_\perp = \lambda TrP_\perp \rho_1 P_\perp + (1-\lambda) TrP_\perp \rho_2 P_\perp$$

The positivity of ρ_1, ρ_2 and the strict positivity of λ and $1-\lambda$ imply that

$$TrP_\perp \rho_1 P_\perp = TrP_\perp \rho_2 P_\perp = 0$$

and by the positivity again we deduce

$$P_\perp \rho_{1,2} P_\perp = 0, \qquad P_\perp \rho_{1,2}^{1/2} = \rho_{1,2}^{1/2} P_\perp = 0$$

(To see this one uses that $(TrA^\dagger A)^{1/2}$ is a norm in $\mathcal{B}(\mathcal{H})$ with the choice $A = \rho^{1/2}P$; and that if the norm of a matrix is zero then the matrix itself is zero) Thus we have

$$\rho_1 = (P_\perp + |\Psi\rangle\langle\Psi|)\rho_1(P_\perp + |\Psi\rangle\langle\Psi|) = (|\Psi\rangle\langle\Psi|)\langle\Psi|\rho_1|\Psi\rangle$$

But $Tr\rho_1 = 1$ so $\langle\Psi|\rho_1|\Psi\rangle = 1$ and $\rho_1 = |\Psi\rangle\langle\Psi|$. The same argument applies to ρ_2 and thus $\rho_1 = \rho_2$. □

The density matrix of a single Qbit. The set of states of a single Qbit can easily be described in terms of 2×2 density matrices as we now show. A basis for all matrices is given by the Pauli matrices $\{I, X, Y, Z\}$,

$$\rho = a_0 I + a_1 X + a_2 Y + a_3 Z$$

We have $Tr\rho = 2a_0$ so we require that $a_0 = \frac{1}{2}$. We rewrite the density matrix as

$$\rho = \frac{1}{2}(I + \mathbf{a}\cdot\Sigma) = \frac{1}{2}\begin{pmatrix} 1+a_3 & a_1 - ia_2 \\ a_1 + ia_2 & 1-a_3 \end{pmatrix}$$

where $\mathbf{a} = (a_1, a_2, a_3)$ and $\Sigma = (X, Y, Z)$ is the vector with the three Pauli matrices as components. We need $\rho^\dagger = \rho$ so the vector $\mathbf{a}$ has real components (Pauli matrices are hermitian). In order to have also $\rho \geq 0$ we necessarily need $det\rho \geq 0$. This is also sufficient because we already have $Tr\rho = 1$ so that both eigenvalues cannot be negative and hence they are both positive. The positivity of the determinant is equivalent to

$$det\rho = 1 - |\mathbf{a}|^2 \geq 0$$

Therefore the space of 2×2 density matrices is

$$\{\rho = \frac{1}{2}(I + \mathbf{a}\cdot\Sigma) | |\mathbf{a}| \leq 1\}$$

Evidently we can identify it to the unit ball $|\mathbf{a}| \leq 1$ and is commonly called the "Bloch sphere". Of course it is convex and the extremal states are those which cannot be written as a non-trivial linear combination, that is the states with

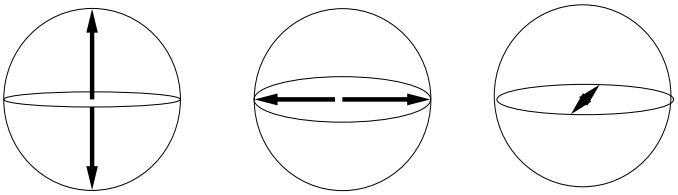


Figure 5.1 Z basis $\{|0\rangle, |1\rangle\}$, Y basis $\{\frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)\}$, X basis $\{\frac{1}{\sqrt{2}}(|0\rangle \pm i|1\rangle)\}$

$|\mathbf{a}| = 1$. Let us check that the later are pure states. We compute

$$\begin{aligned}\rho^2 &= \frac{1}{4}(I + \mathbf{a}\cdot\Sigma)^2 \\ &= \frac{1}{4}(1 + a_1^2X^2 + a_2^2Y^2 + a_3^2Z^2) \\ &+ \frac{1}{4}a_xa_y(XY + YX) + a_xa_z(XZ + ZX) + a_ya_z(YZ + ZY) \\ &+ \frac{1}{4}2\mathbf{a}\cdot\Sigma\end{aligned}$$

The squares of Pauli matrices equal the unit matrix and they anti-commute, so

$$\rho^2 = \frac{1}{4}(1 + |\mathbf{a}|^2) + \frac{1}{2}\mathbf{a}\cdot\Sigma$$

which equal ρ iff $|\mathbf{a}|^2 = 1$.

Figure 1 shows the pure states of the three basis X, Y, Z on the Bloch sphere. For example $|\psi\rangle = |0\rangle$ corresponds to $\rho = |0\rangle\langle 0| = \frac{1}{2}(I + Z)$, i.e $\mathbf{a} = (0, 0, 1)$. For $|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ we have $\rho = |0\rangle\langle 0| + |0\rangle\langle 1| + |1\rangle\langle 0| + |1\rangle\langle 1| = \frac{1}{2}(I + X)$, i.e $\mathbf{a} = (1, 0, 0)$. General pure states can be parametrized by two angles while for mixed states one also needs the length of the vector inside the ball.

5.2 Postulates of QM revisited

We briefly give the postulates of QM in the density matrix formalism.

1. States. A quantum system is described by a Hilbert space $\mathcal{H}$. The state of the system is a density matrix ρ satisfying $\rho = \rho^\dagger$, $\rho \geq 0$ and $Tr\rho = 1$. One may also think of the state as a positive linear functional $A \in \mathcal{B}(\mathcal{H}) \to TrA\rho \in C$. These form a convex set. The extremal points are one dimensional projectors and are called pure states. Other states that are non-trivial linear combinations of one dimensional projectors are called mixed states. Any density matrix is of the form

$$\rho = \sum_n \lambda_n |\phi_n\rangle\langle\phi_n|$$

with $0 \leq \lambda_n \leq 1$ and $\sum_n \lambda_n = 1$.

2. Evolution. The dynamics of the system is given by a unitary matrix acting on the states as

$$\rho(t) = U_t \rho(0) U_t^\dagger$$

Indeed let the initial condition be $\rho(0) = \sum_n \lambda_n |\phi_n\rangle\langle\phi_n|$. At time t each state of the mixture is $U_t|\phi_n\rangle$ thus $\rho(t) = \sum_n \lambda_n U_t|\phi_n\rangle\langle\phi_n|U_t^\dagger = U_t\rho(0)U_t^\dagger$.

3. Observables. They are described by linear self-adjoint operators $A = A^\dagger$. They have a spectral decomposition $A = \sum_n \alpha_n P_n$ with real eigenvalues α_n and an orthonormal set of projectors P_n satisfying the closure or completeness relation $\sum_n P_n = 1$.

4. Measurements. The measurement of an observable A is described by the measurement basis formed by the eigenprojectors of A. When the system is prepared in state ρ the possible outcomes of the measurement are

$$\rho_{\text{after}} = \frac{P_n \rho P_n}{Tr P_n \rho P_n}$$

with probability

$$\text{Prob}(n) = Tr P_n \rho P_n$$

As we will see one can always purify the system, which means constructing a bigger system whose reduced density matrix is ρ. Applying the usual measurement postulate to the purified system leads to the above formulas (we showed this at the very beginning of the chapter).

5. Composite systems. A system composed of two (or more) parts $\mathcal{A} \cup \mathcal{B}$ has a tensor product Hilbert space $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$. A density matrix for this system is of the general form

$$\rho = \sum_n \lambda_n |\phi_n\rangle\langle\phi_n|$$

with $|\phi_n\rangle \in \mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$, $0 \leq \lambda_n \leq 1$ and $\sum_n \lambda_n = 1$. Note that $\rho = \rho_\mathcal{A} \otimes \rho_\mathcal{B}$ only if there are no correlations between the parts.

A remark about the Schroedinger and Heisenberg pictures. In the Schroedinger picture of QM the states evolve as in postulate 2 above and observables stay fixed. The average value of A at time t is given by $Tr A\rho(t)$ where $\rho(t) = U_t \rho U_t^\dagger$. The Heisenberg picture is a mathematically equivalent description where the states ρ stay fixed and the observables evolve according to $A(t) = U_t^\dagger A U_t$. In the Heisenberg picture the average is $Tr A(t)\rho$. Both pictures are equivalent because of the cyclicity of the trace.

5.3 Partial trace and Reduced density matrix

Suppose we have a composite system with Hilbert space $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$ and let it be described by a density matrix ρ. The *reduced density matrix* of $\mathcal{A}$ (resp. $\mathcal{B}$) is

$$\rho_\mathcal{A} = Tr_{\mathcal{H}_\mathcal{B}}\rho \qquad \rho_\mathcal{B} = Tr_{\mathcal{H}_\mathcal{A}}\rho$$

Here the trace is performed over $\mathcal{H}_\mathcal{B}$ only (resp. $\mathcal{H}_\mathcal{A}$ only). This is known as a partial trace and can be defined as follows

$$Tr_\mathcal{B}\underbrace{(|a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2|)}_{operator\ in\ \mathcal{H}_\mathcal{A}\otimes\mathcal{H}_\mathcal{B}} = |a_1\rangle\langle a_2|(Tr|b_1\rangle\langle b_2|) = \underbrace{(|a_1\rangle\langle a_2|)}_{operator\ in\ \mathcal{H}_\mathcal{A}}\ \underbrace{\langle b_2|b_1\rangle}_{\in C}$$

This rule combined with linearity enables one to compute all partial traces in practice. You can translate this rule for computing a partial trace in the usual matrix language but you will see that the Dirac notation is much more powerful at this point. In general if $\rho = \sum_n \lambda_n |\phi_n\rangle\langle\phi_n|$ and $|\phi_n\rangle = \sum_{i,j} a_{ij}^n |\phi_i\rangle_\mathcal{A} \otimes |\chi_j\rangle_\mathcal{B}$, we have

$$\rho = \sum_{n,i,j,k,l} \lambda_n a_{ij}^n (|\phi_i\rangle_\mathcal{A} \otimes |\chi_j\rangle_\mathcal{B})(\langle\phi_k|_\mathcal{A} \otimes \langle\chi_l|_\mathcal{B}) \tag{5.2}$$

$$= \sum_{n,i,j,k,l} \lambda_n a_{ij}^n (|\phi_i\rangle_\mathcal{A}\langle\phi_k|_\mathcal{A}) \otimes (|\chi_j\rangle_\mathcal{B}\langle\chi_l|_\mathcal{B}) \tag{5.3}$$

The partial traces are

$$\rho_\mathcal{A} = Tr_{\mathcal{H}_\mathcal{B}}\rho = \sum_{i,k}\Big(\sum_{n,j,l} \lambda_n a_{ij}^n (\langle\chi_l|\chi_j\rangle_\mathcal{B})\Big)(|\phi_i\rangle_\mathcal{A}\langle\phi_k|_\mathcal{A})$$

and

$$\rho_\mathcal{B} = Tr_{\mathcal{H}_\mathcal{A}}\rho = \sum_{j,l}\Big(\sum_{n,i,k} \lambda_n a_{ij}^n (\langle\phi_k|\phi_i\rangle_\mathcal{A})\Big)(|\chi_j\rangle_\mathcal{B}\langle\chi_l|_\mathcal{B})$$

Examples.

- The partial trace of a tensor product state is a pure state. Indeed let $|\Psi\rangle = |\phi\rangle_\mathcal{A} \otimes |\chi\rangle_\mathcal{B}$. Then one finds

$$\rho_\mathcal{A} = |\phi\rangle_\mathcal{A}\langle\phi|_\mathcal{A}, \qquad \rho_\mathcal{B} = |\chi\rangle_\mathcal{B}\langle\chi|_\mathcal{B}$$

- The partial trace of an entangled pure state is a mixed state (we prove this in full generality later). The reader should check that if $\rho = |B_{00}\rangle$ then

$$\rho_\mathcal{A} = \frac{1}{2} I_\mathcal{A}, \qquad \rho_\mathcal{B} = \frac{1}{2} I_\mathcal{B}$$

- Another instructive calculation is for $\rho = \frac{1}{2}|B_{00}\rangle\langle B_{00}| + \frac{1}{2}|01\rangle\langle 01|$,

$$\rho_\mathcal{A} = \frac{3}{4}|0\rangle_\mathcal{A}\langle 0|_\mathcal{A} + \frac{1}{4}|1\rangle_\mathcal{A}\langle 1|_\mathcal{A}, \qquad \rho_\mathcal{B} = \frac{1}{4}|0\rangle_\mathcal{B}\langle 0|_\mathcal{B} + \frac{3}{4}|1\rangle_\mathcal{B}\langle 1|_\mathcal{B}$$

  The eigenvalues of the two reduced density matrices are the same. Do you think this is a coincidence ?

Physical interpretation. The interpretation of the reduced density matrix is the same as the one discussed in the introduction to this chapter. For a composite system $\mathcal{AB}$ is in the state ρ, the RDM $\rho_{\mathcal{A}}$ describes everything that is accessible by local operations in the part $\mathcal{A}$.

In particular if we measure a local observable $A \otimes I = \sum_n \alpha_n P_n \otimes I$ according to postulate 4) the measured value of the observable is α_n, and the total state collapses to

$$\rho_{\text{after}} = \frac{(P_n \otimes I)\rho(P_n \otimes I)}{Tr(P_n \otimes I)\rho}$$

with probability

$$\text{prob}(n) = Tr(P_n \otimes I)\rho$$

Thus the average value of the observable is $\sum_n \alpha_n \text{prob}(n) = Tr(A \otimes I)\rho$. This is also equal to $TrA\rho$. Since this is true for *any* local observable, from the point of view of a local observer in $\mathcal{A}$, before the measurement the system is in state $\rho_{\mathcal{A}}$ and after it is found in the state

$$\rho_{\mathcal{A},\, after} = Tr_{\mathcal{H}_{\mathcal{B}}} \rho_{\text{after}} = \frac{P_n \rho_{\mathcal{A}} P_n}{TrP_n\rho_{\mathcal{A}}}$$

with probability

$$\text{prob}(n) = TrP_n\rho_{\mathcal{A}}$$

As an example consider the composite system formed of an EPR pair in the state state $|B_{00}\rangle$. Imagine Alice does measurements on her photons and does not communicate with Bob. From the discussions of chapter 4 we know that for any measurement basis $\{|\alpha\rangle, |\alpha_\perp\rangle\}$ (this means she measures any observable $A = \lambda_1|\alpha\rangle\langle\alpha| + \lambda_2|\alpha_\perp\rangle\langle\alpha_\perp\rangle$) she will find outcomes $\alpha\rangle$ or $\alpha_\perp$ each with probability $\frac{1}{2}$. Since this is true for *any* choice of α some thought will show that the only compatible state with the outcomes is the mixed state $\rho_{\mathcal{A}} = \frac{1}{2}I$. Within the density matrix formalism we can arrive at this result in an immediate manner. Indeed the reduced density matrix of the Bell state is indeed $\rho_{\mathcal{A}} = \frac{1}{2}I$. The physical interpretation is that if Alice and Bob share an EPR pair, then Alice (or Bob) cannot learn more than the mixed state $\frac{1}{2}I$ by local measurements. We will see that this has an interesting consequence for the notion of quantum mechanical entropy: the entropy of the composite system is zero (it is in a well defined pure state) but at the same time the entropy of its parts is maximal (it is $\ln 2$). Thus in the quantum world the entropy[3] of a system can be lower than the entropy of its parts. This is one of the effects of entanglement which violates classical inequalities such as Shannon's $H(X,Y) \geq H(X)$.

[3] we will introduce in the next chapter the von Neumann entropy which is a direct generalization of Shannon's entropy

5.4 Schmidt decomposition and purification

The Schmidt decomposition and purification are two useful tools that we will use extensively later on.

THEOREM 5.4.1 *Let $|\Psi\rangle$ be a pure state for a bipartite system with Hilbert space $\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$. then*

a) $\rho_\mathcal{A} = Tr_\mathcal{B}|\Psi\rangle\langle\Psi|$ and $\rho_\mathcal{B} = Tr_\mathcal{A}|\Psi\rangle\langle\Psi|$ have the same non-zero eigenvalues with the same multiplicities. The multiplicity of the zero eigenvalue (if present) may or may not be different. Thus the spectral decompositions of the two reduced density matrices are

$$\rho_\mathcal{A} = \sum_i \rho_i |i\rangle_\mathcal{A}\langle i|_\mathcal{A}, \qquad \rho_\mathcal{B} = \sum_i \rho_i |i\rangle_\mathcal{B}\langle i|_\mathcal{B}$$

with $\rho_i > 0$ and $\sum_i \rho_i = 1$. Note that we do not write explicitly the contribution of the zero eigenvalues since they contribute a vanishing term. Here $|i\rangle_\mathcal{A}$ are orthonormal states of $\mathcal{H}_\mathcal{A}$ and $|i\rangle_\mathcal{B}$ are other orthonormal states of $\mathcal{H}_\mathcal{B}$. Note that they do not form a complete basis unless we include also the eigenstates of the 0 eigenvalues. If the non-zero eigenvalues are not degenerate the vectors $|i\rangle_\mathcal{A}$ and $|i\rangle_\mathcal{B}$ are unique (up to a phase). Otherwise there is freedom in their choice (rotations in the ρ_i subspaces).

b) The pure state $|\Psi\rangle$ has the Schmidt decomposition

$$|\Psi\rangle = \sum_i \sqrt{\rho_i}|i\rangle_\mathcal{A} \otimes |i\rangle_\mathcal{B}$$

This expansion (with positive coefficients) is unique up to rotations in the span of ρ_i.

An immediate consequence is

Corollary 5.4.2 For any $|\Psi\rangle \in \mathcal{H}_\mathcal{A} \otimes \mathcal{H}_\mathcal{B}$ we can form $\rho = |\Psi\rangle\langle\Psi|$ and $\rho_\mathcal{A}$, ρ_B. We have

$$TrF(\rho_\mathcal{A}) = \sum_i F(\rho_i) + g_\mathcal{A}F(0), \qquad TrF(\rho_\mathcal{B}) = \sum_i F(\rho_i) + g_\mathcal{B}F(0)$$

and

$$TrF(\rho_\mathcal{A}) - TrF(\rho_\mathcal{B}) = (g_\mathcal{A} - g_\mathcal{B})F(0)$$

where $g_\mathcal{A}$ and $g_\mathcal{B}$ are the degeneracies of the zero eigenvalues of $\rho_\mathcal{A}$ and $\rho_\mathcal{B}$.

Proof Let us prove the Schmidt theorem. Let $\{|\mu\rangle_\mathcal{A}\}$ be an orthonormal basis of $\mathcal{H}_\mathcal{A}$ and $\{|\mu'\rangle_\mathcal{B}\}$ an orthonormal basis of $\mathcal{H}_\mathcal{B}$. We can expand any pure state in the tensor product basis,

$$|\Psi\rangle = \sum_{\mu,\mu'} a_{\mu\mu'}|\mu\rangle_\mathcal{A} \otimes |\mu'\rangle_\mathcal{B}$$

For each μ set

$$|\tilde{\mu}\rangle_{\mathcal{B}} = \sum_{\mu'} a_{\mu\mu'}|\mu'\rangle_{\mathcal{B}}$$

so that

$$|\Psi\rangle = \sum_{\mu} |\mu\rangle_{\mathcal{A}} \otimes |\tilde{\mu}\rangle_{\mathcal{B}}$$

Note that $\{|\tilde{\mu}\rangle_{\mathcal{B}}\}$ is not necessarily an orthonormal basis so this is not yet a Schmidt decomposition. For the reduced density matrix of the $\mathcal{A}$ part we get

$$\rho_{\mathcal{A}} = \sum_{\mu_1,\mu_2} \langle\tilde{\mu}_2|\tilde{\mu}_1\rangle_{\mathcal{B}}|\mu_1\rangle_{\mathcal{A}}\langle\mu_2|_{\mathcal{A}}$$

Suppose now that

$$\rho_{\mathcal{A}}|i\rangle_{\mathcal{A}} = \rho_i|i\rangle_{\mathcal{A}}$$

For the basis $\{|\mu\rangle_{\mathcal{A}}\}$ we take $\{|i\rangle_{\mathcal{A}}\}$, so

$$\rho_{\mathcal{A}} = \sum_{i_1,i_2} \langle\tilde{i}_2|\tilde{i}_1\rangle_{\mathcal{B}}|i_1\rangle_{\mathcal{A}}\langle i_2|_{\mathcal{A}}$$

But we also have

$$\rho_{\mathcal{A}} = \sum_{i_1} \rho_{i_1}|i_1\rangle_{\mathcal{A}}\langle i_1|_{\mathcal{A}}$$

So for all non zero terms, $\rho_{i_1} \neq 0$, we must have $\langle\tilde{i}_2|\tilde{i}_1\rangle_{\mathcal{B}} = \rho_{i_1}\delta_{i_1 i_2}$. Thus the states $|\tilde{i}\rangle_{\mathcal{B}}$ are orthogonal and we can make them orthonormal by defining

$$|i\rangle_{\mathcal{B}} = \rho_i^{-1/2}|\tilde{i}\rangle_{\mathcal{B}}$$

In this way we obtain the expansion

$$|\Psi\rangle = \sum_i |i\rangle_{\mathcal{A}} \otimes |\tilde{i}\rangle_{\mathcal{B}} = \sum_i \sqrt{\rho_i}|i\rangle_{\mathcal{A}} \otimes |i\rangle_{\mathcal{B}}$$

which is the Schmidt decomposition (statement b)). To obtain statement a) we simply compute the partial traces from this expansion which leads to

$$\rho_{\mathcal{A}} = \sum_i \rho_i|i\rangle_{\mathcal{A}}\langle i|_{\mathcal{A}}, \qquad \rho_{\mathcal{B}} = \sum_i \rho_i|i\rangle_{\mathcal{B}}\langle i|_{\mathcal{B}}$$

These expressions show that $\rho_{\mathcal{A}}$ and $\rho_{\mathcal{B}}$ have the same non zero eigenvalues with the same multiplicities. Now suppose we have a second Schmidt decomposition. This will lead to a second spectral decomposition for $\rho_{\mathcal{A}}$ and $\rho_{\mathcal{B}}$. Thus the unicity of the Schmidt decomposition up to rotations in the span of each ρ_i follows from the same fact for the spectral decomposition. □

Notion of Schmidt number. The number of non-zero coefficients (including multiplicity) in the Schmidt decomposition of $|\Psi\rangle$ is called the *Schmidt number*

of the state. It is invariant under unitary evolutions that do not couple $\mathcal{A}$ and $\mathcal{B}$. Indeed if $U = U_{\mathcal{A}} \otimes U_{\mathcal{B}}$ then

$$U|\Psi\rangle = \sum_i \sqrt{\rho_i} U_{\mathcal{A}}|i\rangle_{\mathcal{A}} \otimes U_{\mathcal{B}}|i\rangle_{\mathcal{B}}$$

which has the same number of non zero coefficients. This number is also the number of non-zero eigenvalues of the reduced density matrices $Tr_{\mathcal{B}}|\Psi\rangle\langle\Psi|$ and $Tr_{\mathcal{A}}|\Psi\rangle\langle\Psi|$. This number can change only if $\mathcal{A}$ and $\mathcal{B}$ interact in some way.

Obviously a tensor product state has Schmidt number equal to 1. Since an entangled state is one which cannot be written as a tensor product state its Schmidt number is necessarily ≥ 2. The Schmidt number is our first attempt to quantify the degree of entanglement.

Purification. This turns out to be a powerful mathematical tool. Given a system S with Hilbert space $\mathcal{H}_S$ and density matrix ρ_S one can view it a part of a bigger system $S \cup R$ with Hilbert space $\mathcal{H}_S \otimes \mathcal{H}_R$ in a *pure* state $|\Psi\rangle_{SR}$ such that

$$\rho_S = Tr_R|\Psi\rangle_{SR}\langle\Psi|_{SR}$$

The Schmidt decomposition can be used to explicitly construct the pure state $|\Psi\rangle_{SR}$. One uses the spectral decomposition

$$\rho_S = \sum \rho_i |i\rangle_{\mathcal{A}}\langle i|_{\mathcal{A}}$$

and takes a copy of the space $\mathcal{H}_S$ - call it $\mathcal{H}_R$. Each vector $|i\rangle_S$ has a copy which we call $|i\rangle_R$. Then form

$$|\Psi\rangle_{SR} = \sum_i \sqrt{\rho_i}|i\rangle_S \otimes |i\rangle_R$$

The reader can easily check that $\rho_S = Tr_R|\Psi\rangle_{SR}\langle\Psi|_{SR}$.

Finally, we remark however that the purification is *not unique*.

6 Quantum entropy

There is a notion of entropy which quantifies the amount of uncertainty contained in an ensemble of Qbits. This is the von Neumann entropy that we introduce in this chapter. In some respects it behaves just like Shannon's entropy but in some others it is very different and strange. As an illustration let us immediately say that as in the classical theory, conditioning reduces entropy; but in sharp contrast with classical theory the entropy of a quantum system can be lower than the entropy of its parts.

The von Neumann entropy was first introduced in the realm of quantum statistical mechanics, but we will see in later chapters that it enters naturally in various theorems of quantum information theory.

6.1 Main properties of Shannon entropy

Let X be a random variable taking values x in some alphabet with probabilities $p_x = \text{Prob}(X = x)$. The Shannon entropy of X is

$$H(X) = \sum_x p_x \ln \frac{1}{p_x}$$

and quantifies the *average uncertainty* about X.

The joint entropy of two random variables X, Y is similarly defined as

$$H(X,Y) = \sum_{x,y} p_{x,y} \ln \frac{1}{p_{x,y}}$$

and the conditional entropy

$$H(X|Y) = \sum_y p_y \sum_{x,y} p_{x|y} \ln \frac{1}{p_{x|y}}$$

where

$$p_{x|y} = \frac{p_{x,y}}{p_y}$$

The conditional entropy is the average uncertainty of X given that we observe $Y = y$. It is easily seen that

$$H(X|Y) = H(X,Y) - H(Y)$$

The formula is consistent with the interpretation of $H(X|Y)$: when we observe Y the uncertainty $H(X,Y)$ is reduced by the amount $H(Y)$. The mutual information between X and Y is the complement of the remaining uncertainty $H(X|Y)$

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X,Y) = I(Y:X) \end{aligned} \tag{6.1}$$

It is easily seen that $I(X;Y) = 0$ iff $p_{x,y} = p_x p_y$.

The Kullback-Leibler divergence, or relative entropy, between two probability distributions p and q is a useful tool

$$D(p\|q) = \sum_x p_x \ln \frac{1}{q_x} - \sum_x p_x \ln \frac{1}{p_x} = \sum_x p_x \ln \frac{p_x}{q_x}$$

Note that this quantity is not symmetric, $D(p\|q) \neq D(q\|p)$. One can also check that

$$I(X;Y) = I(Y;X) = D(P_{X,Y}\|P_X P_Y)$$

Let us list the main inequalities of classical information theory and indicate which become true or false in the quantum domain.

- The maximum entropy state corresponds to the uniform distribution. For an alphabet with cardinality D we have

  $$0 \leq H(X) \leq \ln D$$

  with the upper bound attained iff $p_x = \frac{1}{D}$. Quantum mechanically this is still true.
- $H(X)$ is a concave functional of p_x. This means that if $p_0(x) = \sum_k a_k p_k(x)$, $a_k \geq 0$, $\sum_k a_k = 1$ then

  $$H_0(X) \geq \sum_k a_k H_k(X)$$

  QMly this is still true.
- Entropy is sub-additive,

  $$H(X,Y) \leq H(X) + H(Y)$$

  Equivalently conditioning reduces entropy $H(X|Y) \leq H(X)$, $H(Y|X) \leq H(Y)$, and mutual information is positive $I(X;Y) \geq 0$. QMly all this is true.
- The conditional entropy is positive, the entropy of (X,Y) is higher than that of X (or Y)

  $$H(X|Y) \geq 0, \qquad H(X,Y) \geq H(X), \qquad H(X,Y) \geq H(Y)$$

  with equality if $Y = f(X)$. QMly this is not true ! We will see that (again !) entanglement is responsible for this !

- Conditioning reduces conditional entropy

$$H(X|Y,Z) \leq H(X|Y)$$

This inequality is also called "strong sub-additivity and is equivalent to

$$H(X,Y,Z) + H(Y) \leq H(X,Y) + H(Y,Z)$$

Equality is attained iff $X - Y - Z$ form a Markov chain. This means that $p_{x,z|y} = p_{x|y}p_{z|y}$ or equivalently $p_{x,y,z} = p_{z|y}p_{y|x}p_x$ (a Markov chain is reversible: $Z-Y-X$ is also a Markov chain). We will see that QMly strong sub-additivity still holds. In view of the great gap in difficulty between the classical and quantum proofs it is fair to say that this fact is subtle and remarkable. However the notion of Markov chain is not obvious in the quantum case (there is no natural notion of conditional probability) so it is not easily asserted when equality holds.

- A consequence of strong sub-additivity is the data processing inequality obeyed by Markov chains $X - Y - Z$

$$H(X|Z) \geq H(X|Y)$$

Indeed $H(X|Z) \geq H(X|Z,Y) = H(X|Y)$ where the first inequality is strong sub-additivity and the equality follows from the fact Z and X are independent given Y. Since the notion of Markov chain is not clear QMly the quantum version of the data processing inequality is a subtle matter.

- The relative entropy is positive

$$D(p\|q) \geq 0$$

This is basically a convexity statement which is also true QMly.

- A very useful algebraic identity which follows immediately from definitions, is the chain rule

$$H(X_1,...,X_n|Y) = \sum_{i=1}^{n} H(X_i|X_{i+1},...,X_n,Y)$$

and

$$I(X_1,...,X_n|Y) = \sum_{i=1}^{n} I(X_i|X_{i+1},...,X_n,Y)$$

6.2 Von Neumann entropy and main properties

We assume that the system of interest is described by its density matrix ρ and furthermore we restrict ourselves to the case of a finite dimensional Hilbert space $\dim \mathcal{H} = D$. the von Neumann entropy is by definition

$$S(\rho) = -Tr\rho \ln \rho$$

In physics this quantity gives the right connection between quantum statistical mechanics and thermodynamics when $\rho = e^{-\beta H}/Z$ is the Gibbs state describing a mixture at thermal equilibrium. In quantum information theory this entropy enters in many theorems (data compression, measures of entanglement etc...) and thus acquires a fundamental status.

For the moment we just note that the definition is reasonable in the following sense. Suppose the quantum system is prepared in a mixture of states $\{|\phi_x\rangle; p_x\}$ so that its density matrix is

$$\rho = \sum_x p_x |\phi_x\rangle\langle\phi_x|$$

For the special case where $|\phi_x\rangle$ form an orthonormal basis of $\mathcal{H}$, this is a diagonal operator, so the eigenvalues of $\rho \ln \rho$ are $p_x \ln p_x$, and $S(\rho) = -\sum_x p_x \ln p_x = H(X)$, where X is the random variable with distribution p_x. In an orthogonal mixture all states can be perfectly distinguished so the mixture behaves classically: the quantum and classical entropies coincide.

We emphasize that for a general mixture the states $|\phi_x\rangle$ are not orthonormal so that $S(\rho) \neq H(X)$. In fact we will see that the following holds in full generality

$$S(\rho) \leq H(X)$$

where X is the random variable associated to the "preparation" of the mixture. This bound can be understood intuitively: since the states $|\phi_x\rangle$ cannot be perfectly distinguished (unless they are orthogonal, see chap 2) the quantum uncertainty associated to ρ is less than the classical uncertainty associated to X.

In the case of a pure state $\rho = |\Psi\rangle\langle\Psi|$ we see that the eigenvalues of ρ are 1 (multiplicity one) and 0 (multiplicity $D-1$). Thus

$$S(|\Psi\rangle\langle\Psi|) = 0$$

The entropy of a pure state is zero because there is no uncertainty in this state (in line with the Copenhagen interpretation of QM).

A quantity that plays an important role is also the relative entropy defined by analogy with the KL divergence

$$S(\rho||\sigma) = Tr\rho \ln \rho - Tr\rho \ln \sigma$$

Let us set up some notation concerning the entropy of composite systems and their parts. For a bipartite system $\mathcal{AB}$ with density matrix $\rho_{\mathcal{AB}}$ we write

$$S(\mathcal{AB}) = -Tr\rho_{\mathcal{AB}} \ln \rho_{\mathcal{AB}}$$

and for its parts described by the reduced density matrices $\rho_{\mathcal{A}} = Tr_{\mathcal{B}}\rho_{\mathcal{AB}}$ and $\rho_{\mathcal{B}} = Tr_{\mathcal{A}}\rho_{\mathcal{AB}}$,

$$S(\mathcal{A}) = -Tr\rho_{\mathcal{A}} \ln \rho_{\mathcal{A}}, \qquad S(\mathcal{B}) = -Tr\rho_{\mathcal{B}} \ln \rho_{\mathcal{B}}$$

One could try to pursue further the analogies with the classical case and define conditional entropies as $S(\mathcal{A}|\mathcal{B}) = S(\mathcal{AB}) - S(\mathcal{B})$, $S(\mathcal{B}|\mathcal{A}) = S(\mathcal{AB}) - S(\mathcal{A})$ and

mutual information as $I(\mathcal{A};\mathcal{B}) = I(\mathcal{B};\mathcal{A}) = S(\mathcal{A}) + S(\mathcal{B}) - S(\mathcal{AB})$. However it is not clear that these are of any fundamental use since they do not enter (yet) in any theorem of quantum information theory. Perhaps two more serious argument for suspicion are that first, as we will see $S(\mathcal{AB}) - S(\mathcal{B})$ can be negative, and second it is not at all clear how to define the quantum analog of conditional probabilities.

Let us now proceed to the statements and proofs of the basic inequalities satisfied by von Neumann's entropy.

- **Uniform distribution maximizes entropy.** Any ρ can be diagonalized and has positive eigenvalues ρ_x which sum to one. Thus $S(\rho) = -\sum \rho_x \ln \rho_x$, a quantity which is maximized for the distribution $\rho_x = \frac{1}{D}$ (as in the classical case). Thus in the basis where it is diagonal $\rho = \frac{1}{D} I$, and this is also true in any basis. We conclude

$$0 \leq S(\rho) \leq \ln D$$

  where the upper bound is attained for the "fully mixed" (or most disordered, or uniform) state $\rho = \frac{1}{D} I$. The lower bound is attained for pure states (check !).
- **Concavity.** Let ρ and σ be two density matrices. then

$$S(t\rho + (1-t)\sigma) \geq tS(\rho) + (1-t)S(\sigma), \qquad 0 \leq t \leq 1$$

  The proof follows the same lines as the classical one which uses convexity of $x \to x \ln x$. We prove below that $\rho \to Tr\rho \ln \rho$ is a convex functional and this immediately implies concavity of von Neumann's entropy.

  Lemma 6.2.1 (**Klein's inequality**) Let A and B self-adjoint and f convex from $\mathbf{R} \to \mathbf{R}$. We have

$$Tr(f(A) - f(B) - (A-B)f'(B)) \geq 0$$

  Proof Let $A|\phi_i\rangle = a_i|\phi_i\rangle$ and $B|\psi_i\rangle = b_i|\psi_i\rangle$. Then

$$Tr(f(A) - f(B) - (A-B)f'(B)) = \sum_i \langle\phi_i|f(A) - f(B) - (A-B)f'(B)|\phi_i\rangle \tag{6.2}$$

  Each term in the sum equals

$$f(a_i) - \langle\phi_i|f(B)|\phi_i\rangle - a_i\langle\phi_i|f'(B)|\phi_i\rangle + \langle\phi_i|Bf'(B)|\phi_i\rangle \tag{6.3}$$

  Using the closure relation

$$1 = \sum_j |\psi_j\rangle\langle\psi_j|$$

  equation (6.2) can be rewritten as

$$\sum_j |\langle\phi_i|\psi_j\rangle|^2 \Big(f(a_i - f(b_j) - (a_i - b_j)f'(b_j) \Big)$$

Now since $f : \mathbf{R} \to \mathbf{R}$ is convex we have

$$f(a_i) - f(b_j) \geq (a_i - b_j) f'(b_j)$$

which proves the statement. □

Corollary 6.2.2 Let A and B self-adjoint and positive (positive means that all eigenvalues are positive or equivalently that all diagonal averages $\langle\psi|A|\psi\rangle$ are positive for any $|\psi\rangle$). Then

$$TrA \ln A - TrA \ln B \geq Tr(A - B)$$

Proof Take $f(t) = t \ln t$ and apply Klein's inequality. □

Now choose $A = \rho$ and $B = t\rho + (1-t)\sigma$. From the corollary

$$Tr\rho \ln \rho - Tr\rho \ln(t\rho + (1-t)\sigma) \geq (1-t)Tr(\rho - \sigma) = 0$$

Choose $A = \sigma$ and $B = t\rho + (1-t)\sigma$. Then

$$Tr\sigma \ln \sigma - Tr\sigma \ln(t\rho + (1-t)\sigma) \geq tTr(\sigma - \rho) = 0$$

Multiplying the first inequality by t and the second by $(1-t)$ and adding them yields

$$Tr(t\rho + (1-t)\sigma) \ln(t\rho + (1-t)\sigma) \leq tTr\rho \ln \rho + (1-t)Tr\sigma \ln \sigma$$

which proves the concavity of entropy.

- **Positivity of relative entropy.** Choose $A = \rho$ and $B = \sigma$ and apply the corollary,

$$S(\rho||\sigma) = Tr\rho \ln \rho - Tr\rho \ln \sigma \geq Tr(\rho - \sigma) = 0$$

- **Sub-additivity.** In the classical case one has

$$H(X) + H(Y) - H(X,Y) = D(p_{x,y}||p_x p_y) \geq 0$$

In the quantum case the proof is similar, but we detail the steps

$$\begin{aligned} S(\mathcal{A}) + S(\mathcal{B}) - S(\mathcal{AB}) &= -Tr_{\mathcal{A}}\rho_{\mathcal{A}} \ln \rho_{\mathcal{A}} - Tr_{\mathcal{B}}\rho_{\mathcal{B}} \ln \rho_{\mathcal{B}} + Tr\rho_{cab} \ln \rho_{\mathcal{AB}} \\ &= -Tr\rho_{\mathcal{AB}} \ln \rho_{\mathcal{A}} \otimes I_{\mathcal{B}} - Tr\rho_{\mathcal{AB}} \ln I_{\mathcal{A}} \otimes \rho_{\mathcal{B}} + Tr\rho_{\mathcal{AB}} \ln \rho_{\mathcal{AB}} \\ &= Tr\rho_{\mathcal{AB}} \ln \rho_{\mathcal{AB}} - Tr\rho_{\mathcal{AB}}(\ln \rho_{\mathcal{A}} \otimes I_{\mathcal{B}} + \ln I_{\mathcal{A}} \otimes \rho_{\mathcal{B}}) \\ &= Tr\rho_{\mathcal{AB}} \ln \rho_{\mathcal{AB}} - Tr\rho_{\mathcal{AB}} \ln \rho_{\mathcal{A}} \otimes \rho_{\mathcal{B}} \\ &= S(\rho_{\mathcal{AB}}||\rho_{\mathcal{A}} \otimes \rho_{\mathcal{B}}) \geq 0 \end{aligned}$$

Note that sub-additivity can also formally be written as $S(\mathcal{A}|\mathcal{B}) \leq S(\mathcal{A})$ in terms of the naive conditional entropy. We may say that conditioning reduces quantum entropy, as in the classical case.

Exercise: check the identity $\ln \rho_{\mathcal{A}} \otimes I_{\mathcal{B}} + \ln I_{\mathcal{A}} \otimes \rho_{\mathcal{B}} = \ln \rho_{\mathcal{A}} \otimes \rho_{\mathcal{B}}$ by using spectral decompositions.

- **Araki-Lieb bound.** Classically $H(X,Y) \geq H(X)$ (the whole is more disordered than the parts). But quantum mechanically this can be completely wrong as the following counterexample shows. In quantum mechanics it is not true that the naive conditional entropy is always non-negative. Let

$$\rho_{\mathcal{AB}} = |B_{00}\rangle\langle B_{00}|$$

This is a pure state so $S(\mathcal{AB}) = 0$. However we have for the two parts

$$\rho_{\mathcal{A}} = \frac{1}{2} I_{\mathcal{A}}, \qquad \rho_{\mathcal{B}} = \frac{1}{2} I_{\mathcal{B}}$$

which have maximal entropies $S(\mathcal{A}) = S(\mathcal{B}) = \ln 2$. The two parts of the EPR pair when looked upon locally are as disordered as they can be, however the global state is highly correlated.

Is there a general good lower bound for $S(\mathcal{AB})$ in terms of the entropies of the parts ? The answer is provided by

THEOREM 6.2.3 (Araki-Lieb)

$$S(\mathcal{AB}) \geq |S(\mathcal{A}) - S(\mathcal{B})|$$

Proof The proof is a nice application of the purification idea and the Schmidt decomposition theorem. We introduce a third system R such that $\mathcal{ABR}$ is a purification of $\mathcal{AB}$. That is

$$\rho_{\mathcal{ABR}} = |\mathcal{ABR}\rangle\langle\mathcal{ABR}|, \qquad Tr_{\mathcal{R}}\rho_{\mathcal{ABR}} = \rho_{\mathcal{AB}}$$

By sub-additivity

$$S(\mathcal{AR}) \leq S(\mathcal{A}) + S(\mathcal{R}) \tag{6.4}$$

Now since $\rho_{\mathcal{ABR}}$ is a pure state the non-zero eigenvalues of $\rho_{\mathcal{AB}}$ and $\rho_{\mathcal{R}}$ are equal; and also the non zero eigenvalues of $\rho_{\mathcal{AR}}$ and $\rho_{\mathcal{B}}$ are equal (Schmidt theorem). Thus

$$S(\mathcal{AB}) = S(\mathcal{R}), \qquad S(\mathcal{AR}) = S(\mathcal{B})$$

Replacing in (6.4) we get

$$S(\mathcal{B}) - S(\mathcal{A}) \leq S(\mathcal{AB})$$

Since $\mathcal{A}$ and $\mathcal{B}$ play a symmetric role we can exchange them which ends the proof. □

- **Strong sub-additivity.** let $\mathcal{ABC}$ be a quantum system formed of three parts $\mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{B}} \otimes \mathcal{H}_{\mathcal{C}}$. We have similarly to the classical case

$$S(\mathcal{ABC}) + S(\mathcal{B}) \leq S(\mathcal{AB}) + S(\mathcal{BC})$$

This can be written also as $S(\mathcal{C}|\mathcal{AB}) \leq S(\mathcal{C}|\mathcal{B})$ in terms of "naive" conditional entropies. So one may say that further conditioning reduces conditional entropy (although the "conditional" entropy is not necessarily positive). As in classical information theory, this inequality plays an important role.

Classically the proof of this inequality is based on the positivity of the KL divergence. It turns out that quantum mechanically the proof is much more difficult. We will omit it here except for saying that one can base it on the *joint concavity* of the functional

$$f(A, B) = TrM^{\dagger}A^{s}MB^{(1-s)}$$

for any matrix M (not necessarily self-adjoint) and any $0 \leq s \leq 1$. This fact was a conjecture of Wigner-Yanase-Dyson for many years until Lieb found a proof (1973). Later, Lieb and Ruskai realized that it implies strong sub-additivity.

6.3 Useful bounds on the entropy of a mixtures

This section is devoted to the proof of the following important theorem

THEOREM 6.3.1 *Let X be a random variable with distribution p_x and $\rho = \sum_x p_x\rho_x$ where ρ_x are mixed states. We have*

$$S(\rho) \leq \sum_x p_x S(\rho_x) + H(X)$$

This inequality has a clear interpretation: the uncertainty about ρ cannot be greater than the average uncertainty about each ρ_x plus the uncertainty about the classical preparation described by X. If in particular $\rho_x = |\phi_x\rangle\langle\phi_x|$ are pure states we have $S(\rho_x) = 0$ so, as announced at the beginning of the chapter,

$$S(\rho) \leq H(X)$$

Proof First we deal with a mixture of pure states. For convenience we call this mixture $\mathcal{A}$ and set

$$\rho_{\mathcal{A}} = \sum_x p_x |\phi_x\rangle_{\mathcal{A}}\langle\phi_x|_{\mathcal{A}}$$

Let $\mathcal{H}_{\mathcal{R}}$ a space whose dimension is equal to the number of terms in the mixture and with orthonormal basis labeled as $|x\rangle_{\mathcal{R}}$. The pure state

$$|\mathcal{AR}\rangle = \sum_x \sqrt{p_x}|\phi_x\rangle_{\mathcal{A}} \otimes |x\rangle_{\mathcal{R}}$$

is a purification of $\rho_{\mathcal{A}}$ because

$$Tr_{\mathcal{R}}|\mathcal{AR}\rangle\langle\mathcal{AR}| = \sum_x p_x |\phi_x\rangle\langle\phi_x|_{\mathcal{A}} = \rho_{\mathcal{A}}$$

We also have that

$$\rho_{\mathcal{R}} = Tr_{\mathcal{A}}|\mathcal{AR}\rangle\langle\mathcal{AR}| = \sum_{x,x'} \sqrt{p_x}\sqrt{p_{x'}}\langle\phi_x|\phi_{x'}\rangle_{\mathcal{A}}|x\rangle_{\mathcal{R}}\langle x'|_{\mathcal{R}} \tag{6.5}$$

By the Schmidt theorem we know that $\rho_{\mathcal{A}}$ and $\rho_{\mathcal{R}}$ have the same non zero eigenvalues, thus

$$S(\rho_{\mathcal{A}}) = S(\rho_{\mathcal{R}})$$

Consider now

$$\rho'_{\mathcal{R}} = \sum_x p_x|x\rangle_{\mathcal{R}}\langle x|_{\mathcal{R}}$$

and look at the relative entropy

$$S(\rho_{\mathcal{R}}||\rho'_{\mathcal{R}}) = Tr\rho_{\mathcal{R}}\ln\rho_{\mathcal{R}} - Tr\rho_{\mathcal{R}}\ln\rho'_{\mathcal{R}} \geq 0$$

Thus

$$S(\rho_{\mathcal{A}}) = S(\rho_{\mathcal{R}}) \leq -Tr\rho_{\mathcal{R}}\ln\rho'_{\mathcal{R}} \tag{6.6}$$

It remains to compute the right hand side. Since $|x\rangle_{\mathcal{R}}$ is an orthonormal basis

$$\ln\rho'_{\mathcal{R}} = \sum_x (\ln p_x)|x\rangle_{\mathcal{R}}\langle x|_{\mathcal{R}}$$

which implies

$$Tr\rho_{\mathcal{R}}\ln\rho'_{\mathcal{R}} = \sum_x (\ln p_x)Tr\rho_{\mathcal{R}}|x\rangle_{\mathcal{R}}\langle x|_{\mathcal{R}} = \sum_x (\ln p_x)\langle x|\rho_{\mathcal{R}}|x\rangle$$

From the expression of $\rho_{\mathcal{R}}$ (6.5) we remark that

$$\langle x|\rho_{\mathcal{R}}|x\rangle = p_x$$

Thus (6.6) becomes

$$S(\rho_{\mathcal{A}}) \leq -\sum p_x \ln p_x = H(X)$$

Consider now the general case of a mixture of mixed states $\rho = \sum_x p_x\rho_x$. each mixed state has a spectral decomposition

$$\rho_x = \sum_j \lambda_j^{(x)}|e_j^{(x)}\rangle\langle e_j^{(x)}|$$

so

$$\rho = \sum_{x,j} p_x\lambda_j^{(x)}|e_j^{(x)}\rangle\langle e_j^{(x)}|$$

Note that this is a convex combination of one dimensional projectors so that we

can apply the previous result

$$\begin{aligned} S(\rho) &\leq -\sum_{x,j} p_x \lambda_j^{(x)} \ln p_x \lambda_j^{(x)} \\ &= -\sum_{x,j} p_x \lambda_j^{(x)} \ln p_x - \sum_{x,j} p_x \lambda_j^{(x)} \ln \lambda_j^{(x)} \\ &= -\sum_{x} p_x \ln p_x - \sum_{x} p_x \sum_{j} \lambda_j^{(x)} \ln \lambda_j^{(x)} \\ &= H(X) + \sum_x p_x S(\rho_x) \end{aligned}$$

In the last equality we used $S(\rho_x) = \sum_j \lambda_j^{(x)} \ln \lambda_j^{(x)}$.

□

6.4 Measuring without learning the measurement outcome cannot decrease entropy

Suppose we are given a mixed state ρ and a measurement apparatus with measurement basis $\{|x\rangle\langle x|\}$. According to the measurement postulate the possible outcomes are pure states

$$|x\rangle, \qquad \text{with probability } p_x = \langle x|\rho|x\rangle$$

[Note that $\sum_x \langle x|\rho|x\rangle = 1$]. If we observe the measurement result we know that we have some $|x\rangle$ with zero entropy.

Now imagine that we do the measurement but do not record the measurement result (subsequently we will call this a "blind" measurement). Then our description of the state of the system is a mixture $\{|x\rangle, p_x\}$ with diagonal density matrix

$$\rho_{\text{blind}} = \sum_x \langle x|\rho|x\rangle |x\rangle\langle x|$$

Note that this diagonal density matrix is equivalent to a classical state. If we look at the relative entropy

$$S(\rho||\rho_{\text{blind}}) \geq 0$$

we find, by a small calculation[1],

$$S(\rho) \leq H(\langle x|\rho|x\rangle) = S(\rho_{\text{blind}})$$

Thus blind measurements can only increase the entropy or leave it constant.

To conclude the chapter consider again a composite system $\mathcal{AB}$ where Alice and Bob are very far apart and do not communicate. A local measurement (with an apparatus $\{|i\rangle_{\mathcal{A}}\langle i|_{\mathcal{A}}\}$) is done by Alice on part $\mathcal{A}$ which is blind to Bob. Thus according to the previous inequality $S(\rho_{\mathcal{B}}^{\text{blind}}) - S(\rho_{\mathcal{B}}) \geq 0$. However a true

[1] identical to the one in the proof of the upper bound in the previous section

(immediate) increase would violate locality and it is very reassuring to check that $S(\rho_{\mathcal{B}}^{\text{blind}}) = S(\rho_{\mathcal{B}})$

After Alice's measurement the possible outcomes for the total system are

$$\frac{(|i\rangle_{\mathcal{A}}\langle i|_{\mathcal{A}} \otimes I_{\mathcal{B}})\rho_{\mathcal{AB}}(|i\rangle_{\mathcal{A}}\langle i|_{\mathcal{A}} \otimes I_{\mathcal{B}})}{Tr(|i\rangle_{\mathcal{A}}\langle i|_{\mathcal{A}} \otimes I_{\mathcal{B}})\rho_{\mathcal{AB}}(|i\rangle_{\mathcal{A}}\langle i|_{\mathcal{A}} \otimes I_{\mathcal{B}})}$$

or equivalently

$$\rho_{\mathcal{AB}}^{(i)} = \frac{(|i\rangle_{\mathcal{A}}\langle i|_{\mathcal{A}} \otimes I_{\mathcal{B}})\rho_{\mathcal{AB}}(|i\rangle_{\mathcal{A}}\langle i|_{\mathcal{A}} \otimes I_{\mathcal{B}})}{\langle i|\rho_{\mathcal{A}}|i\rangle}$$

with probability (we set $|i\rangle_{\mathcal{A}} = |i\rangle$ to alleviate the notation)

$$\langle i|\rho_{\mathcal{A}}|i\rangle$$

Since this is a blind measurement for Bob the reduced density matrix is (a mixture of mixed states)

$$\rho_B^{\text{blind}} = \sum_i \langle i|\rho_{\mathcal{A}}|i\rangle Tr_A \rho_{\mathcal{AB}}^{(i)}$$

A short calculation shows that this equals

$$\rho_B^{\text{blind}} = \sum_i \langle i|\rho_{\mathcal{A}}|i\rangle \frac{\langle i|\rho_{\mathcal{AB}}|i\rangle}{\langle i|\rho_A|i\rangle} = \sum_i \langle i|\rho_{\mathcal{AB}}|i\rangle = Tr_A \rho_{\mathcal{AB}} = \rho_B$$

So after Alice's measurement not only Bob's entropy is unchanged but even his density matrix is left the same as it was before the measurement. This provides a completely general proof that Bob does not notice Alice's measurements. On Alice's side if she does not record her measurement outcome her entropy is greater.

7 Accessible information

In this chapter we prove a very important bound of quantum information theory, namely Holevo's bound. Suppose we are given a system prepared in a mixed state represented by a density matrix. Information about the *preparation* of this state can be retrieved by making measurements on the system. The Holevo bound gives an upper bound on the maximaum possible *information that can be extracted from the mixed state by a measurement process*. Holevo was a precursor, indeed this is the first information theoretic estimate that was derived and involves von neumann entropy as a basic ingredient. In fact we will see an important new quantity that enters in this estimate and is nowadys called the *Holevo quantity*. We will see in a later chapter that it has important applications in channel coding theory, and play there it plays the role of a mutial information.

7.1 Notion of accessible information

We argued in chapter 2 that non-orthogonal quantum states are not perfectly distinguishable. The Holevo bound quantifies this statement. Suppose a system with Hilbert space $\mathcal{H}$ is prepared in a mixed state $\{p_x, \rho_x\}$ where ρ_x are density matrices (hence the preparation of the system is a mixture of mixed states). The total density matrix of the system is

$$\rho = \sum_x p_x \rho_x$$

We imagine that Alice has prepared the mixture $\{p_x, \rho_x\}$ but "gives only ρ" to Bob who wants to extract information about the preparation by performing measurements on ρ. Let us formalize the problem.

- The preparation of Alice is described by a classical random variable X taking value x with $\text{Prob}(X = x) = p_x$. For example Alice flip a coin: if Face is obtained with $p_F = \frac{1}{2}$ she prepares a photon in state $\rho_F = |0\rangle\langle 0|$, while if Tail is obtained with $p_T = \frac{1}{2}$ she prepares a photon in state $\rho_T = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)(\frac{1}{\sqrt{2}}(\langle 0| + \langle 1|)$. In this way Alice prepares an ensemble of photons (trapped in a cavity say).
- Bob is given the enesemble of photons described by the mixed state ρ. However he does not know the details $\{p_x, \rho_x\}$ of the preparation. He has full access

to ρ in the sense that he can manipulate and measure the state, and wants to retrieve information about the preparation.

In the example

$$\begin{aligned}\rho &= \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)(\frac{1}{\sqrt{2}}(\langle 0| + \langle 1|) \\ &= \frac{3}{4}|0\rangle\langle 0| + \frac{1}{4}|0\rangle\langle 1| + \frac{1}{4}|1\rangle\langle 0| + \frac{1}{4}|1\rangle\langle 1| \\ &= \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}\end{aligned}$$

- Bob makes measurements (on the ensemble of photons) with an apparatus corresponding to a measurement basis $\{P_y\}$ where $P_y^2 = P_y$ are projectors and $\sum_y P_y = 1$. The outcome of the measurement is a random variable Y

$$\text{Prob}(Y = y) = \sum_x p_x Tr \rho_x P_y = Tr \rho P_y \qquad (= p_y)$$

Note that the last equation also follows directly from the measurement postulate applied to ρ.

Here we can a natural *define* a conditional probability distribution from

$$\text{Prob}(Y = y | X = x) = Tr \rho_x P_y \qquad (= p_{y|x})$$

This also allows to define a joint probability distribution

$$\text{Prob}(X = x, Y = y) = p_x p_{y|x} = p_x Tr \rho_x P_y \qquad (= p_{x,y})$$

It is good to check that p_y is a marginal of $p_{x,y}$ and the marginal

$$\text{Prob}(Y = y) = \sum_x p_x Tr \rho_x P_y = Tr \rho P_y \qquad (= p_y)$$

and of course that p_x is the other marginal.

In the example, suppose that Bob uses a measurement apparatus corresponding to the canonical basis $\{|0\rangle\langle 0|; |1\rangle\langle 1|\}$. One obtains for the distribution of (measurements) Y

$$p_y = \begin{bmatrix} \frac{3}{4} \\ \frac{1}{4} \end{bmatrix}. \tag{7.1}$$

For the conditional distribution one obtains

$$p_{y|x} = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix}, \tag{7.2}$$

for the joint distribution

$$p_{x,y} = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} \end{bmatrix}, \tag{7.3}$$

- The mutual information $I(X;Y)$ defined from $p_{x,y}$ is the information about X that Bob can extract from ρ by his measurement outcomes Y. We define the *accessible information* as the maximum possible mutual information obtained by the best possible measurement

$$\mathrm{Acc}(\{p_x,\rho_x\}) = \sup_{\{P_y\}} I(X;Y)$$

  In the example we have $H(X) = \ln 2$, $H(Y) = \ln 4 - \frac{3}{4}\ln 3$ and $H(X,Y) = \frac{3}{2}\ln 2$. Thus for the particular measurement in the canonical basis $I(X;Y) = \frac{3}{2}\ln 2 - \frac{3}{4}\ln 3 = 0.215$. This equals $0.31 \ln 2$ so Bob retrieves 0.31 bits from this type of measurement. He can do better by choosing a more clever basis but, since the states ρ_F and ρ_T are not perfectly distinguishable, his accessible information will always be strictly smaller than 1 bit (the entropy of X). An interesting question partly answered in the next paragraph is : how much smaller is it ?

  Note that if ρ_x are pure orthogonal states they form a subset of a basis of the Hilbert space. Thus by choosing *this basis* as a measurement basis Bob gets $Y = X$ so that $I(X;Y) = H(X)$. This means that a mixture of orthogonal states behaves as a classical probability distribution and can be perfectly known by suitable measurements.

7.2 The Holevo bound

In general it is very difficult to compute the supremum over all possible measurement basis, involved in the definition of the accessible information. Holevo (following pioneering works of Gordon and Levitin) gave a bound which gives us an estimate that is independent of the measurement basis. In general this bound is loose and is not achievable by a measurement basis. The achievability holds for special mixtures, as briefly discussed in the next paragraph, and plays an important role in channel coding theorems.

Theorem [Holevo bound]. Let X be a classical random variable $\{p_x = \mathrm{Prob}(X = x)\}$ and $\{p_x,\rho_x\}$ a mixture of mixed quantum states. Let Y be the random variable describing outcomes of measurements on the state $\rho = \sum_x p_x\rho_x$ in the basis $\{P_y\}$ (these can be measurements of any observable that has the spectral decomposition $A = \sum_y a_y P_y$). Then

$$I(X;Y) \le \chi(\{p_x,\rho_x\}), \qquad \text{so also} \qquad \mathrm{Acc}(\{p_x,\rho_x\}) \le \chi(\{p_x,\rho_x\})$$

where

$$\chi(\{p_x,\rho_x\}) = S(\rho) - \sum_x p_x S(\rho_x)$$

In the example ρ_F and ρ_T are pure so their individual entropies vanish, and $\chi(\{p_x,\rho_x\}) = S(\rho)$. The eigenvalues of ρ are $\rho_\pm = \frac{1}{2} \pm \frac{\sqrt{2}}{4}$. So the von Neumann

entropy is

$$S(\rho) = -(\frac{1}{2} + \frac{\sqrt{2}}{4}) \ln(\frac{1}{2} + \frac{\sqrt{2}}{4}) - (\frac{1}{2} - \frac{\sqrt{2}}{4}) \ln(\frac{1}{2} - \frac{\sqrt{2}}{4}) = 0.41 = 0.59 \ln 2$$

We can conclude that there are no measurements that would retrieve more than 0.59 bits of information from X.
Exercise: compute $Acc(\{p_x, \rho_x\})$.

Proof of the Holevo Bound. Bob is given the mixed state $\rho_Q = \sum_x p_x \rho_x$ which we view as a state belonging to $\mathcal{H}_Q$. We introduce a larger Hilbert space $\mathcal{H}_X \otimes \mathcal{H}_Q \otimes \mathcal{H}_Y$ and a state

$$\rho_{XQY} = \sum_x p_x |x\rangle\langle x| \otimes \rho_x \otimes |0\rangle\langle 0|$$

The interpretation of this state is as follows: $|x\rangle\langle x|$ are mutual orthogonal states describing Alice's preparation (or r.v X) and $|0\rangle\langle 0|$ is a blank state where Bob will record his measurement outcomes. Note that $dim\mathcal{H}_X$ = number of values of x, $dim\mathcal{H}_Q$ is the dimension of the Hilbert space in which Bob's state lives (e.g 2 if this is a single Qbit) and $dim\mathcal{H}_Y = dim\mathcal{H}_Q$ since $\mathcal{H}_Y$ records the measurement outcomes. For the measurement basis of Bob we take $\{P_y = |y\rangle\langle y|\}$.

We introduce the unitary operation

$$U_{XQY} = Id \otimes U_{QY}$$

where

$$U_{QY}|\phi\rangle_Q \otimes |a\rangle_Y = \sum_y P_y|\phi\rangle_Q \otimes |a \oplus y\rangle_Y$$

Here $a \oplus y$ is computed modulo $dim\mathcal{H}_Y$. Let us check that this is a unitary operation. We have

$$\begin{aligned}
\langle\psi| \otimes \langle b|U_{QY}^\dagger U_{QY}|\phi\rangle \otimes |a\rangle &= \sum_{y,y'} \langle\psi|P_{y'} \otimes \langle b \oplus y'|P_y\phi\rangle \otimes |a \oplus y\rangle \\
&= \sum_{y,y'} \langle\psi|P_{y'}P_y|\phi\rangle\langle b \oplus y'|a \oplus y\rangle \\
&= \sum_{y,y'} \delta_{y,y'} \langle\psi|P_y|\phi\rangle\langle b \oplus y|a \oplus y\rangle \\
&= \langle\psi|\phi\rangle\langle b|a\rangle
\end{aligned}$$

Thus $Id \otimes U_{QY}$ preserves the inner product and is unitary.

Now we define

$$\rho'_{XQY} = U_{XQY}\rho_{XQY}U_{XQY}^\dagger = \sum_{x,y,y'} p_x|x\rangle\langle x| \otimes P_y\rho_x P_{y'} \otimes |y\rangle\langle y'|$$

The two density matrices ρ_{XQY} and ρ'_{XQY} have the same eigenvalues (since they are unitarily related) therefore their von Neumann entropies are the same

$$S(\rho_{XQY}) = S(\rho'_{XQY})$$

The two partial density matrices

$$\rho_{QY} = Tr_X \rho_{XQY} = \sum_x p_x \rho_x \otimes |0\rangle\langle 0| = \rho \otimes |0\rangle\langle 0|$$

and

$$\rho'_{QY} = Tr_X \rho'_{XQY} = \sum_x p_x P_y \rho_x P'_y \otimes |y\rangle\langle y'|$$

are also unitarily related because of the tensor product form of $U_{XQY} = Id \otimes U_{QY}$. Thus we also have

$$S(\rho_{QY}) = S(\rho'_{QY})$$

From the strong sub-additivity (in the form $S(X|QY) \leq S(X|Y)$ say)

$$S(\rho'_{XQY}) - S(\rho'_{QY}) \leq S(\rho'_{XY}) - S(\rho'_Y),$$

thus we get

$$S(\rho_{XQY}) - S(\rho_{QY}) \leq S(\rho'_{XY}) - S(\rho'_Y).$$

The rest of the proof is a computation of all the entropies appearing in this last inequality. For the first one we have (since the pure part $|0\rangle\langle 0|$ has zero von Neumann entropy)[1]

$$S(\rho_{XQY}) = S\big(\sum_x p_x |x\rangle\langle x| \otimes \rho_x\big)$$

To compute this entropy we use the spectral decomposition $\rho_x = \sum_{a_x} \lambda_{a_x} |a_x\rangle\langle a_x|$. Then

$$\sum_x p_x |x\rangle\langle x| \otimes \rho_x = \sum_{x,a_x} p_x \lambda_{a_x} |x\rangle\langle x| \otimes |a_x\rangle\langle a_x|$$

Since this is a convex combination of mutually orthogonal states we have that its entropy is

$$\begin{aligned} -\sum_{x,a_x} p_x \lambda_{a_x} \ln p_x \lambda_{a_x} &= H(X) - \sum_x p_x \sum_{a_x} \lambda_{a_x} \ln \lambda_{a_x} \\ &= H(X) + \sum_x p_x S(\rho_x) \end{aligned} \tag{7.4}$$

Thus

$$S(\rho_{XQY}) = H(X) + \sum_x p_x S(\rho_x)$$

For the second entropy since $\rho_{QY} = \rho \otimes |0\rangle\langle 0|$ we simply have

$$S(\rho_{QY}) = S(\rho).$$

For the third one, we first compute the reduced density matrix

$$\rho'_{XY} = Tr_Q \rho'_{XQY} = \sum_{x,y,y'} p_x |x\rangle\langle x| \otimes |y\rangle\langle y'| Tr_Q P_y \rho_x P_{y'}$$

[1] More formally use $\mathrm{Tr}\rho_A \otimes \rho_B \ln \rho_A \otimes \rho_B = \mathrm{Tr}\rho_A \ln \rho_A + \mathrm{Tr}\rho_B \ln \rho_B$).

By the cyclicity of the trace

$$Tr_Q P_y \rho_x P_{y'} = Tr_Q P_{y'} P_y \rho_x = \delta_{yy'} Tr_Q P_y \rho_x = \delta_{yy'} p_{y|x}$$

Thus we find

$$\rho'_{XY} = \sum_{x,y} p_{x,y} |x\rangle\langle x| \otimes |y\rangle\langle y|$$

The states $|x\rangle\langle x| \otimes |y\rangle\langle y|$ are mutually orthogonal. Thus this density matrix is just another representation for the random variable (X, Y). The von Neumann entropy is

$$S(\rho'_{XY}) = H(X, Y)$$

Now it remains to compute the last entropy $S(\rho'_Y)$. We have

$$\rho'_Y = Tr\rho'_{XY} = \sum_{x,y} p_{x,y} |y\rangle\langle y| = \sum_y p_y |y\rangle\langle y|$$

therefore

$$S(\rho'_Y) = H(Y)$$

Collecting all these entropies and replacing them in the strong sub-additivity inequality we obtain

$$H(X) + \sum_x p_x S(\rho_x) - S(\rho) \leq H(X, Y) - H(Y)$$

which is the same as

$$I(X; Y) \leq S(\rho) - \sum_x p_x S(\rho_x) = \chi(X; \rho)$$

This ends the proof of Holevo's bound.

7.3 Remarks on the achievability of Holevo's bound

Given a measurement basis $\{P_y\}$ we have

$$\begin{aligned} I(X; Y) =& H(Y) - H(Y|X) \\ =& -\sum_y (Tr P_y \rho) \ln(Tr P_y \rho) + \sum_x p_x \sum_y (Tr P_y \rho_x) \ln(Tr P_y \rho_x) \end{aligned}$$

The Holevo bound states that for any $\{P_y\}$ this expression is less than

$$S(\rho) - \sum_x p_x S(\rho_x)$$

In general, given a mixture $\{p_x, \rho_x\}$ it is difficult to assess if there exists a measurement basis $\{P_y\}$ such that the bound is achieved. A positive answer can be given in special important cases.

For a mixture of pure states $\{p_x, |\phi_x\rangle\langle\phi_x|\}$ and a measurement basis $P_y = |y\rangle\langle y|$ we have

$$I(X;Y) = -\sum_y \big(\sum_x p_x |\langle y|\phi_x\rangle|^2\big) \ln\big(\sum_x p_x |\langle y|\phi_x\rangle|^2\big) + \sum_x p_x \sum_y |\langle y|\phi_x\rangle|^2 \ln |\langle y|\phi_x\rangle|^2$$

If the states $|\phi_x\rangle$ are mutually orthonormal, and we choose a measurement basis containing all these states, we find

$$I(X;Y) = H(X)$$

But since $S(\rho) \leq H(X) + \sum_x p_x S(\rho_x)$ (a general bound proved in chapter 6) we always have

$$\chi(p_x;\rho_x) \leq H(X)$$

Therefore we see that (for mixtures of mutually orthonormal states) the equality is achieved for mixtures of orthonormal states, by a measurement basis containing these states. This result is an expression of the fact that orthonormal states can be perfectly distinguished: we gain the maximum possible amount of mutual information by doing the right measurements.

These arguments can be generalized to the case of a mixture such that the density matrices ρ_x are mutually orthonormal in the sense that

$$Tr\rho_x\rho_{x'} = 0, \qquad x \neq x'$$

This means that for no zero eigenvalues the eigenprojectors of ρ_x and $\rho_{x'}$ are mutually orthogonal. If we set $\rho_x = \sum_j \lambda_j P_{j,x}$ for the spectral decomposition, we have (for non zero λ_j's)

$$TrP_{j,x}P_{j',x'} = \delta_{j,j'}\delta_{x,x'}$$

This can be checked by replacing the spectral decompositions of the density matrices in the trace and noting that all terms in the sum are non-negative. We leave it as an exercise for the reader to check that if the measurement basis $\{P_y\}$ contains $\{P_{j,x}\}$ one gets

$$I(X;Y) = S(\rho) - \sum_x p_x S(\rho_x)$$

Summarizing, when the density matrices are mutually orthogonal, the Holevo bound can again be attained by an appropriate measurement basis.

A more sophisticated case were achievability can be proven is the following. Take a finite "alphabet" of density matrices ρ_x (for example $|0\rangle\langle 0|$ and $\frac{1}{4}|0\rangle\langle 0| + \frac{3}{4}|1\rangle\langle 1|$) and fix a classical distribution p_x. Take M tensor products of n elements picked in the alphabet: $\{\rho_{x_1} \otimes \cdots \otimes \rho_{x_n}\}$. Now given $\{p_x, \rho_x\}$, as long as M is

not too large it is possible to find M tensor products that are asymptotically mutually orthogonal

$$Tr(\rho_{x_1} \otimes \cdots \otimes \rho_{x_n})(\rho_{x'_1} \otimes \cdots \otimes \rho_{x'_n}) \to 0, \qquad n \to \infty$$

The proof uses the probabilistic method, in the spirit of Shannon's theory. One picks the tensor product strings randomly according to the distribution $p_{x_1}...p_{x_n}$ to first show that the tensor products are mutually orthogonal on average. Then usual arguments show that there must exist one such choice. The proof shows that one should have $\frac{\log_2 M}{n} \leq \chi(\{p_x, \rho_x\})$. Combining this result with the arguments of the previous paragraph we see that in such a situation there exists a measurement basis in $\mathcal{H}^{\otimes n}$ that asymptotically achieves the Holevo bound as $n \to +\infty$,

$$\frac{1}{n}\left| I(X_1...X_n; Y_1...Y_n) - \left(S(\rho^{\otimes n}) - \sum_{x_1...x_n} p_{x_1}...p_{x_n} S(\rho_{x_1} \otimes \cdots \otimes \rho_{x_n}) \right) \right| \to 0 \quad (7.5)$$

This remark is at the basis of one of the capacity theorems for communication of classical messages accross quantum channels.

8 Compression of a Quantum State

We are now ready derive the direct analog of Shannon's lossless source coding theorem, that was first analyzed by Schumacher.

In the classical case we are given a memoryless source which produces strings $x_1x_2...x_n$ where each letter $x \in \{0,1\}$ and occurs with probability $\text{Prob}(X = x) = p_x$. One shows that for n sufficiently large, the outputs of the source can be faithfully be described by nR bits as long as $R > H(X)$ and also that this is not possible if $R < H(X)$. Thus length n messages can compressed be to length nR messages with negligible error as $n \to +\infty$.

In the quantum case a memoryless source produces tensor product states $\rho_{x_1} \otimes \cdots \rho_{x_n}$ each "letter" ρ_x belonging to a finite set $\mathcal{A}$ of $d \times d$ density matrices (the quantum alphabet) and occurring with probability p_x (so $\sum_{x\in\mathcal{A}} p_x = 1$). The quantum state of the source is therefore

$$\sum_{x_1...x_n} p_{x_1}...p_{x_n}\rho_{x_1}...\rho_{x_n} = \big(\sum_x p_x\rho_x\big)^{\otimes n} = \rho^{\otimes n} \tag{8.1}$$

This is a density matrix of dimension $d^n \times d^n$ (the Hilbert space of pure states has dimension d^n; for example $d = 2$ for Qbits, $d = 3$ for "quantum trits" ect). We want to compress the source: this means that we want to represent it faithfully by states (or density matrices) of a Hilbert space of dimension d^{nR}.

In general this problem is open. It is known that it is not possible to achieve a compression rate $R < \chi(\{p_x, \rho_x\})$, but it is not known that any rate higher than the Holevo quantity is achievable. However, Schumacher solved the special case where the alphabet letters $\rho_x \in \mathcal{A}$ are pure states $\rho_x = |\phi_x\rangle\langle\phi_x|$. Namely any rate $R > S(\rho)$ is achievable while it is not possible to faithfully compress at rates $R < S(\rho)$. Note that if the alphabet consists of orthonormal states (say $\{|0\rangle, |1\rangle\}$) $S(\rho) = H(X)$ so one recovers the classical Shannon theorem. This should be so, since orthonormal states are perfectly distinguishable, so that the problem becomes equivalent to the classical one. In the sequel we concentrate on sources of pure states that live in $\mathbf{C}^d$.

8.1 Notion of typical subspace

In the classical case, the space of length n strings is partitioned into $\mathcal{T}_{n,\epsilon} \cup \mathcal{T}^c_{n,\epsilon}$ where

$$\mathcal{T}_{n,\epsilon} = \left\{ \{x_1 ... x_n| \left| \frac{1}{n} \sum_{i=1}^{n} \log_d \frac{1}{p(x_i)} - H(X) \right| \leq \epsilon \right\} \tag{8.2}$$

is called the space of (weakly) typical sequences. This definition implies that all typical sequences have approximately the same probability, namely

$$d^{-n(H(X)+\epsilon)} \leq p_{x_1} ... p_{x_n} \leq d^{-n(H(X)-\epsilon)} \tag{8.3}$$

By the law of large numbers, for any ϵ and δ small positive we can find n large enough such that

$$1 - \delta \leq \text{Prob}(\mathcal{T}_{n,\epsilon}) \leq 1 \tag{8.4}$$

Summing (8.3) over typical sequences and using (8.4) we also deduce an estimate on the number of typical sequences

$$(1 - \delta) d^{n(H(X)-\epsilon)} \leq |\mathcal{T}_{n,\epsilon}| \leq d^{n(H(X)+\epsilon)} \tag{8.5}$$

Finally, any set $S_{n,\epsilon}$ of sequences that is too small in the sense that $|S_{n,\epsilon}| \leq d^{nR}$ with $R \leq H(X) - \epsilon$ has negligible probability,

$$\text{Prob}(S_{n,\epsilon}) \leq \delta + d^{-n(H(X)-\epsilon-R)} \tag{8.6}$$

To see this write

$$S_{n,\epsilon} = (S_{n,\epsilon} \cap \mathcal{T}_{n,\epsilon}) \cup (S_{n,\epsilon} \cap \mathcal{T}^c_{n,\epsilon}) \tag{8.7}$$

and use (8.3) with the union bound.

These properties immediately suggest to encode only the typical sequences and to throw away or code non-typical ones into a junk state. Because of (8.4) this scheme will incur a decoding error with probability at most δ. Because of (8.5) it is enough to use $n(H(X) + \epsilon)$ nats for the encoding. Moreover because of (8.6) using less than $n(H(X) - \epsilon)$ nats will incur a finite probability of error[1].

In the quantum case one defines a similar notion of typicality. Consider a memoryless source that outputs with probability p_x letters $|\phi\rangle_x \in \mathcal{A}$ which belong to the Hilbert space $\mathcal{H} = \mathbf{C}^d$. The density matrix for the source is

$$\sum_{x_1 ... x_n} p_{x_1} ... p_{x_n} |\phi_{x_1}\rangle\langle\phi_{x_1}| \otimes ... \otimes |\phi_{x_n}\rangle\langle\phi_{x_n}| = \Big(\sum_x p_x |\phi_x\rangle\langle\phi_x|\Big)^{\otimes n} = \rho^{\otimes n} \tag{8.8}$$

One can find the spectral decomposition of this density matrix. Indeed let

$$\rho = \sum_a \lambda_a P_a \tag{8.9}$$

[1] See Cover and Thomas for more details

be the spectral decomposition for the length one case (here we assume for simplicity non-degeneracy of the eigenvalues, so $P_a = |a\rangle\langle a|$). Then

$$\rho^{\otimes n} = \sum_{a_1 \ldots a_n} \lambda_{a_1} \ldots \lambda_{a_n} P_{a_1} \otimes \ldots \otimes P_{a_n} \tag{8.10}$$

The eigenvalues λ_a are positive and sum to one, thus define a probability distribution. Moreover the projectors P_a are mutually orthogonal, thus distinguishable. Therefore the density matrix $\rho^{\otimes n}$ is also the density matrix of a "mathematical" *memoryless classical source* that outputs letters a (or P_a or $|a\rangle$) with probabilities p_a. We stress that this is not the physical preparation of the state $\rho^{\otimes n}$. We can define a set of typical sequences of eigenvalues and/or eigenstates

$$\mathcal{T}_{n,\epsilon} = \left\{ a_1 \ldots a_n | \left| \frac{1}{n} \sum_{i=1}^{n} \log_d \frac{1}{\lambda_{a_i}} - S(\rho) \right| \leq \epsilon \right\} \tag{8.11}$$

Definition: typical subspace. *Consider the projector*

$$P_{n,\epsilon} = \sum_{a_1 \ldots a_n \in \mathcal{T}_{n,\epsilon}} P_{a_1} \otimes \ldots \otimes P_{a_n} \tag{8.12}$$

The subspace $P_{n,\epsilon}\mathcal{H}^{\otimes n}$ is called the typical subspace. We have

$$\rho^{\otimes n} = P_{n,\epsilon}\rho^{\otimes n}P_{n,\epsilon} + (I - P_{n,\epsilon})\rho^{\otimes n}(I - P_{n,\epsilon}) \tag{8.13}$$

The source coding scheme described in the next section is based on the following theorem, which is the quantum analog of (8.3), (8.5) and (8.6).

THEOREM 8.1.1 **[typical subspace theorem]** *Fix ϵ and δ positive, small. For n sufficiently large,*

- *the density matrix has almost all its support on the typical subspace*

$$1 - \delta \leq Tr P_{n,\epsilon}\rho^{\otimes n} \leq 1, \tag{8.14}$$

- *the dimension of the typical subspace is approximately $d^{nS(\rho)}$*

$$(1 - \delta)d^{n(S(\rho)-\epsilon)} \leq Tr P_{n,\epsilon} \leq d^{n(S(\rho)+\epsilon)}, \tag{8.15}$$

- *let $S_{n,\epsilon}$ be a projector on a subspace of dimension less than d^{nR} with $R \leq S(\rho) - \epsilon$. In other words $Tr S_{n,\epsilon} \leq d^{nR}$ with $R \leq S(\rho) - \epsilon$. For such a projector we have*

$$Tr S_{n,\epsilon}\rho^{\otimes n} \leq \delta + d^{-n(S(\rho)-\epsilon-R)}. \tag{8.16}$$

Proof The basic difference with the classical case is that one has to deal a bit more carefully with operator inequalities for the third statement[2]

[2] We recall: a hermitian matrix $A = A^\dagger$ is said to be (semi-definite) positive iff $\langle\phi|A|\phi\rangle \geq 0$ for any $|\phi\rangle$; $A \geq B$ iff $(A - B) \geq 0$; and $A \geq 0$ implies $C^\dagger AC \geq 0$.

First statement. Observe that

$$P_{n,\epsilon}\rho^{\otimes n}P_{n,\epsilon} = \sum_{a_1\ldots a_n\in\mathcal{T}_{n,\epsilon}} \lambda_{a_1}\ldots\lambda_{a_n}P_{a_1}\otimes\ldots\otimes P_{a_n} \tag{8.17}$$

So

$$TrP_{n,\epsilon}\rho^{\otimes n} = \sum_{a_1\ldots a_n\in\mathcal{T}_{n,\epsilon}} \lambda_{a_1}\ldots\lambda_{a_n} \tag{8.18}$$

which is the probability of the set $T_{n,\epsilon}$. The statement follows by the law of large numbers (as in the classical case).

Second statement. Observe that

$$TrP_{n,\epsilon} = \sum_{a_1\ldots a_n\in\mathcal{T}_{n,\epsilon}} 1 = |\mathcal{T}_{n,\epsilon}| \tag{8.19}$$

so the statement again follows like in the classical case. Note that here we have assumed that the eigenvalues are not degenerate (if $TrP_a = g_a$ we have to modify the definition of typical sequences according to $\frac{1}{\lambda_a} \to \frac{g_a}{\lambda_a}$).

Third statement. We use the decomposition (8.13) to write $TrS_{n,\epsilon}\rho^{\otimes n}$ as a sum of two contributions.

For the first one we have,

$$\begin{aligned} TrS_{n,\epsilon}P_{n,\epsilon}\rho^{\otimes n}P_{n,\epsilon} &= TrS_{n,\epsilon}P_{n,\epsilon}\rho^{\otimes n}P_{n,\epsilon}S_{n,\epsilon} \\ &\leq d^{-n(S(\rho)-\epsilon)}TrS_{n,\epsilon} \\ &\leq d^{-n(S(\rho)-\epsilon-R)} \end{aligned} \tag{8.20}$$

In the first equality we use the cyclicity of the trace and for the first inequality we use the operator inequality $P_{n,\epsilon}\rho^{\otimes n}P_{n,\epsilon} \leq d^{-n(S(\rho)-\epsilon)}\, I$.

For the second contribution we observe that $M = (I - P_{n,\epsilon})\rho^{\otimes n}(I - P_{n,\epsilon})$ is a positive operator so by the cyclicity of the trace

$$\begin{aligned} TrS_{n,\epsilon}(I - P_{n,\epsilon})\rho^{\otimes n}(I - P_{n,\epsilon}) &= Tr\sqrt{M}S_{n,\epsilon}\sqrt{M} \\ &\leq Tr\sqrt{M}\, I\, \sqrt{M} = Tr\, M \\ &= Tr\rho^{\otimes n}(I - P_{n,\epsilon}) \\ &\leq \delta \end{aligned} \tag{8.21}$$

In the inequality we used that $S_{n,\epsilon} \leq I$ (true for any projector). □

8.2 Compression scheme

The source outputs words of length n,

$$|\phi_{x_1}\rangle \otimes \ldots \otimes |\phi_{x_n}\rangle \tag{8.22}$$

with probability $p_{x_1}\ldots p_{x_n}$. We specify a bock coding scheme for these words: we would like to encode these words which belong to the Hilbert space $\mathcal{H}^{\otimes n}$ by

states in a Hilbert space $\mathcal{H}^{\otimes nR}$ were $R < 1$. This encoding should be faithful in the sense that that it should be possible to recover, most of the time, the original words by some decoding procedure.

Encoding procedure. We have to rely on a slightly more general encoding process that encodes source states into density matrices. An encoding map is a map from states of dimension d^n to density matrices of dimension $d^{nR} \times d^{nR}$

$$\begin{aligned} \mathcal{E}_n : \mathcal{H}^{\otimes n} &\to \mathcal{DM}(\mathcal{H}^{\otimes nR}) \\ |\phi_{x_1}\rangle \otimes ... \otimes |\phi_{x_n}\rangle &\to \mathcal{E}(|\phi_{x_1}\rangle \otimes ... \otimes |\phi_{x_n}\rangle) \end{aligned}$$

Here $\mathcal{DM}(\mathcal{H}^{\otimes nR})$ is the space of density matrices of dimension $d^{nR} \times d^{nR}$. The compression rate per letter is $R = \frac{nR}{n}$.

Decoding procedure. Ideally we should map back the density matrix $\mathcal{E}(|\phi_{x_1}\rangle \otimes ... \otimes |\phi_{x_n}\rangle)$ to the input word $|\phi_{x_1}\rangle \otimes ... \otimes |\phi_{x_n}\rangle$. This cannot be done exactly, so we allow for a slightly more general definition,

$$\begin{aligned} \mathcal{D}_n : \mathcal{DM}(\mathcal{H}^{\otimes nR}) &\to \mathcal{DM}(\mathcal{H}^{\otimes n}) \\ \sigma &\to \mathcal{D}(\sigma) \end{aligned}$$

Reliability criterion. The scheme $(\mathcal{E}_n, \mathcal{D}_n)$ should be faithful. Let

$$\rho_{\text{output}} = \mathcal{D}(\mathcal{E}(|\phi_{x_1}\rangle \otimes ... \otimes |\phi_{x_n}\rangle)) \tag{8.23}$$

We define a *fidelity* as the overlap between the input and output states,

$$F(|\phi_{x_1}\rangle \otimes ... \otimes |\phi_{x_n}\rangle) = \langle \phi_{x_1}, ..., \phi_{x_n} | \rho_{\text{output}} | \phi_{x_1}, ..., \phi_{x_n} \rangle \tag{8.24}$$

The average fidelity is

$$\bar{F}_n = \sum_{x_1,...,x_n} p_{x_1} ... p_{x_n} \langle \phi_{x_1}, ..., \phi_{x_n} | \rho_{\text{output}} | \phi_{x_1}, ..., \phi_{x_n} \rangle \tag{8.25}$$

One has to be careful with the notation here: in each term of the sum ρ_{output} depends on the input state $|\phi_{x_1}, ..., \phi_{x_n}\rangle$.

The intuitive meaning of the fidelity can be better understood by looking at the classical case. If the letters $|\phi_x\rangle$ are orthonormal then we are reduced to a classical situation and the encoding-decoding operations can be done by looking at a "look-up table". In other words for a typical source word we have perfect recovery so $\rho_{\text{output}} = |\phi_{x_1}, ..., \phi_{x_n}\rangle\langle\phi_{x_1}, ..., \phi_{x_n}|$ and $F = 1$; while for a non typical source word we have $\rho_{\text{output}} = |\text{junk}\rangle\langle\text{junk}|$ and the decoder simply declares an error and sets $F = 0$ (simply assume that $|\text{junk}\rangle$ is orthogonal to all source words). We see that in the classical case the average fidelity is precisely equal to $1 - \text{Prob(error)}$.

THEOREM 8.2.1 **[Schumacher's theorem]** *Fix $\epsilon > 0$, $\delta > 0$ small.*

- *Fix $R > S(\rho) + \epsilon$. Then one can find encoding-decoding schemes $(\mathcal{E}_n, \mathcal{D}_n)$ such*

that for n large enough $\bar{F}_n \geq 1-2\delta$. *So asymptotically loss-less compression is possible.*

- *Fix* $R < S(\rho) - \epsilon$. *Then for any encoding-decoding scheme* $(\mathcal{E}_n, \mathcal{D}_n)$ *we have* $\bar{F}_n \leq \delta + d^{-n(S(\rho)-\epsilon-R)}$. *Loss-less compression is not possible.*

This theorem says that compression rates above $S(\rho)$ are (faithfully) achievable, while this is not the case for compression rates below $S(\rho)$. Note that in general $S(\rho) \leq H(X)$. The fact that a quantum source is more compressible than a classical one should not surprise the reader: this is an expression of the fact that non-orthogonal alphabet letters cannot be perfectly distinguished so that a quantum source word is more redundant than its classical counterpart.

8.3 Proof of the source coding theorem

First we prove the achievability part and then proceed to the converse.

Achievability part. We specify the encoding map $\mathcal{E}$. Take the measurement apparatus defined by the two orthogonal projectors $\{P_{n,\epsilon}, I - P_{n,\epsilon}\}$ on the typical subspace and its orthogonal complement. Given a source word $|\phi_{x_1}, ..., \phi_{x_n}\rangle$ perform a measurement. According to the measurement postulate the outcome is

$$\frac{P_{n,\epsilon}|\phi_{x_1}, ..., \phi_{x_n}\rangle}{\langle\phi_{x_1}, ..., \phi_{x_n}|P_{n,\epsilon}|\phi_{x_1}, ..., \phi_{x_n}\rangle^{1/2}}, \text{ with prob } \langle\phi_{x_1}, ..., \phi_{x_n}|P_{n,\epsilon}|\phi_{x_1}, ..., \phi_{x_n}\rangle \tag{8.26}$$

or

$$\frac{(I - P_{n,\epsilon})|\phi_{x_1}, ..., \phi_{x_n}\rangle}{\langle\phi_{x_1}, ..., \phi_{x_n}|I - P_{n,\epsilon}|\phi_{x_1}, ..., \phi_{x_n}\rangle^{1/2}}, \text{ with prob } \langle\phi_{x_1}, ..., \phi_{x_n}|I - P_{n,\epsilon}|\phi_{x_1}, ..., \phi_{x_n}\rangle \tag{8.27}$$

Now the first state is in the typical subspace $P_{n,\epsilon}\mathcal{H}^{\otimes n}$ so it can be described by $nS(\rho)$ quantum nats (because of theorem 1 the dimension of the typical subspace is $d^{nS(\rho)}$). One can find a basis of $\mathcal{H}^{\otimes n}$ such that this typical subspace is described by the first $nS(\rho)$ terms of the tensor product. In other words we can find a unitary operation U that transforms the state (8.26) to the form (this unitary depends only on the original basis and the typical space, not on the particular input state)

$$\sum_{b_1...b_m} c^{b_1...b_n}_{x_1...x_n} | \underbrace{b_1...b_m}_{nR \text{ terms}}, \underbrace{0, 0, ..., 0}_{n(1-R) \text{ terms}} \rangle = |\psi_{\text{compressed}}\rangle \otimes | \underbrace{0, 0, ..., 0}_{n(1-R) \text{ terms}} \rangle \tag{8.28}$$

The state $|0_{n(1-R)}\rangle$ is then discarded. The second possible outcome is not coded since it lies in the non typical subspace. More precisely we describe all such states as $|\text{junk}\rangle$ a single specified quantum state (in the typical subspace, say).

We assume that the outcome is not observed during the compression stage so its state is described by the mixture

$$\begin{aligned}\mathcal{E}(|\phi_{x_1}...\phi_{x_n}\rangle) =& \langle\phi_{x_1},...,\phi_{x_n}|P_{n,\epsilon}|\phi_{x_1},...,\phi_{x_n}\rangle|\psi_{\text{compressed}}\rangle\langle\psi_{\text{compressed}}| \\ &+ \langle\phi_{x_1},...,\phi_{x_n}|I-P_{n,\epsilon}|\phi_{x_1},...,\phi_{x_n}\rangle|\text{junk}\rangle\langle\text{junk}| \end{aligned} \tag{8.29}$$

For the decoding operation one first appends $n(1-R)$ quantum letters in the $|0_{n(1-R)}\rangle$ state that was discarded. then one performs the inverse unitary operation $U^\dagger$. So the decoder map is given by

$$\begin{aligned}&\mathcal{D}(\mathcal{E}(|\phi_{x_1}...\phi_{x_n}\rangle)) \\ &= P_{n,\epsilon}|\phi_{x_1}...\phi_{x_n}\rangle\langle\phi_{x_1},...,\phi_{x_n}|P_{n,\epsilon} \\ &+ \langle\phi_{x_1},...,\phi_{x_n}|I-P_{n,\epsilon}|\phi_{x_1},...,\phi_{x_n}\rangle U^\dagger\,|\text{junk},0_{n(1-R)}\rangle\langle\text{junk},0_{n(1-R)}|\,U\end{aligned} \tag{8.30}$$

We now estimate the fidelity associated to this scheme $(\mathcal{E}_n,\mathcal{D}_n)$. We replace ρ_{output} given by (8.30) in the definition of the average fidelity. The contribution from the first term is

$$\begin{aligned}&\sum_{x_1...x_n} p_{x_1}...p_{x_n}\langle\phi_{x_1},...,\phi_{x_n}|P_{n,\epsilon}|\phi_{x_1}...\phi_{x_n}\rangle^2 \\ &\geq \left\{\sum_{x_1...x_n} p_{x_1}...p_{x_n}\langle\phi_{x_1},...,\phi_{x_n}|P_{n,\epsilon}|\phi_{x_1}...\phi_{x_n}\rangle\right\}^2 \\ &= (Tr\rho^{\otimes n}P_{n,\epsilon})^2 \\ &\geq (1-\delta)^2\end{aligned} \tag{8.31}$$

The first inequality is Cauchy-Schwartz, and the second comes from theorem 1. Finally the contribution from the second term is trivially positive (write it down and see !). Thus we conclude that

$$\bar{F} \geq (1-\delta)^2 \geq 1-2\delta \tag{8.32}$$

Converse part. Let

$$\mathcal{E}_N : \ |\phi_{x_1}...\phi_{x_n}\rangle\langle\phi_{x_1}...\phi_{x_n}| \ \rightarrow \ \sigma \tag{8.33}$$

be a completely general encoding scheme (so σ is any $d^{nR}\times d^{nR}$ density matrix). The first step of the decoder is to append $|0_{n(1-R)}\rangle\langle 0_{n(1-R)}|$ to get a state

$$\sigma\otimes|0_{n(1-R)}\rangle\langle 0_{n(1-R)}| \tag{8.34}$$

in the original Hilbert space. Here we restrict the proof to the special case of *unitary decoders*[3]. So let

$$\mathcal{D} : \ \sigma\otimes|0_{nR}\rangle\langle 0_{nR}| \ \rightarrow \ U\sigma\otimes|0_{nR}\rangle\langle 0_{nR}|U^\dagger \tag{8.35}$$

[3] More general ones would correspond to a mappings between density matrices and would require a more complicated proof.

The density matrix (8.34) is constructed out of states of a d^{nR} dimensional subspace of $\mathcal{H}^{\otimes n}$. Let S_n be the projector on that subspace, and note that

$$U\sigma \otimes |0_{nR}\rangle\langle 0_{nR}|U^\dagger = US_n\Big(\sigma \otimes |0_{nR}\rangle\langle 0_{nR}|\Big)S_nU^\dagger \tag{8.36}$$

Now, the average fidelity is

$$\begin{aligned}\bar{F} &= \sum_{x_1\ldots x_n} p_{x_1}\ldots p_{x_n}\langle\phi_{x_1}\ldots\phi_{x_n}|US_n\Big(\sigma \otimes |0_{nR}\rangle\langle 0_{nR}|\Big)S_nU^\dagger|\phi_{x_1}\ldots\phi_{x_n}\rangle \qquad (8.37)\\ &\leq \sum_{x_1\ldots x_n} p_{x_1}\ldots p_{x_n}\langle\phi_{x_1}\ldots\phi_{x_n}|US_nU^\dagger|\phi_{x_1}\ldots\phi_{x_n}\rangle \\ &= Tr(\rho^{\otimes n}US_nU^\dagger)\end{aligned}$$

We first used that any density matrix is smaller than the identity matrix, so $\sigma \otimes |0_{nR}\rangle\langle 0_{nR} \leq I$, and then the cyclicity of the trace. Clearly $US_nU^\dagger$ is a projector on some d^{nR} dimensional subspace of $\mathcal{H}^{\otimes n}$ with $R < S(\rho) - \epsilon$. Then, the third statement of theorem 1 implies

$$\bar{F} \leq \delta + d^{-n(S(\rho)-\epsilon-R)} \tag{8.38}$$

This achieves the proof of the converse part for the class of unitary decoders.

9 Capacities of Quantum Channels

Part III

Quantum Computation

10 Quantum Computation and Circuit Model

10.1 Brief historical introduction

A computation is ultimately performed by a physical device. Then it is a natural question to ask what are the fundamental limitations that the laws of physics impose on a computation. An early work on such issues was that of Landauer who argued that a bit erasure - a logically irreversible process - is always accompanied by heat dissipation and is thus a thermodynamically irreversible process. Consequently any computation using logically irreversible gates (AND; OR) will dissipate heat. But is there a fundamental principle that requires a minimum amount of heat dissipation in computation? A negative answer to this question was put forward by Bennett, Benioff and others. More precisely any logically irreversible computation can be made logically reversible, with appropriate elementary gates, provided we are willing to increase the work space. And there is no physical principle that requires a minimum amount of heat dissipation for logically reversible computation.

If no heat is generated by reversible computations, then as the physical support of bits becomes smaller and smaller the quantum mechanical behavior of matter and notably quantum coherence effects might become important. It is natural to ask the question: what are the effects of the quantum mechanical behavior of matter on computation? Do they help, or do they bring any new limitations?

These issues where raised and discussed by Feynman, Benioff, Manin and others in the early 1980's. In principle QM does not bring any new limitations. On the contrary the superposition principle applied to many particle systems (many qubits) enables us to perform "parallel computations", thereby speeding up classical computations. This was recognized very early by Feynman who argued that classical computers cannot simulate efficiently quantum mechanical processes. This lead him to suggest that we should buid "quantum machines" to simulate efficiently quantum processes.

The basic reason why classical computation cannot simulate efficiently quantum processes is the following. A general quantum state for N quantum bits involves a superposition of 2^N classical states :

$$|\psi\rangle = \sum_{b_1 \ldots b_N \in \ \{0,1\}^N} c_{b_1,\ldots,b_N} |b_1, \ldots, b_N\rangle$$

A classical simulation of the evolution of $|\psi\rangle$ must perform essentially 2^N computations for the evolution of each "classical state" $|b_1 \dots b_N\rangle$ (states of the computational basis are mutually othogonal and can be considered as "classical"). On the contrary, the unitary quantum dynamics U acts on $|\psi\rangle$ as a whole (or on each $|b_1 \dots b_N\rangle$ in parallel). So physical devices performing a unitary dynamics on $|\psi\rangle$ can be viewed as devices performing a quantum computation.

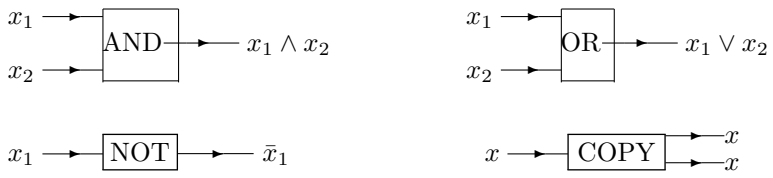
Feymman developed the concept of quantum computation in the language of Hamiltonian dynamics. It turns out that this is not very practical. A practical classical model of computation (that has an intuitively clear quantum counterpart) can be represented by a circuit model built out of a given set of elementary gates acting in a recursive way on the input of the computation. Around 1985 David Deutsch showed that the same holds in the quantum case. Namely, any unitary evolution can be approximated well enough by some set of universal elementary quantum gates.

Nowaday it is the Deutsch model - a quantum circuit model - of a quantum computer that is the most popular. The subject of this chapter is to explain this model. One reason for its popularity is that it is a universal model: in principle any quantum computation can be cast as an instance of a quantum circuit.

There is also a notion of quantum Turing machine (which is analogous to classical Turing machines) and is the natural and most convenient framework to discuss quantum complexity classes. It has been shown that the Quantum Turning machine model is equivalent to the quantum circuit model. Complexity issues are (almost) not discussed here.

10.2 Classical circuit model of Computation

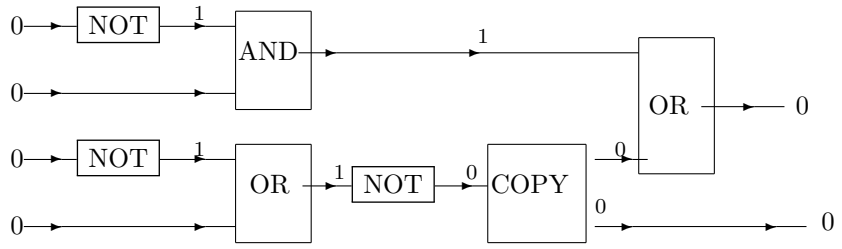
We begin with a discussion of classical computations done with classical boolean circuits. Consider the basis set of logical gates acting on bits $x_i \in \mathbb{F}_2 = \{0, 1\}$.



Note that COPY is also sometimes called FANOUT. We can use these logical gates to define a boolean circuit.

DEFINITION 10.2.1 A boolean circuit is a directed acyclic graph with n input bits and m output bits. The input can always be initialized to $(0 \dots 0)$ because any $(x_1 \dots x_m)$ is obtained by a series of appropriate NOT gates.

The following figure is an example illustrating this definition.



A celebrated theorem of Emil Post (circa 1950) says that for any function $f : \mathbb{F}_2^m \to \mathbb{F}_2^m$ one can construct a Boolean circuit that computes it.

THEOREM 10.2.2 *For any function $f : \mathbb{F}_2^m \to \mathbb{F}_2^m$ there exists a Boolean circuit that maps inputs $(x_1 \dots x_m)$ to outputs $(y_1 \dots y_m) = f(x_1 \dots x_m)$. The Boolean circuit is constructed out of* NOT, AND, OR, COPY *and is a directed acyclic graph.*

One says that the set of gates $\{\text{NOT}, \text{AND}, \text{OR}, \text{COPY}\}$ is universal. Note that AND, OR are not reversible. We come back later to the issue of reversibility.

Proof A function $f : \mathbb{F}_2^m \to \mathbb{F}_2^m$ can be represented by component functions $f_i : \mathbb{F}_2^m \to \mathbb{F}_2$, $i = 1 \dots m$. So if we can COPY the input, m times we just need to show that there exists a Boolean circuit for each $f_i : \mathbb{F}_2^m \to \mathbb{F}_2$. The problem is then reduced to finding Boolean circuits for functions $f : \mathbb{F}_2^m \to \mathbb{F}_2$.

For each $\vec{a} = (a_1 \dots a_m) \in \mathbb{F}$ we define

$$C_{\vec{a}}(x_1 \dots x_m) = \phi_1(x_1) \wedge \phi_2(x_2) \wedge \dots \wedge \phi_m(x_m)$$

where

$$\begin{cases} \phi_i(x_i) = \bar{x}_i \text{ if } a_i = 0 \\ \phi_i(x_i) = x_i \text{ if } a_i = 1 \end{cases}$$

This is built out of AND, NOT gates only, and since $\wedge$ is associative it can be done recursively (directed acyclic graph). We note that $C_{\vec{a}}(x_1 \dots x_m) = 1$ if $(x_1 \dots x_m) = (a_1 \dots a_m)$.

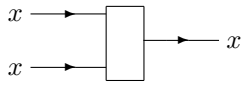
Now given a function $f : \mathbb{F}_2^m \to \mathbb{F}_2$ let $\{\vec{a}^{(1)}, \dots, \vec{a}^{(k)}\}$ be the set of inputs in $\mathbb{F}_2^m$ for which f takes values 1. For all other input f takes value 0. A little thought shows that

$$f(x_1 \dots x_m) = C_{\vec{a}^{(1)}}(x_1 \dots x_m) \vee C_{\vec{a}^{(2)}}(x_1 \dots x_m) \vee \dots \vee C_{\vec{a}^{(K)}}(x_1 \dots x_m)$$

It remains to see that $\vee$ is associative and can thus be done in a recursive way (i.e with a directed acyclic graph). So f is computed from OR and COPY. $\square$

10.3 Reversibility versus irreversibility

The NOT gate is logically reversible. This means that from the output one can recover the input. However the AND, OR gates are logically irreversible, and this implies heat dissipation. The COPY operation is logically reversible. However its naive implementation (the FANOUT) is thermodynamically irreversible. Indeed the reversed operation



erases a bit. Thus, according to Landauer's analysis, there is heat dissipation and the gate is not thermodynamically reversible. However we will see below that the agte can be implemented in a way that is also thermodynamically reversible.

We will now show that any Boolean circuit can be simulated by a *logically* reversible circuit. Moreover a universal set of reversible gates exists.

From $f : \mathbb{F}_2^m \to \mathbb{F}_2$ we construct $\tilde{f} : \mathbb{F}_2^m \oplus \mathbb{F}_2 \to \mathbb{F}_2^m \oplus \mathbb{F}_2$ as follow :

$$\tilde{f}(x_1 \dots x_m, y) = (x_1 \dots x_m, f(x_1 \dots x_m) \oplus y)$$

Now, $\tilde{f}$ is invertible since from $(x_1 \dots x_m, f(x_1 \dots x_m) \oplus y)$ we can recover $x_1 \dots x_m$ and $f(x_1 \dots x_m)$ [since we have a circuit for f] and then

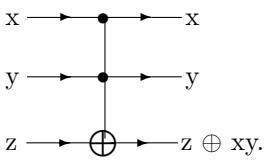
$$y = (f(x_1 \dots x_m) \oplus y) \oplus f(x_1 \dots x_m).$$

Thus any f can be computed in a reversible way from the circuit for $\tilde{f}$. To compute f reversibly we start with the input $(x_1 \dots x_m, 0)$ compute $\tilde{f}(x_1 \dots x_m, 0) = (x_1 \dots x_m, f(x_1 \dots x_m))$ copy[1] the last bit $f(x_1 \dots x_m)$ and run back the computation to get

$$\tilde{f}^{-1}(x_1 \dots x_m, f(x_1 \dots x_m)) = (x_1 \dots x_m, 0).$$

In this way we have $f(x_1 \dots x_m)$ and the circuit is left in its initial state $(x_1, \dots, x_m, 0)$.

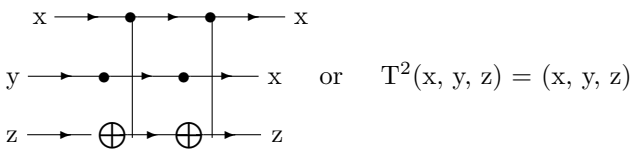
It remains to be shown that $\tilde{f}$ can be represented by a circuit containing only irreversible elementary gates. We already know that $\tilde{f}$ can be represented by a circuit containing AND, OR, NOT, COPY. We want to replace AND, OR by a reversible gate. This can be achieved by using the 3 bit gate known as *Toffoli gate* which is a CCNOT (controlled-controlled NOT):



or $\quad \mathrm{T}(x, y, z) = (x, y, z \oplus xy).$

[1] This step is logically reversible but not necessarily thermodynamically so. However as explained below it can be made thermodynamically reversible by a proper implementation of COPY

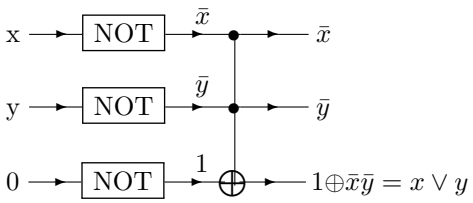
This gate flips the target bit z if both control bits x and y are equal to 1. Otherwise z is left unchanged. The Toffoli gate is reversible because:



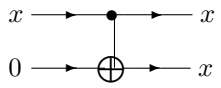
or $\mathrm{T}^2(\mathrm{x}, \mathrm{y}, \mathrm{z}) = (\mathrm{x}, \mathrm{y}, \mathrm{z})$

If we set $z = 0$, $T(x, y, 0) = (x, y, xy)$ outputs $x \wedge y$ in the third bit. Thus the AND gates can be replaced by a Toffoli gate provided we increase our workspace to have a target input bit $z = 0$ and the additional outputs x & y.

For the OR gate we can use



Finaly the COPY gate can be replaced by



which is a CNOT gate (reversible) with the target bit set to 0 in the input.

Summarizing we have shown that a Boolean circuit made of the universal set $\{\mathrm{AND}, \mathrm{OR}, \mathrm{COPY}, \mathrm{NOT}\}$ can be simulated by a reversible circuit made of the universal set $\{\mathrm{CNOT}; \mathrm{Toffoli}; \mathrm{NOT}\}$.

The set $\{\mathrm{AND}, \mathrm{OR}, \mathrm{COPY}, \mathrm{NOT}\}$ involves single and two bit gates. On the other hand $\{\mathrm{CNOT}; \mathrm{Toffoli}; \mathrm{NOT}\}$ involves single, two bit, and the Toffoli three bit gate. Is it possible to build reversible circuits using only single and two bit gates? It is possible to show that the answer to this question is no. In fact it suffices to produce a counterexample: the Toffoli gate cannot be simulated reversibly with single & two bit gates. We will see that (perhaps surprisingly) in the quantum case single and two bit gates suffice for reversible computation.

10.4 Deutsch model of quantum circuits.

As we will see the quantum circuits are built out of a small set of gates. We fisrt start by listing a few of the most important gates that we will encounter.

10.4.1 Single Qbit gates

- The three Pauli-gates $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$; $Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ and $Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

  $|b\rangle \longrightarrow \boxed{\text{X}} \longrightarrow |\bar{b}\rangle$ is the quantum NOT gate.

  Quantum mechanicaly the input can also be a coherent superposition of the states $\{|0\rangle, |1\rangle\}$. For example

  $$X(\alpha|0\rangle + \beta|1\rangle) = \alpha|1\rangle + \beta|0\rangle.$$

- The Hadamard gate $H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$

  $|b\rangle \longrightarrow \boxed{\text{H}} \longrightarrow H|b\rangle = \frac{1}{\sqrt{2}}(|0\rangle + (-1)^b|1\rangle)$

- The "$\frac{\pi}{8}$ gate" $T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}$

  $|b\rangle \longrightarrow \boxed{\text{T}} \longrightarrow e^{ib\pi/4}|b\rangle = \begin{cases} |0\rangle \to |0\rangle \\ |1\rangle \to e^{i\pi/4}|1\rangle \end{cases}$

  for superpositions: $\alpha|0\rangle + \beta|1\rangle \to \alpha|0\rangle + \beta e^{i\pi/4}|1\rangle$

- The gate $S = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$ $\quad \begin{cases} |0\rangle \to |0\rangle \\ |1\rangle \to i|1\rangle. \end{cases}$

An important Lemma that we give here without proof is

Lemma 10.4.1 (Approximation of single Qbit gates by H and T.) Any single bit unitary U can be approximated to arbitrary precision δ by a concatenation of "Hadamard H" and "$\frac{\pi}{8} - T$ gates". Moreover if V is the concatenation of H and T gates approximating U and $||U - V|| < \delta$ we need at most $O(\ln \delta)$ gates H and T [This last statement is known as the Solevay-Kitaev theorem].

The main idea of the proof is to represent U by a succession of rotations, themselves represented by Pauli-gates, themselves represented by H and T. It is not very difficult to prove that a circuit size $O(1/\delta)$ is sufficient. The Solovay-Kitaev $O(\ln \delta)$ is more difficult.

10.4.2 Controlled two-bit gates.

- The CNOT gate (controlled not) is the prototypical two-bit gate:

  $|c\rangle \longrightarrow \bullet \longrightarrow |c\rangle$ control bit.

  $|t\rangle \longrightarrow \oplus \longrightarrow |c \oplus t\rangle$ target bit.

In the basis

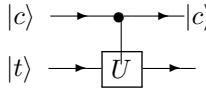
$$|0\rangle \otimes |0\rangle\ ; \quad |0\rangle \otimes |1\rangle\ ; \quad |1\rangle \otimes |0\rangle\ ; \quad |1\rangle \otimes |1\rangle$$

$$\begin{pmatrix}1\\0\\0\\0\end{pmatrix} \quad \begin{pmatrix}0\\1\\0\\0\end{pmatrix} \quad \begin{pmatrix}0\\0\\1\\0\end{pmatrix} \quad \begin{pmatrix}0\\0\\0\\1\end{pmatrix}$$

the matrix representation is

$$CNOT = \begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&0&1\\0&0&1&0\end{pmatrix}$$

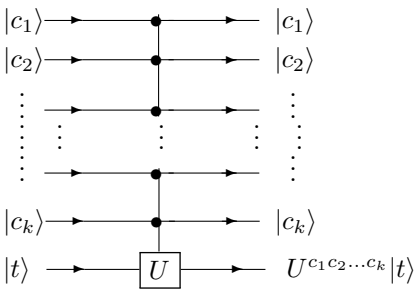
- The controlled-U gate where U is a single bit operation.



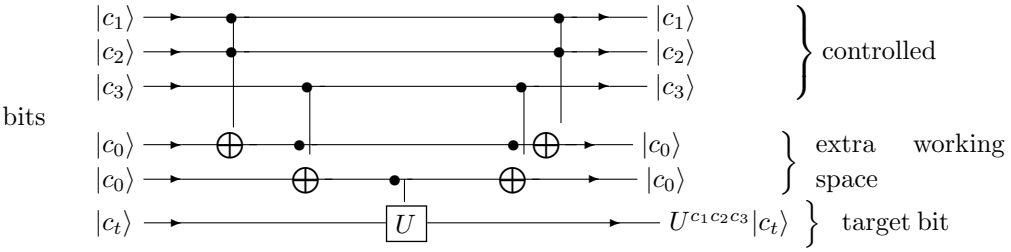
$$U^c|t\rangle = \begin{cases} |t\rangle \text{ if } c = 0 \\ U|t\rangle \text{ if } c = 1 \end{cases}$$

10.4.3 Multi-controlled gates.

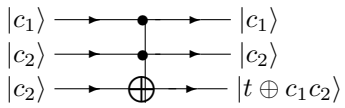
A generalization of the preview controlled-U gate is the multi-controlled-U



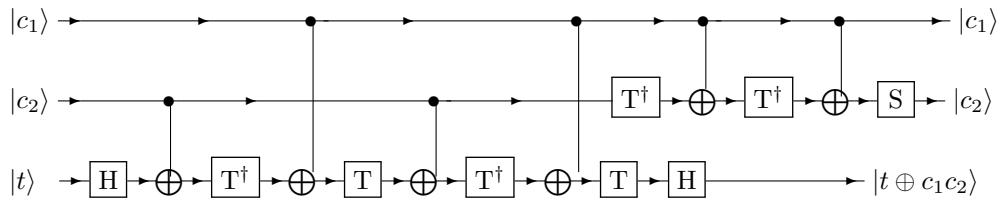
So U acts on the target bit if and only if all control bits are set to 1. By increasing the work space this gate can be represented by a concatenation of controlled-controlled-NOT and a controlled U. Indeed:



The controlled-controlled-NOT gate is nothing else than the Toffoli gate. Remarkably it can be represented by one and two-bit quantum gates $\{T, S, H, CNOT\}$. Remember that classicaly this is not possible! The reader can check that:



is equivalent to the following circuit:



Summarizing we arrive at the following Lemma:

Lemma 10.4.2 Any multibit-Controlled single bit gate U acting on N Qbits ($2^N \times 2^N$ matrix), can be represented by the set $\{T, S, H, CNOT, U\}$ where U acts on the last Qbit (2×2 matrix).

10.4.4 A universal set of quantum gates and the circuit model

An important lemma that we give here without proof is:

Lemma 10.4.3 Any unitary U acting on N-Qbits states, i.e states in $\mathbb{C}^2 \otimes \mathbb{C}^2 ... \otimes \mathbb{C}^2$ can be decomposed as a finite product of "two qubit unitaries":

$$U = U^{(i_1 j_1)} \otimes U^{(i_2 j_2)} ... \, U^{(i_K j_K)}$$

where $U^{(ij)}$ acts from $\mathbb{C}^2 \otimes ... \otimes \mathbb{C}^2 \to \mathbb{C}^2 \otimes ... \otimes \mathbb{C}^2$ on spaces i and j (and trivialy on all other spaces).

Lemma 10.4.4 Any two level unitary $U^{(ij)}$ acting as U on bits i and j and as the identity on all others can be implemented by a concatenation of CNOT and a multicontrolled single bit U.

One can show that there exist N-Qbits unitary matrices ($2^N \times 2^N$ matrices) such that the decomposition obtained by lemmas 3 and 4 requires $O(e^N)$ gates. For

Combining Lemmas 1, 2, 3, 4 we arrive at the following basic theorem on which the quantum Circuit Model of quantum Computation is based:

THEOREM 10.4.5 *Any $2^N \times 2^N$ unitary matrix U acting on $\mathbb{C}_2 \otimes ... \otimes \mathbb{C}_2$ can be represented to arbitrary accuracy by a concatenation of the finite set of single and two-bit gates $\{T, S, H, CNOT\}$.*

If the required accuracy is δ one can argue that the maximal number of gates is of the form $O((\ln \delta)e^N)$. One can show that there exists unitary U for which it is not possible to have $O((\ln \delta)\text{poly}(N))$. On the other hand for some special problems such as factoring we will see that $O(\text{poly}N)$ suffices for the quantum circuit model (even though it does not suffice for a classical circuit).

The basic theorem just explained justifies the following model for quantum computation.

DEFINITION 10.4.6 (Definition of a quantum circuit) a. A quantum circuit is a directed acyclic graph whore vertices are gates among the finite set $\{T, S, H, CNOT\}$. The wires "carry" single Qbits $(\alpha|0\rangle + \beta|1\rangle)$.

b. The input is set to the simple product state

$$|0\rangle \otimes ... \otimes |0\rangle$$

c. The output is the result of the unitary evolution operating on the input. The output is in general a state of the form

$$|\Psi\rangle = \sum_{c_1...c_N} A(c_1...c_N)|c_1c_2...c_N\rangle$$

d. Finally a measurement is performed on $|\Psi\rangle$ with an apparatus measuring in the computational basis $\{|0\rangle, |1\rangle\}^{\otimes N}$. The outcome of the measurement is "the result of the computation" $|c_1...c_N\rangle$ obtained with probability $|A(c_1...c_N)|^2$.

A few remarks about this model are in order:

1. Acting on "Qtrits" instead of "Qbits" would not change anything fundamental (e.g the size or complexity of the circuit).
2. Performing measurements at intermediates stages instead of at the end does not change anything.
3. Performing measurements in another basis simply amounts to first unitarily rotate the basis so this can viewed as an adjunction to the circuit and finaly does not change anything. However this may add complexity.
4. Starting with another initial condition amounts to start from the $|0\rangle \otimes ... \otimes |0\rangle$ initial condition and adding extra unitary gates. However this may add complexity.
5. Other sets of universal gates exist. It may be surprising that in the classical care three bit gates are needed whereas this is not the case for quantum computation. But from a more physical point of view this is not surprising because the classical three bit gates can be viewed as an effect of "two-body interaction" [see Billiard-Ball-Model and Fredkim gate].
6. A quantum computation is reversible as long as the measurement has not been performed.

7. A reversible classical computation can be represented by a unitary operator. Indeed
$$\tilde{f}(x_1...x_N, y) = (x_1...x_N, y \oplus f(x_1...x_N))$$
induces the unitary
$$U_f|x_1...x_N, y\rangle = |x_1...x_N, y \oplus f(x_1...x_N)\rangle.$$
That U_f is a unitary, is easily checked by checking that it preserves the scalar product.
8. Any classical reversible computation is included in the model of quantum computation.
9. The power of quantum computation comes from the simultaneous action of the unitary evolution on all "classical" strings $|c_1...c_N\rangle$ of a many-qubit state. The complexity of the calculation is given by the size of the circuit. Since the result is obtained with some probability, typicaly one must repeat a certain number of times the computation to get a result with high probability (hopefully). This repetition may add to the complexity.

10.5 The Deutsch-Josza problem.

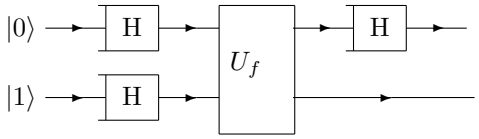
This problem illustrates nicely the notion of quantum parallelism and the power of quantum computation might be. We start with a special case due to Deutsch.

10.5.1 The Deutsch problem

We are given a black box representing $f : \{0, 1\} \to \{0, 1\}$ and are assured that f is either constant i.e $f(0) = f(1)$ or balanced $f(0) \neq f(1)$. The black box gives us an output when it is queried with an input. How many queries are needed to decide if f is constant or balanced? Classicaly we need two queries. Indeed we present the input 0 and get $f(0)$. Then we present the input 1 and get $f(1)$ and check wether $f(0) = f(1)$ or $f(0) \neq f(1)$.

We will show that quantum mechanically one query suffices! This is because we can present a query $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and get a global answer about f even though we do not learn what specific f is in the black box.

Consider the circuit:



[measurement of first bit].

Here H is the Hadamard gate and
$$U_f|x, y\rangle = |x, y \oplus f(x)\rangle.$$

Note that any U_f can be represented by $T, S, H, CNOT$ but this is not the point here; we view U_f as a black-box or Oracle which gives us an output for a given input.

We query the black-box with $|0\rangle \otimes |1\rangle$. Let us compute the action of the circuit step by step:

$$
\begin{aligned}
&(H \otimes \mathbf{1}) U_f (H \otimes H)|0\rangle \otimes |1\rangle \\
&\quad = (H \otimes \mathbf{1}) U_f (H|0\rangle \otimes H|1\rangle) \\
&\quad = (H \otimes \mathbf{1}) U_f \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \\
&\quad = (H \otimes \mathbf{1}) \frac{1}{2}(U_f|00\rangle - U_f|01\rangle + U_f|10\rangle + U_f|11\rangle) \\
&\quad = (H \otimes \mathbf{1}) \frac{1}{2}(|0 f(0)\rangle - |01 + f(0)\rangle + |1 f(1)\rangle + |11 + f(1)\rangle) \\
&\quad = (H \otimes \mathbf{1}) \frac{1}{2}((-1)^{f(0)}|00\rangle - (-1)^{f(0)}|01\rangle + ((-1)^{f(1)}|10\rangle - (-1)^{f(1)}|11\rangle \\
&\quad = (H \otimes \mathbf{1}) \frac{1}{2}((-1)^{f(0)}|0\rangle \otimes (|0\rangle - |1\rangle) + (-1)^{f(1)}|1\rangle \otimes (|0\rangle - |1\rangle)) \\
&\quad = (H \otimes \mathbf{1}) \frac{1}{2}((-1)^{f(0)}|0\rangle + (-1)^{f(1)}|1\rangle) \otimes (|0\rangle - |1\rangle) \\
&\quad = \frac{1}{2\sqrt{2}}((-1)^{f(0)}(|0\rangle + |1\rangle) + (-1)^{f(1)}(|0\rangle - |1\rangle)) \otimes (|0\rangle - |1\rangle) \\
&\quad = \{\frac{1}{2}((-1)^{f(0)} + (-1)^{f(1)})|0\rangle + \frac{1}{2}((-1)^{f(0)} - (-1)^{f(1)})|1\rangle\} \otimes \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)
\end{aligned}
$$

This last state is the output of the circuit. We may do a measurement of the first Qbit:

$$\text{Prob}(0) = \frac{1}{4}|(-1)^{f(0)} + (-1)^{f(1)}|^2 = \begin{cases} 1 \text{ if } f(0) = f(1) \\ 0 \text{ if } f(0) \neq f(1) \end{cases}$$

and

$$\text{Pro}(1) = \frac{1}{4}|(-1)^{f(0)} - (-1)^{f(1)}|^2 = \begin{cases} 0 \text{ if } f(0) = f(1) \\ 1 \text{ if } f(0) \neq f(1) \end{cases}$$

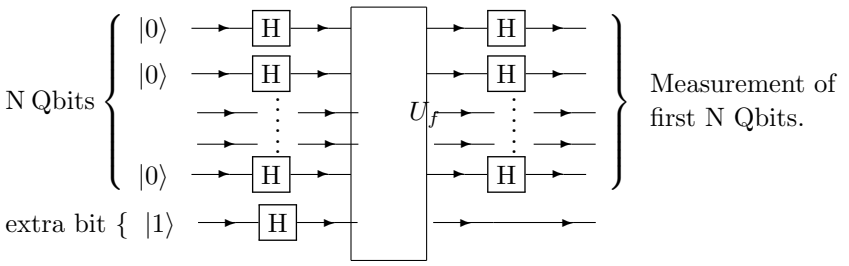
Thus if the function happens to balanced the measurement surely yields $|1\rangle$ and if it happens to be constant the measurement surely yields $|0\rangle$. Therefore with one "query" we can learn if f is constant or balanced.

10.5.2 The Deutsch-Josza Problem

This is a generalization of the previous problem to functions $f : \{0,1\}^N \to \{0,1\}$. We are assured that f is either constant or balanced. Here balanced means that it takes the value 0 for half of the inputs and the value 1 for the other half. If f happens to be constant, classically we have to query the black box $\frac{2^N}{2} + 1 = 2^{N-1} + 1$ times to determine if it is constant. On the other hand if f

happens to be balanced we need at least 2 queries and at most $\frac{2^N}{2}+1 = 2^{N-1}+1$ queries.

We will now show that there is a quantum circuit for which only one query suffices! Consider the generalization of the previous circuit:



We analyse the efect of the circuit step by step.

The initial state is $|0\rangle \otimes ... \otimes |0\rangle \otimes |1\rangle$.

The Hadamard gates transform the input as

$$\begin{aligned} & H\otimes ... \otimes H|0\rangle \otimes ... \otimes |0\rangle \otimes |1\rangle \\ & \quad = H|0\rangle \otimes ...H|0\rangle \otimes H|1\rangle \\ & \quad = \frac{1}{2^{N/2}}\big(|0\rangle + |1\rangle\big) \otimes ... \otimes \big(|0\rangle + |1\rangle\big) \otimes \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \\ & \quad = \frac{1}{2^{N/2}} \sum_{b_1...b_N \in \{0,1\}^N} |b_1...b_N\rangle \otimes \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle). \end{aligned}$$

We query the U_f box with this input. The output is:

$$\begin{aligned} & \frac{1}{2^{N/2}} \sum_{b_1...b_N \in \{0,1\}^N} \left(\frac{1}{\sqrt{2}}|b_1...b_N, f(b_1...b_N)\rangle - \frac{1}{\sqrt{2}}|b_1...b_N, 1+f(b_1...b_N)\rangle \right) \\ & \quad = \frac{1}{2^{N/2}} \sum_{b_1...b_N} \frac{1}{\sqrt{2}}(-1)^{f(b_1...b_N)} \left(|b_1...b_N, 0\rangle - \frac{1}{\sqrt{2}}(-1)^{f(b_1...b_N)}|b_1...b_N, 1\rangle \right) \\ & \quad = \frac{1}{2^{N/2}} \left(\sum_{b_1...b_N} (-1)^{f(b_1...b_N)}|b_1...b_N\rangle \right) \otimes \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \end{aligned}$$

We apply again the Hadamard gates $H \otimes \cdots \otimes H$ on the first N Qbits and leave the last Qbit intact. This yields:

$$\frac{1}{2^{N/2}} \sum_{b_1...b_N} (-1)^{f(b_1...b_N)}(H|b_1\rangle \otimes ... \otimes H|b_N\rangle) \otimes \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$$

Since

$$H|b_i\rangle = \frac{1}{\sqrt{2}}(|0\rangle + (-1)^{b_i}|1\rangle)$$

we obtain

$$H|b_1\rangle \otimes ... \otimes H|b_N\rangle = \frac{1}{2^{N/2}} \sum_{a_1...a_N} (-1)^{\vec{a}\cdot\vec{b}} |a_1...a_N\rangle$$

where

$$\vec{a} \cdot \vec{b} = \sum_{i=1}^{N} a_i b_i$$

The net result for the output of the circuit is therefore:

$$\frac{1}{2^N}\left(\left\{ \sum_{b_1...b_N} (-1)^{f(b_1...b_N)} (-1)^{\vec{a}\cdot\vec{b}} \right\} \sum_{a_1...a_N} |a_1...a_N\rangle \right) \otimes \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$$

Now we measure the first N Qbits: the probability of an outcome $(a_1...a_N) = (0...0)$ is

$$\text{Prob}(0...0) = \frac{1}{2^{2N}} | \sum_{b_1...b_N} (-1)^{f(b_1...b_N)} |^2$$

We conclude that

- If f happens to be constant we find $\text{Prob}(0...0) = 1$ so the measurement will surely yield $(0...0)$.
- If f happens to be balanced we find $\text{Prob}(0...0) = 0$ so the measurement will surely yield some $a_1...a_N \neq (0...0)$.

Summarizing, we see that with only one query of the quantum circuit we can learn if f is constant or balanced! We have explored all possible classical questions simultaneously, a spectacular illustration of quantum parallelism.

11 Hidden Group, Period Finding and Factoring

In this chapter we begin with a variation of the Deutsch-Josza problem, that was originally proposed by Simon. As we will see the quantum Fourier transform will appear naturally as an attempt to generalize Simon's algorithm to the search for the (hidden) period of a function. The factorization of an integer can be reduced to the search of the period of an arithmetic functio (the modular exponential). Shor's factorization algorithm exploits this fact and is an instance of the period-finding quantum algorithm.

11.1 Simon's problem and hidden groups

Let us first give the statement of the simplest version originally considered by D. Simon. Let $f : \{0,1\}^n \to X$ where X is a finite set of values. We suppose that we are assured that f satisfies "Simon's promise": $f(\underline{x}) = f(\underline{y})$ if and only if $\underline{x} = \underline{y}$ or $\underline{x} = \underline{y} \oplus \underline{a}$ for some fixed $\underline{a} \in \{0,1\}^n = \mathbb{F}_2^n$. Simon's problem is: given a black box that computes f find $\underline{a}$ in a minimal number of queries of f.

It can be shown that if the black box is classical then we need at least $O(2^{cn})$ queries to find $\underline{a}$ with finite probability $O(1)$. We will show that if the black box is quantum, there is an algorithm such that $O\big(\text{poly}(n)\big)$ queries suffices to find $\underline{a}$ with probability equal to $1 - O(2^{-cn})$.

We note that $\{\underline{0}, \underline{a}\}$ forms a subgroup of $(\mathbb{F}_2^n, \oplus)$ (in fact this is a one-dimensional vector sub-space). This motivates the more general Hidden Subgroup Problem. Let G be a finite group (abelian) and H a subgroup. Let $\rho : G \to X$ (X a finite set) with satisfies Simon's promise: ρ is constant on the cosets G/H (equivalence classes). The problem is to find a set of generators for H.

The following figure (?) illustrates a function that is constant over the cosets G/H. The group G can be partioned in equivalence classes (cosets) with equal cardinalities.

If the elements of G (abelian) admit a binary representation of size $O(n); n = \log |G|$ then there exist a quantum algorithm solving the problem with probability $1 - O(2^{-cn})$ with $O(n)$ queries. For G non-abelian the problem is still open.

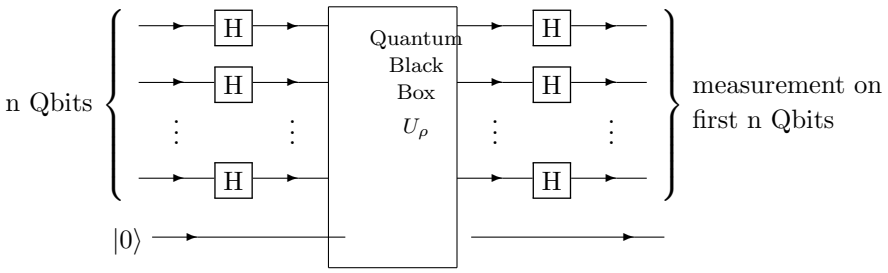
Here we will limit ourselves to $G = (\mathbb{F}_2^n, \oplus)$ and $H \subset G$ is a subgroup of $\mathbb{F}_2^n$. In fact H is also a vector sub-space since $\mathbb{F}_2^n$ is a vector space. That is, H has dimension $dimH$ and we look for a set of basis vectors. Equivalently we

could look for a set of basis vectors of $H^\perp$ (the orthogonal complement of H: $\mathbb{F}_2^n = H \oplus H^\perp$). Indeed once a basis of $H^\perp$ is known it is possible to find a basis for H in $O\big(\text{poly}(n)\big)$ steps by methods of linear algebra (e.g Gauss-Jordan).

For later use we note that the number of vectors in $\mathbb{F}_2^n$ is 2^n; the number of vectors in H is $2^{\dim H}$; the number of vectors in $H^\perp$ is $2^{\dim H^\perp}$. The cosets $\mathbb{F}_2^n/H$ all have one representative that we denote by $\underline{t}$. Given a coset with representative $\underline{t}$, all vectors in the coset are $\{\underline{t} + \underline{h} \mid \underline{h} \in H\}$. So the number of vectors in each coset is $2^{\dim H}$ and the number of cosets is $\frac{2^n}{2^{\dim H}} = 2^{\dim H^\perp}$.

11.2 Simon's algorithm

The quantum circuit for Simon's algorithm is



Let us perform all operations step by step.

Initialisation: the initial state is $|0\rangle \otimes ... \otimes |0\rangle \otimes |0\rangle$. Any quantum algorithm may be initialized in this way at the expense of adding unitary gates effecting the desired preparation. Here this preparation is performed by the next step.

Hadamard transformations: these prepare transform the initial state into a coherent superposition of all possible "classical states",

$$\begin{aligned}(H|0\rangle \otimes ... \otimes H|0\rangle) \otimes |0\rangle &= \frac{1}{2^{n/2}}(|0\rangle + |1\rangle) \otimes ... \otimes (|0\rangle + |1\rangle) \otimes |0\rangle \\ &= \frac{1}{2^{n/2}} \sum_{\underline{b}} |\underline{b}\rangle \otimes |0\rangle \quad ; \quad \underline{b} = (b_1 ... b_n) \quad ; \quad b_i \in \{0,1\} \\ &= \frac{1}{2^{n/2}} \sum_{\underline{t} \in \mathbb{F}_2^n/H} \; \sum_{\underline{h} \in H} |\underline{t} + \underline{h}\rangle \otimes |0\rangle\end{aligned}$$

Quantum Black Box: on the computational basis states the black box acts as $U_\rho |\underline{b}\rangle \otimes |0\rangle = |\underline{b}\rangle \otimes |\rho(\underline{b})\rangle$. As usual this can be checked to be unitary. Using that ρ is constant on cosets $\mathbb{F}_2^n/H$, we find that the output of U_ρ on the state prepared by the Hadamard transforms is

$$\frac{1}{2^{n/2}} \sum_{\underline{t} \in \mathbb{F}_2^n/H} \; \sum_{\underline{h} \in H} |\underline{t} + \underline{h}\rangle \otimes |\rho(\underline{t})\rangle$$

Final Hadamard transforms: applying $\underbrace{H \otimes H ... \otimes H}\otimes \mathbf{1}$ we get the output quantum superposition:

$$\frac{1}{2^{n/2}} \sum_{\underline{t}\in \mathbb{F}_2^n/H} \quad \sum_{\underline{h}\in H} (H|t_1 + h_1\rangle \otimes ... \otimes H|t_n + h_n\rangle) \otimes |\rho(\underline{t})\rangle$$

Using

$$H|x_i\rangle = \frac{1}{\sqrt{2}} \sum_{y_i=0,1} (-1)^{x_i y_i} |y_i\rangle$$

we obtain

$$\frac{1}{2^{n/2}} \sum_{\underline{t}\in \mathbb{F}_2^n/H} \quad \sum_{\underline{h}\in H} \frac{1}{2^{n/2}} \sum_{\underline{y}} (-1)^{(\underline{t}+\underline{h})\cdot \underline{y}} |y\rangle \otimes |\rho(\underline{t})\rangle$$
$$= \frac{1}{2^n} \sum_{\underline{t}\in \mathbb{F}_2^n/H} \quad \sum_{\underline{y}} \Big(\sum_{\underline{h}\in H} (-1)^{\underline{h}\cdot \underline{y}} \Big) (-1)^{\underline{t}\cdot \underline{y}} |y\rangle \otimes |\rho(\underline{t})\rangle$$

Since H is a group:

$$\sum_{\underline{h}\in H} (-1)^{\underline{h}\cdot \underline{y}} = \begin{cases} |H| = 2^{\dim H} & \text{if } \underline{y} \in H^{\perp} (\text{i.e } \underline{h}\cdot \underline{y} = 0) \\ 0 \quad \text{otherwise} \end{cases}$$

Note that to get the "otherwise" we use the group property:

$$\sum_{\underline{h}\in H} (-1)^{\underline{h}\cdot \underline{y}} = \sum_{\underline{h}\in H} (-1)^{(\underline{h}+\underline{g})\cdot \underline{y}} = (-1)^{\underline{g}\cdot \underline{y}} \sum_{\underline{h}\in H} (-1)^{\underline{h}\cdot \underline{y}}$$

for some fixed $\underline{g} \in H$. If $\underline{y} \notin H^{\perp} \quad (-1)^{\underline{g}\cdot \underline{y}} \neq 0$ so that $\sum_{\underline{h}\in H} (-1)^{\underline{h}\cdot \underline{y}} = 0$.

Summarizing, the output of the quantum circuit is:

$$|\phi\rangle = \frac{2^{\dim H}}{2^n} \cdot \sum_{\underline{t}\in \mathbb{F}_2^n/H} \quad \sum_{\underline{y}\in H^{\perp}} (-1)^{\underline{t}\cdot \underline{y}} |y\rangle \otimes |\rho(t)\rangle$$

This is the state just before the measurement. The quantum black box has been queried *only once* to prepare this state.

Measurement of first n Qbits: we measure in the computational basis $\{|0\rangle, |1\rangle\}$ so that the outcome is $\underline{y} \in H^{\perp}$. From the measurement postulate

$$\text{Prob}(y) = \langle \phi | (\underbrace{|\underline{y}\rangle\langle \underline{y}| \otimes \mathbf{1}}) |\phi\rangle$$
$$= \frac{2^{2\dim H}}{2^{2m}} \sum_{\underline{t},\underline{t}'} (-1)^{\underline{t}\cdot \underline{y}} \, (-1)^{\underline{t}'\cdot \underline{y}} \, \langle \rho(\underline{t})| \, \rho(\underline{t}')\rangle$$

Now since ρ takes different values on different cosets we have

$$\langle \rho(\underline{t})| \, \rho(\underline{t}')\rangle = \delta_{\underline{t},\underline{t}'}$$

so that the probability of outcome y becomes:

$$\begin{aligned}\text{Prob}(y) &= \frac{2^{2\dim H}}{2^{2n}} \cdot \#(\text{of cosets}) \\ &= \frac{2^{\dim H}}{2^{2n}} \cdot 2^n \\ &= \frac{1}{2^{\dim H^\perp}} = \frac{1}{|H^\perp|}\end{aligned}$$

So we get a state $|\underline{y}\rangle$ for some $y \in H^\perp$ with uniform probability $\frac{1}{2^{\dim H^\perp}}$. In other words, when we query the quantum black box we obtain surely a vector in $H^\perp$. This is a random vector in $H^\perp$. So if we query $O(n)$ times the quantum black box we get a set of random vectors $\underline{y}_1, \ldots \underline{y}_{O(n)}$ surely in $H^\perp$.

It remains to estimate the probability that $\underline{y}_1, \ldots \underline{y}_{O(n)}$ form a linearly independent set.

Lemma 11.2.1 Let $\underline{y}_1, \ldots \underline{y}_k$ randomly chosen vectors from $H^\perp$ with a uniform distribution. Then for $k \leq \dim H^\perp$,

$$\text{Prob}(\underline{y}_1 \ldots \underline{y}_k \text{linearly independent}) \geq \frac{1}{4}$$

Proof Choose $\underline{y}_1$. We have

$$\text{Prob}(\underline{y}_1 \neq 0) = \frac{2^{\dim H^\perp} - 1}{2^{\dim H^\perp}}$$

Suppose that $(\underline{y}_1 \ldots \underline{y}_{i-1})$ have been chosen linearly independent. Choose $\underline{y}_i$.

$$\text{Prob}(\underline{y}_i \text{linearly independent of} (\underline{y}_1 \ldots \underline{y}_{i-1})) = \frac{2^{\dim H^\perp} - 2^{i-1}}{2^{\dim H^\perp}}$$

Indeed the # (of vector lin dep in $\underline{y}_1 \ldots \underline{y}_{i-1}$ subspace)$= 2^{i-1}$. So $\underline{y}_i$ must be in the complement. Thus we get

$$\begin{aligned}\text{Prob}(y_1 \ldots y_k \text{ lin indep}) &= \prod_{i=0}^{k-1} \frac{2^{\dim H^\perp - 2^i}}{2^{\dim H^\perp}} \\ &= \prod_{i=0}^{k-1} (1 - 2^{i - \dim H^\perp})\end{aligned}$$

For $k \leq \dim H^\perp$ we can show that this product is $\geq \frac{1}{4}$. This is conveniently done by writing the product as the exponential of a logarithm and looking at the first order Taylor expansion of the logarithm. □

We conclude that if we run the algorithm M times we assured that $\sim \frac{M}{4}$ times we will have an independent set of vectors spanning $H^\perp$. Each time we make a check to see if we are successful (If we are we stop; if we are not we continue). The total number of queries is $O(M)$. If $M \sim n$ we can show that we will be successful with probability exponentially close to 1.

11.3 Period Finding and Quantum Fourier Transform

Consider the following variation of the Hidden Subgroup problem. Let $G = (\mathbb{Z}, +)$ and $H = \frac{\mathbb{Z}}{r\mathbb{Z}}$, where r some fixed unknown integer. We have a function $f : \mathbb{Z} \to \mathbb{Z}$ which is r-periodic:

$$f(x) = f(x + r), x \in \mathbb{Z}$$

and we want to find the period. Here the group G is infinite so that Simon's algorithm has to be adapted. The first idea is to assume that we know $r < M$ for some very large integer M and then restrict $G = \mathbb{Z}$ to the set $\{0, 1, ..., N-1\}$ for some $N >> M$ (later on we will choose $N \sim M^2$). The problem now is that $\{0, 1, ..., N-1\}$ does not have the group structure so Simon's algorithm does not quite work. We will see that a very similar algorithm works which is based on the Quantum Fourier Transform. The net result will be that we can find the period r with exponentially high probability $1 - O(2^{-cn})$ in poly(n) time where $n = \log_2 N$.

We use the following rotation. Integers in $\{0, 1, ..., N-1\}$ are denoted by x, y. Corresponding quantum states are $|x\rangle$, $|y\rangle$. If $N = 2^n$ these states are in the space $\mathbb{C}^2 \otimes ... \otimes \mathbb{C}^2$. More precisely if we use the binary expansion

$$x = 2^{n-1}x_{n-1} + ... + 2^2x_2 + 2x_1 + 2^0x_0, \quad x_i \in \{0, 1\}$$

we define the computational basis states

$$\begin{aligned} |x\rangle &= |x_{n-1}...\, x_2\ x_1\ x_0\rangle \\ &= |x_{n-1} \otimes ... \otimes |x_2\rangle \otimes |x_1\rangle \otimes |x_0\rangle \end{aligned}$$

As usual we can form superpositions of such states, that have no classical analog.

The quantum Fourier transform is a unitary operator defined on basis vectors as

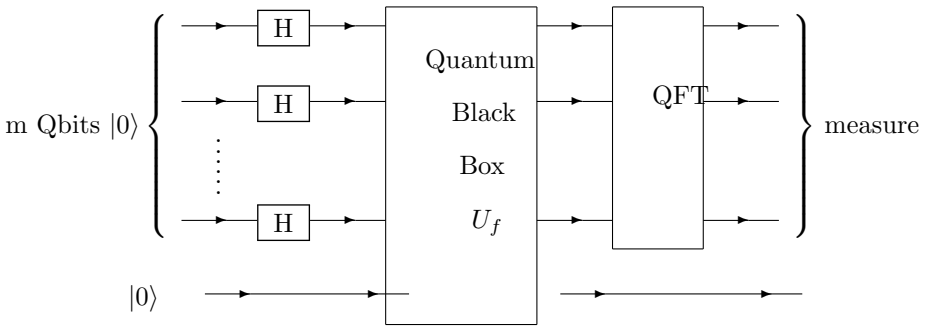
$$\text{QFT } |x\rangle = \frac{1}{\sqrt{N}} \sum_{y=0}^{N-1} e^{2\pi i \frac{xy}{N}} |y\rangle, \quad N = 2^n$$

Its action on a general state is obtained by linearity:

$$\text{QFT}\Big(\sum_{x=0}^{N-1} C_x |x\rangle \Big) = \sum_{x=0}^{N-1} C_x \text{ QFT } |x\rangle$$

Thanks to the general results of lecture 10 we know that it can be implemented using only one and two bit elementary gates. In fact as we will show later the size of the circuit for QFT can be chosen $O(n^2) = O((\log_2 N)^2)$.

For the moment we just suppose that we have such a circuit at our disposal and look at the algorithm for period inding. The quantum circuit for Period Finding is:



With respect to Simon's algorithm we see that the first series of Hadamard gates is the same. It serves to transform the initial state into a coherent superposition of all possible classical inputs. The second column of Hadamard gates has been traded for a QFT.

Let us now look at the operations performed by this circuit at each stage.

The input is $\underbrace{|0\rangle \otimes ... \otimes |0\rangle}\ \otimes |0\rangle$.

After the Hadamard transforms we get the state

$$\frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle \otimes |0\rangle$$

One query with quantum black box (unitary) yields

$$\frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} U_f(|x\rangle \otimes |0\rangle) = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle \otimes |f(x)\rangle$$

Since f is r-periodic we can rewrite this state as

$$\frac{1}{\sqrt{N}} \sum_{j=0}^{r-1} \sum_{j=0}^{A_{x_0}-1} |x_0 + jr\rangle |f(x_0\rangle$$

where A_{x_0} is an integer such that

$$N - r \leq x_0 + (A-1)r < N.$$

In the sequel we write A instead of A_{x_0} for simplicity of notation. But it is important to remenber that A depends on x_0.

Now we apply the QFT. We find the final state (just before the measurement)

$$\begin{aligned} &\frac{1}{N} \sum_{x_0=0}^{r-1} \sum_{j=0}^{A-1} \sum_{y=0}^{N-1} e^{2\pi i \frac{(x_0+jr)y}{N}} |y\rangle |f(x_0)\rangle \\ &= \frac{1}{N} \sum_{x_0=0}^{r-1} \sum_{y_0=0}^{N-1} e^{2\pi i \frac{x_0 y}{N}} \sum_{j=0}^{A-1} e^{2\pi i \frac{jy}{N/r}} |y\rangle |f(x_0)\rangle \end{aligned}$$

The last step is the measurement of the first register. According to the measurement postulate when we measure the first n Qbits in the computational basis $\{|y\rangle\}$ we get an outcome $|y\rangle$ with probability:

$$\text{Prob}(y) = \frac{1}{N^2} \sum_{x_0=0}^{r-1} \left| e^{2\pi i \frac{x_0 y}{N}} \sum_{j=0}^{A-1} e^{2\pi i \frac{jy}{N/r}} \right|^2$$

In other words the outcome is $y \in \{0, \ldots, N-1\}$ with

$$\text{Prob}(y) = \frac{1}{N^2} \sum_{x_0=0}^{r-1} \left| \sum_{j=0}^{A-1} e^{2\pi i \frac{jy}{N/r}} \right|^2$$

where A is an integer such that $N - r \leq x_0 + (A-1)r < N$.

The following remark is useful for the ensuing analysis. We have $\frac{N}{r} \leq \frac{x_0}{r} + A < \frac{N}{r} + 1$. From $x_0/r < 1$ we deduce $A < \frac{N}{r} + 1$ and $\frac{N}{r} < A + 1$. So A is in general the unique integer such that $A - 1 < \frac{N}{r} < A + 1$.

We now show that from the outcome y we can find the period r with finite probability. For this we distinguish two cases: $\frac{N}{r}$ is integer (easy but unrealistic case) and $\frac{N}{r}$ is not integer (harder realistic case).

Easy unrealistic case: suppose $\frac{N}{r}$ is a integer. Then

$$e^{2\pi i \frac{jy}{N/r}} = 1 \text{ if } y = k\frac{N}{r} \quad \text{with} k \in \{0, 1, ..., r-1\}$$

Since $A = N/r$ we have

$$\text{Prob}(y) = \begin{cases} \frac{1}{N^2} \cdot r \cdot \left(\frac{N}{r}\right)^2 = \frac{1}{r} & \text{for } y \in \{0, \frac{N}{r}, \frac{2N}{r}, ..., (r-1)\frac{N}{r}\} \\ 0 \quad \text{otherwise} \end{cases}$$

This means that with probability equal to one we have

$$\frac{y}{N} = \frac{k}{r}, \qquad k \in \{0, 1, 2, ..., r-1\}$$

If $\text{PGCD}(k, r) = 1$ we get k and r in a univocal manner by simplifying the fraction $\frac{y}{N}$. In practice we do not known if the PGCD equals one. But it is enough to simplify the fraction and check if the denominator which is obtained is indeed a period of the function.

The probability of success is the probabilisty that $\text{PGCD}(k, r) = 1$. This probability is equal to $\frac{\varphi(r)}{r}$ where $\varphi(r)$ is Euler's fonction ($\sharp$ of coprimes to r that are $< r$). It is known that

$$\varphi(r) \geq \frac{r}{4\ln(\ln r)}$$

Thus we have

$$\text{Prob } (PGCD(k, r) = 1) \geq \frac{1}{4 \ln \ln r} \geq \frac{1}{4 \ln \ln n}$$

and we can conclude that with one query of the circuit we succeed with probability at least $\frac{1}{4 \ln \ln n}$. By repeating the query of the quantum circuit $O(\ln \ln n)$ we

can amplify this probability to $O(1)$. Thus we have a polynomial time algorithm which finds the period with a probability close to 1.

Realistic case: $\frac{N}{r}$ is not an integer. The formula for the probability favors values of $y \in \{0, ..., N-1\}$ that are close to $k\frac{N}{r}$. We can prove the following:

Lemma 11.3.1

$$\text{Prob}(-\frac{1}{2} \leq y - \frac{kN}{r} \leq \frac{1}{2} \text{ for some } k) \geq \frac{2}{5}$$

The factor $2/5$ can be made as close as we want to $4/\pi^2$.

Proof By summing the complex exponentials with the help of the formula for a geometric series we get:

$$\begin{aligned}\text{Prob}(y) &= \frac{1}{N^2} \sum_{x_0=0}^{r-1} \frac{\sin^2(\frac{\pi y r A}{N})}{\sin^2(\frac{\pi y r}{N})} \\ &= \frac{1}{N^2} \sum_{x_0=0}^{r-1} \frac{\sin^2 \frac{\pi A}{N}(yr - kN)}{\sin^2 \frac{\pi}{N}(yr - kN)} \qquad (\forall\ k \text{ by periodicity})\end{aligned}$$

Now fix $k \in \{0, ..., r-1\}$ such that $-\frac{r}{2} \leq yr - kN \leq \frac{r}{2}$. Each ratio in the last sum takes its minimum for $yr - kN = \frac{r}{2}$. So:

$$\text{Prob}(|y - \frac{kN}{r}| \leq \frac{1}{2} \text{ for some } k) \geq \frac{r}{N^2} \sum_{x_0=0}^{r-1} \frac{\sin^2(\frac{\pi A r}{2N})}{\sin^2(\frac{\pi r}{2N})}$$

Since as remarked earlier $\frac{N}{r} - 1 < A < \frac{N}{r} + 1$ we have

$$\frac{\pi}{2}(1 - \frac{r}{n}) \leq \frac{\pi A r}{2N} \leq \frac{\pi}{2}(1 + \frac{r}{N})$$

which implies

$$\text{Prob}(|y - \frac{kN}{r}| \leq \frac{1}{2}) \ \geq\ \frac{r^2}{N^2} \cdot \frac{\sin^2 \frac{\pi}{2}(1 - \frac{r}{N})}{\sin^2 \frac{\pi}{2} \frac{r}{N}} \ \geq\ \frac{1}{N^2} \frac{r^2}{\frac{\pi^2}{4} \frac{r^2}{N^2}} = \frac{4}{\pi^2}$$

since $\frac{r}{N} < \frac{M}{M^2} = \frac{1}{M}$ small. □

We can now conclude the analysis. With one query of the quantum circuit we observe an integer y s.t for some $k \in \{0, ..., r-1\}$ we have $|y - k\frac{N}{r}| \leq \frac{1}{2}$ with probability at least $\frac{2}{5}$. So with this probability:

$$|\frac{y}{N} - \frac{k}{r}| \leq \frac{1}{2N} \leq \frac{1}{2M^2} < \frac{1}{2r^2} \qquad (\text{Remember } \ r << M\ ;\ N = M^2)$$

A standard result of number theory following from Euclid's algorithm states that:

Lemma 11.3.2 Let $\frac{y}{N}$ and $\frac{k}{r}$ s.t $|\frac{y}{N} - \frac{k}{r}| \leq \frac{1}{2r^2}$. If $\text{PGCD}(k, r) = 1$ then $\frac{k}{r}$ belongs to the set of convergents[1] of truncated of the fraction $\frac{y}{N}$. All convergents

[1] The set of convergents of a fraction is given by all truncations of the continuous fraction expansion.

can be found by Euclid's algorithm from the continuous fraction expansion of $\frac{y}{N}$ in $O((\log_2 N)^3)$ steps (i.e $O(n^3)$.

Note that in practice we do not know if the PGCD is equal to 1, so at each step of the continuous fraction expansion we check if the denominator is a period of the function.

What is the probablity of success? Combining this Lemma with the previous results we deduce that with one query yielding y we can successfully compute r with a

$$\text{Prob} \geq \frac{2}{5} \cdot \frac{1}{4(\ln \ln r)} \qquad \text{in } O(n^3) \text{ steps.}$$

So by repeating the procedure $O(\ln \ln n)$ times we can amplify the probability and get r with a probability close to 1. The total time is $O(n^3 \ln \ln n)$.

11.4 Quantum circuit for the quantum Fourier transform

Recall that the QFT is defined by its action on basis vectors $\{|0\rangle, |1\rangle, \ldots, |N-1\rangle\}$ of an N-dimensional Hilbert space $\mathcal{H} = \text{span}\{|0\rangle, \ldots, |N-1\rangle\}$:

$$\text{QFT}|x\rangle = \frac{1}{\sqrt{N}} \sum_{y=0}^{N-1} \exp(2\pi i \frac{xy}{N})|y\rangle$$

We first look at special cases to unravel the structure of this transformation.

Case $N = 2$**.** The QFT reduces to the usual Hadamard gate:

$$(\text{QFT})_{N=2}|x\rangle = \frac{1}{\sqrt{2}} \sum_{y=0}^{1} \exp(\pi i x y)|y\rangle = \frac{1}{\sqrt{2}} \sum_{y=0}^{1} (-1)^{xy}|y\rangle$$

$$= \frac{1}{\sqrt{2}}(|0\rangle + (-1)^x|1\rangle)$$

Of course the circuit in this case is simply

$$|x\rangle \longrightarrow \boxed{\text{H}} \longrightarrow H|x\rangle$$

Case $N = 4$**.** We have 4 states in the Hilbert space: $\mathcal{H} = \text{span}\{|0\rangle, |1\rangle, |2\rangle, |3\rangle\}$.

$$\text{QFT}|x\rangle = \frac{1}{\sqrt{4}}\Big\{e^{2\pi i \frac{x\cdot 0}{4}}|0\rangle + e^{2\pi i \frac{x\cdot 1}{4}}|1\rangle + e^{2\pi i \frac{x\cdot 2}{4}}|2\rangle + e^{2\pi i \frac{x\cdot 3}{4}}|3\rangle\Big\}$$

We can represent $y = 0, 1, 2, 3$ in binary rotation:

$$|0\rangle = |0\,0\rangle \;;\; |1\rangle = |0\,1\rangle \;;\; |2\rangle = |1\,0\rangle \;;\; |3\rangle = |1\,1\rangle$$

Then

$$\text{QFT}|x\rangle = \frac{1}{\sqrt{4}}(|0\,0\rangle + e^{i\frac{\pi}{2}x}|0\,1\rangle) + e^{i\pi x}|1\,0\rangle + e^{3i\frac{\pi}{2}x}|1\,1\rangle)$$

$$= \frac{1}{\sqrt{2}}(|0\rangle + e^{i\pi x}|1\rangle) \otimes (|0\rangle + e^{i\frac{\pi}{2}x}|1\rangle)$$

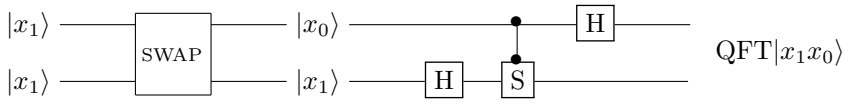
Now we use the binary expansion of x, i.e $x \in \{0;1;2;3\} = 2 \cdot x_1 + x_0$ with $x_0, x_1 \in \{0,1\}$, to remark that

$$e^{i\pi x} = e^{2\pi i x_1} e^{i\pi x_0} = (-1)^{x_0}, \qquad e^{i\frac{\pi}{2}x} = e^{i\pi x_1} e^{i\frac{\pi}{2}x_0} = (-1)^{x_1} e^{i\frac{\pi}{2}x_0}$$

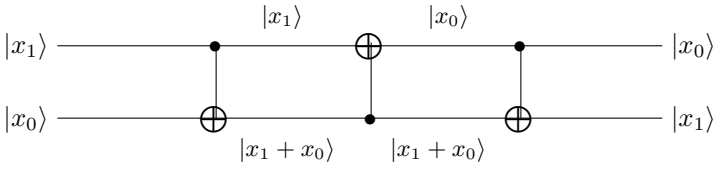
Thus we have

$$QFT|x\rangle QFT|x_0x_1\rangle = \frac{1}{\sqrt{2}}(|0\rangle + (-1)^{x_0}|1\rangle) \otimes (|0\rangle + (-1)^{x_1} e^{i\frac{\pi}{2}x_0}|1\rangle)$$

A circuit realizing this operation is:



where the SWAP operation is realized as follows:



Once the SWAP operation is performed on $|x_1x_0\rangle$ we act with H on the second Qbit:

$$H \text{ SWAP } |x_1x_0\rangle = (I \otimes H)|x_0x_1\rangle = |x_0\rangle \cdot \frac{1}{\sqrt{2}}(|0\rangle + (-1)^{x_1}|1\rangle)$$

Then we act with a controlled $S = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\frac{\pi}{2}} \end{pmatrix}$ gate:

$$(CS)(I \otimes H) \text{ SWAP } |x_1x_0\rangle = \text{CS } |x_0\rangle \otimes \frac{1}{\sqrt{2}}(|0\rangle + (-1)^{x_1}|1\rangle) \tag{11.1}$$

$$= |x_0\rangle \otimes \frac{1}{\sqrt{2}}(|0\rangle + (-1)^{x_1} e^{i\frac{\pi}{2}x_0}|1\rangle) \tag{11.2}$$

In fact the matrix for CS (a two bit gate) is

$$CS = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\frac{\pi}{2}} \end{pmatrix}$$

so that $e^{i\frac{\pi}{2}}$ acts only on $|1\,1\rangle = \begin{pmatrix}0\\0\\0\\1\end{pmatrix}$. Another way to express CS is

$$CS = |0\rangle\langle 0| \otimes \mathbf{1} + |1\rangle\langle 1| \otimes S = \begin{pmatrix}1 & 0\\0 & 0\end{pmatrix} \otimes \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix} + \begin{pmatrix}0 & 0\\0 & 1\end{pmatrix} \otimes \begin{pmatrix}1 & 0\\0 & e^{i\frac{\pi}{2}}\end{pmatrix}$$

The last Hadamard gate acts on the first bit and yields:

$$\begin{aligned}(H \otimes I)(CS)(I \otimes H)\text{SWAP}\ |x_1x_0\rangle &= H|x_0\rangle \otimes \frac{1}{\sqrt{2}}(|0\rangle + (-1)^{x_1}e^{i\frac{\pi}{2}x_0}|1\rangle)\\ &= \frac{1}{\sqrt{2}}(|0\rangle + (-1)^{x_0}|1\rangle) \otimes \frac{1}{\sqrt{2}}(|0\rangle + (-1)^{x_1}e^{i\frac{\pi}{2}x_0}|1\rangle)\end{aligned}$$

So we have the decomposition (for $N = 4$)

$$\text{QFT}\ = (H \otimes I)(CS)(I \otimes H)(\text{SWAP})$$

Generalisation to any $N = 2^n$. We begin with a lemma that will allow us to find the general circuit.

Lemma 11.4.1 Let $x \in \{0, 1, \ldots, N-1\}$ with $N = 2^n$.

$$\text{QFT}|x\rangle = \prod_{l=1}^{n}(|0\rangle + e^{\frac{i\pi x}{2^{l-1}}}|1\rangle)$$

Proof Use the binary representation for $|y\rangle = |y_{n-1}\ldots y_0\rangle$ where

$$y = 2^{n-1}y_{n-1} + 2^{n-2}y_{n-2} + \ldots + 2^0 \cdot y_0$$

with $y_i \in \{0, 1\}$. Take the definition of $\text{QFT}|x\rangle$ and split the sum over $y \in \{0, \ldots, N-1\}$ into a sum over even terms ans a sum over odd terms:

$$\begin{aligned}\text{QFT}|x\rangle &= \frac{1}{2^{\frac{n}{2}}}\sum_{y\ \text{even}} e^{2\pi i\frac{xy}{2^n}}|y\rangle + \sum_{y\ \text{odd}} e^{2\pi i\frac{xy}{2^n}}|y\rangle\\ &= \sum_{y'=0}^{2^{n-1}-1} e^{2\pi i\frac{x2y'}{2^n}}|y_{n-1}, \ldots, y_1, 0\rangle + \sum_{y'=0}^{2^{n-1}-1} e^{2\pi ix\frac{2y'+1}{2^n}}|y_{n-1}, \ldots, y_1, 1\rangle\end{aligned}$$

where we used the facts that:

- If $y = 2y'$ and $y = 2^{n-1}y_{n-1} + \ldots + 2^1y_1 + 2^0y_0$, then $y' = 2^{n-2}y_{n-1} + \ldots + 2^0y_1$ and $y_0 = 0$.
- If $y = 2y'+1$ and $y = 2^{n-1}y_{n-1}+\ldots+2^1y_1+2^0y_0$, then $y' = 2^{n-2}y_{n-1}+\ldots+2^0y_2$ and $y_0 = 1$.

With this decomposition we conclude that

$$\text{QFT}|x\rangle = \Big(\frac{1}{2^{\frac{(n-1)}{2}}}\sum_{y=0}^{2^{n-1}-1} e^{\frac{2\pi ixy}{2^{n-1}}}|y\rangle\Big) \otimes \frac{1}{\sqrt{2}}(|0\rangle + e^{\frac{\pi ix}{2^{n-1}}}|1\rangle)$$

By repeating the same decomposition again and again on the first parenthesis we obtain the result of the Lemma. □

The l-th term in the tensor product of the Lemma is:

$$|0\rangle + e^{i\frac{\pi}{2^{l-1}}x}|1\rangle$$

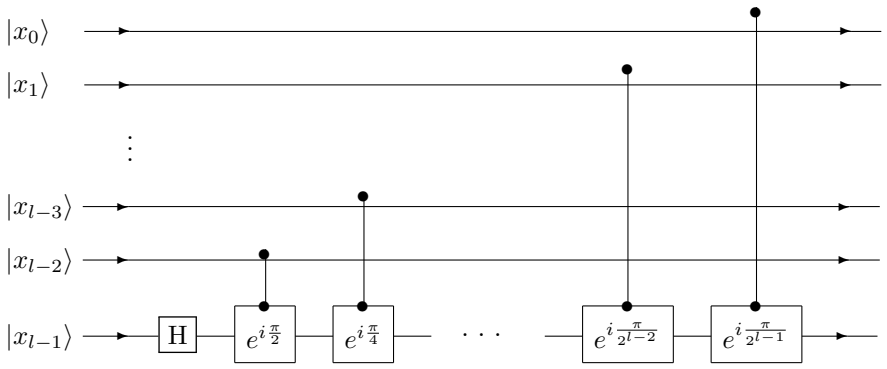
Let us look at the phase factor more closely. The binary expansion of x is:

$$x = 2^{n-1} \cdot x_{n-1} + ... + 2^2 \cdot x_2 + 2^1 \cdot x_1 + 2^0 \cdot x_0$$

and this implies that

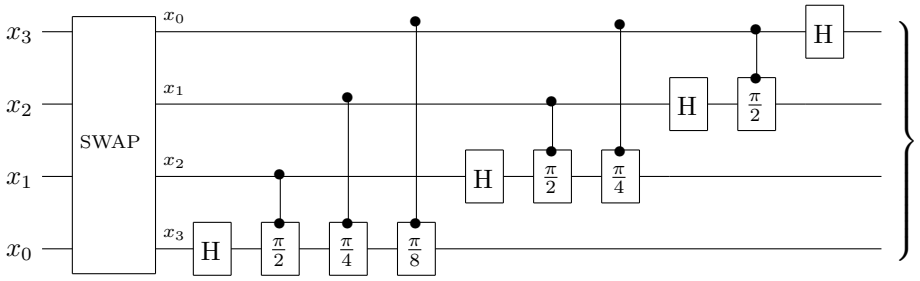
$$e^{i\frac{\pi}{2^{l-1}}x} = (-1)^{x_{l-1}} \cdot e^{i\frac{\pi}{2}x_{l-2}} \cdot e^{i\frac{\pi}{4}x_{l-3}} ... e^{i\frac{\pi}{2^{l-1}}x_0}$$

So to obtain the $l-th$ term in the product we may use the operations (Hadamard and double control phases).



The output of this circuit is $|x_0\rangle|x_1\rangle...|x_{l-2}\rangle \cdot \frac{1}{\sqrt{2}}(|0\rangle + e^{i\frac{\pi}{2^{l-1}}x}|1\rangle)$.

From these observations we deduce the full circuit for the QFT found by Shor (1994). Here we depict the example of $n = 4$ Qbits ($N = 2^4 = 16$). The circuit represents the 16×16 QFT matrix.



Let us briefly discuss the complexity of this circuit. To perform the SWAP here we can first SWAP x_0 and x_3 and then x_1 and x_2. In any case the SWAP operation requires $O(3n)$ CNOT gates. The rest of the QFT requires (H and phase gates):

$$n + (n-1) + ... 1 = \frac{n(n-1)}{2} \quad \text{gates}$$

So the overall size of the QFT circuit is $O(n^2) = O((\ln N)^2)$. We also remark that in practice one may want to acheive only a finite accurarcy: then one can neglect the phase gates $\frac{\pi}{2^K}$ for $\frac{1}{2^K} < \epsilon$, and then the number of gates becomes $O(n) = O(\ln N)$ [the coefficient will be ϵ-dependent].

11.5 Shor's factorization algorithm

In this section we review the main application of period fincding and of the QFT, namely Shor's famous algorithm for factoring an integer N. The "size" of the integer is given by $n = O(\ln N)$, i.e the length of its decimal or binary representation. Classicaly there is no know classical polynomial method (the best methods are superpolynomial but still subexponential). Note however that there is no proof that a polynomial classical method of complexity $\mathrm{O}((\ln N)^a)$ does not exist!

As we will see Shor's quantum algorithm has a total complexity of $O(n^3 \ln n)$ The strategy is as follows: first we reduce factoring to a probabilistic method for order finding in modular arithmetic. This goes back to Miller ($\sim$ 1976). Then we will recognize that order finding is a particular case of period finding for the modular exponential function. We will be able to show that the "black box" representing the modular exponential can be realized with a polynomial number of gates. Combining these results with the QFT leads immediately to Shor's algorithm.

11.5.1 Reduction of Factoring to Order Finding

Let N be an integer to be factored. We will suppose that

$$N = p_1^{e_1} p_2^{e_2} ... p_k^{e_k} \qquad \text{with} \quad p_i \neq 2 \text{ and } k \geq 2$$

Indeed powers of 2 are easily recognized and extracted (even numbers) and if $N = p^e$ there exist efficient classical methods to find p and e. We give the steps of the factoring algorithm based on order finding:

1. Choose randomly with uniform probability a $\in \{2, ..., N-1\}$ and compute

$$d = \mathrm{GCD}(a, N)$$

This greatest common divisor can be computed by Euclid's algorithm in $\sim (\ln N)^3$ steps.

2. If $d > 1$: we have a factor of N. We keep this factor and start again at [a].

3. If $d = 1$ we find the "order of a mod N" i.e we find the smallest integer r such that:

$$a^r \equiv 1 \mod \mathrm{N}$$

For this step there is no known polynomial method and that is where Shor's algorithm enters.

4. Suppose r is odd: output failure and go back to [a].

5. Suppose r is even. Then

$$a^r - 1 = (a^{\frac{r}{2}} - 1)(a^{\frac{r}{2}} + 1)$$

Since N divides $a^r - 1$ we have three possibilities:

a. N divides $a^{\frac{r}{2}} - 1$ But this is impossible since we would have $a^{\frac{r}{2}} \equiv 1 \pmod{N}$ so r would not be the (smallest) order.

b. N divides $a^{\frac{r}{2}} + 1$. Then output failure and go back to [a].

c. N shares non-trivial factors with both $(a^{\frac{r}{2}} - 1)$ and $(a^{\frac{r}{2}} + 1)$. In other words: $d_{\pm} = \mathrm{GCD}\ (a^{\frac{r}{2}} \pm 1, N)$ is non trivial and we have two factors d_+ and d_- of N. This step can be done again in $(\ln N)^3$ steps thanks to Euclid's algorithm.

6. Go back to [a].

What is the probability of success for one run? The answer is provided by the following Lemma (proven by using the Chinese Remainder Theorem). For the proof we refer to the literature [e.g Appendix in book of Chueng & Nielsen].

Lemma 11.5.1 Let $N = p_1^{e_1} ... p_k^{e_k}$; $p_i \neq 2$; $k \geq 2$. Choose a randomly uniformly in $\{2, ..., N-1\}$. Then

$$\mathrm{Prob}(\{r \text{ is even and } a^{\frac{r}{2}} \not\equiv -1 \bmod N\}) \geq \frac{1}{2}.$$

This is enough to ensure success by running the algorithm a large number of times (but finite with respect to $\log N$).

11.5.2 Quantum algorithm for order finding

Given $a \in \{2, ..., N-1\}$ we want to find the smallest integer r such that

$$a^r \equiv 1 \ (\mathrm{mod} N)$$

We recognize that r is the period of the *modular exponential*. This is the number theoretic function:

$$f_a : \frac{\mathbb{Z}}{N\mathbb{Z}} \to \frac{\mathbb{Z}}{N\mathbb{Z}} \tag{11.3}$$

$$x \longmapsto f_a(x) = a^x (\mathrm{mod}\ N) \tag{11.4}$$

Indeed $f_a(x + r) = a^{x+r} = a^x a^r = a^x \mathrm{mod}\ N$, and r is the smallest such integer, so it is the period of teh function f_a.

Suppose now we are given a black box performing U_{f_a}:

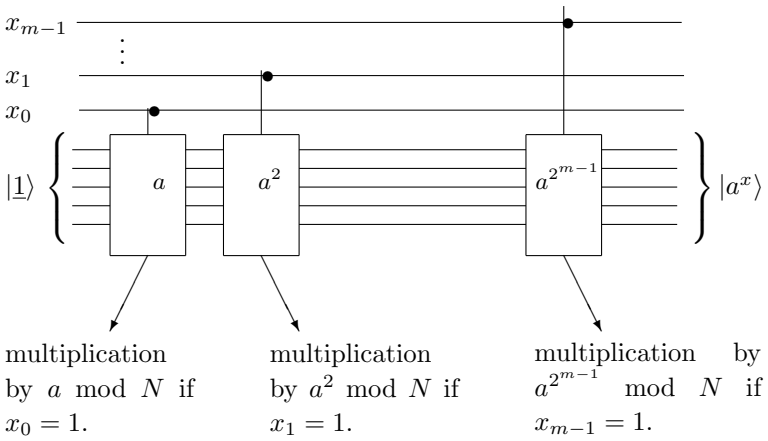
$$U_{f_a}|x,0\rangle \;=\; |x,a^x\rangle$$

Then we can simply use the algorithm (circuit) for period finding. What is the complexity of the total circuit? We know that the QFT can be realized with $O((\ln N)^2)$ gates. We will now show that U_{f_a} can be realized with $O((\ln N)^3)$ gates, so that this operation dominates the circuit complexity.

Let $x = x_{m-1}2^{m-1} + ... + 2^0 \cdot x_0$ and set $M = 2^m$. Look at $a^x \mod N$. We have

$$a^x (\text{mod } N) = (a^{2^{m-1}})^{x_{m-1}} ... (a^{2^0})^{x_0} (\text{mod } N)$$

So if we know all even powers of a, namely $a, a^2, \cdots, a^{2^{m-1}}$ We can use the circuit:



The "controlled multiplications" by a^{2^j} (if the control bit $x_j = 1$) can themselves be represented by standard classical gates. From the techniques exposed in Chapter 10 these are made *reversible* and can thus be implemented as unitary quantum gates.

This last circuit uses uses $O(m)$ controlled multiplications. Each multiplication can be done in $O(m^2)$ steps (or gates). Thus the modular exponential

$$U_{f_a}(|x\rangle \otimes |\underline{1}\rangle) = |x\rangle \otimes |a^x \text{ mod } M\rangle$$

can be realized by a circuit of size $O(m^3) = O((\ln M)^3)$. Finally, remember that in period finding with $r < N$ we choose to work with m bits where $2^m = N \sim M^2$. So the total size of this circuit is $O((\ln N)^3)$ again.

11.5.3 Shor's algorithm

We summarize the algorithm resulting from the discussion of the previous paragraphs. We input an odd number N with at least two distinct prime factors. The algorithm outputs a non trivial factor of N.

1. Choose randomly uniformly $a \in \{2, \dots, N-1\}$.
2. Compute $d = \text{GCD}$ (a, N) by Euclid's algorithm. If $d > 1$ output the factor d.
3. If $d = 1$ compute the order of a mod N (i.e $a^1 = 1 \bmod N$) by using Shor's quantum circuit. Work with m Qbits where $M = 2^m \sim N^2$.
4. Check that the output of Shor's circuit i.e the number r satisfies $a^r \equiv 1 \bmod N$. If not output "failure".
5. If $a^r \equiv 1 \bmod N$ then check of r is odd or if $a^{r/2} \equiv -1 \bmod N$. Output "failure" if this happens.
6. Otherwise (no "failure") compute GCD $(a^{r/2} \pm 1, N)$ by Euclid's Algorithm.

What is the probability of success for one experiment? The period finding works with prob $\frac{1}{\ln \ln N}$ as we saw (point 4). Moreover (point 5) does not happen with prob $> \frac{1}{2}$. So the probability of success of one run is $O(\frac{1}{\ln \ln N})$. By making $O(\ln \ln N)$ runs we can amplify this probability close to 1.

What is the total complexity? Let us first answer this question for one run. Step 2 has complexity $O((\ln N)^3)$. Indeed this the complexity of Euclid's algorithm. Step 3 has complexity $O((\ln N)^3) + O((\ln N)^2)$ which is the sum of the complexities for the modular exponential and the QFT. Steps 4 and 5 have complexities $\sim O((\ln N)^2)$. Finally step 6 which uses Euclid's algorithm has complexity $O((\ln N)^3)$.

So for one run we have $O((\ln N)^3)$ complexity in total. Since we need $O(\ln \ln N)$ runs to amplify the success probability, the total time needed to find one non trivial factor is of the order of $\sim O((\ln N)^3 \ln \ln N)$.

12 Search Problems and Grover Algorithm

12.1 Formulation of the search problem

Let us begin with the example of searching an item in a database. As a concrete example suppose that you are given a phonebook (the database) and a phone number $\mathcal{N}$ (the entry). Your problem is to find the person $\mathcal{P}$ corresponding to this entry. This is not easy because the phone numbers are not sorted out in any specified order. Of course the reverse problem where you are given a person's name (the entry) and you have to find its phone number is easy because the entries (the person's names) are sorted out in alphabetical order. For the initial problem, in the worst case, an exhaustive search would requires N queries of the database assuming there are N entries and N persons. One might think that a probabilistic search would improve the situation but this is not so. A simple probabilistic argument shows that we still need $O(N)$ queries if we want a non-vanishing probability of success.

Another example of search problems are decision problems. Take the example of 3-SAT. We have a boolean function

$$f(x_1 ... x_n) = \phi_1 \wedge \phi_2 ... \wedge \phi_M$$

which is a conjunction of M clauses. Each clause is a disjunction of 3 literals (boolean variables or bits) $\phi_i = x_{i_1} \vee \bar{x}_{i_2} \vee x_{i_3}$, $\phi_j = \bar{x}_{j_1} \vee \bar{x}_{j_2} \vee \bar{x}_{j_3}$ etc (here $\bar{x} = 1 - x$). The 3-SAT decision problem is to decide whether there exist an assignement $(x_1^* ... x_m^*)$ to the n bits such that

$$f(x_1^* ... x_n^*) \text{ is satisfied i.e equals 1.}$$

The space of all possible assignements has $N = 2^n$ elements. This problem can be shown to be NP-complete and no polynomial (in $m = \log_2 N$) time method of solution is known[1].

Both the database and 3-SAT problem are special cases of the following class of problems. Let $f : \{0,1\}^n \to \{0,1\}$ and $2^n = N$. We want to find a solution

[1] Here we assume that f has no obvious or no structure as in the unsorted database problem. We also remark that if we are given an Oracle that tells us the solution of the decision problem for each n, we can find the solutions $\bar{x_1} ... \bar{x_n}$ in linear time $O(n)$. Indeed set $x_1 = 1$ and ask the oracle if the reduced formula has a solution. If yes continue with $x_2 = 1$ etc. If not set $x_1 = 0$ and continue. In this fashion we have a solution in linear time

$x_1^\star ... x_n^\star$ such that

$$f(x_1^\star ... x_n^\star) = 1$$

The function f is assumed to be a "black box function" on which we know nothing special. We just assume that we have an oracle that we can query with an input $x_1 ... x_n$ and which tells us whether $f(x_1 ... x_n) = 0 or 1$. Classicaly, as argued before, we need to query the oracle $O(N) = O(2^n)$ times to get a solution with finite probability. We will now explain Grover's quantum algorithm which needs $O(\sqrt{N})$ queries in order to find a solution. This is not an exponential speedup but merely a quadratic one.

Remark that factoring allows for an exponential speedup because of the hidden structure or symmetry behind the search problem: reduction to period finding of an arithmetic function! Finally we mention without proof that it is known that certain classical problems do not admit any speedup by using QM.

12.2 Grover's quantum search algorithm

We are allowed to query a "quantum black box" and this counts as "one computational step". The quantum oracle is represented by a unitary operator U_f:

$$U_f|x\rangle|b\rangle = |x\rangle \otimes |b \oplus f(x)\rangle$$

where $|x\rangle \in \{|0\rangle, ..., |N-1\rangle\}$, $|b\rangle = |0\rangle, |1\rangle$ and $f(x) \in \{0, 1\}$. The goal is to find a solution to $f(x_1{}^\star, ..., x_N{}^\star) = 1$ in a minimal number of queries.

12.2.1 Derivation of the algorithm

In order to query with all classical inputs simultaneously we first prepare a superposition state:

$$(H \otimes ... \otimes H) \underbrace{(|0\rangle \otimes ... \otimes |0\rangle)}_{\text{n - Qbits}} = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle$$

Note

$$U_f \; H^{\otimes n}|0 ... 0\rangle \otimes |0\rangle = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle \otimes |f(x)\rangle$$

and

$$U_f \; H^{\otimes n}|0 ... 0\rangle \otimes |1\rangle = \frac{1}{N} \sum_{x=0}^{N-1} |x\rangle \otimes |1 \oplus f(x)\rangle$$

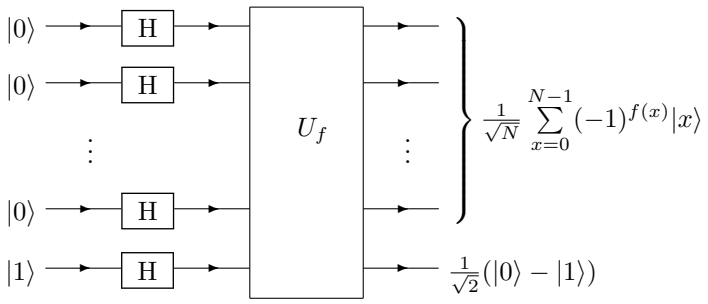
Note also

$$|f(x)\rangle - |1 \oplus f(x)\rangle = \begin{cases} |0\rangle - |1\rangle \text{ if } f(x) = 0 \\ |1\rangle - |0\rangle \text{ if } f(x) = 1 \end{cases}$$

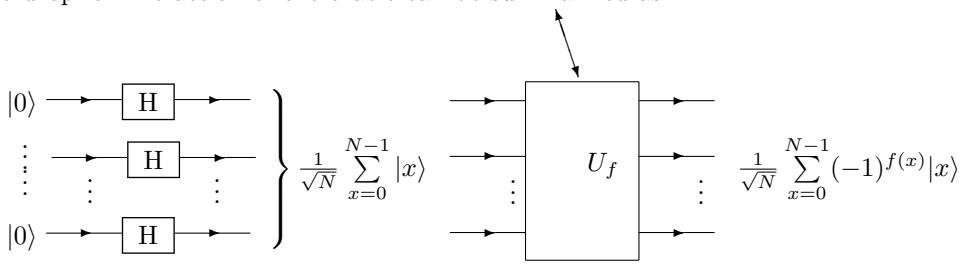
thus $|f(x)\rangle - |1 \oplus f(x)\rangle = (-1)^{f(x)}(|0\rangle - |1\rangle)$. These remarks imply

$$U_f H^{\otimes n}|0 \dots 0\rangle \otimes H|1\rangle = \frac{1}{\sqrt{N}} \sum_{x=0}^{N} (-1)^{f(x)}|x\rangle \otimes \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$$

The circuit representation of this last identity is



The extra bit $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ will play a trivial role in the sequel and it is customary to drop it. The action of the oracle can be summarized as



We may say that the oracle recognizes solutions $f(x) = 1$ and "marks them" with a phase (-1) while it leaves the phase (+1) to non-solutions.

Now suppose there are M solutions to $f(\vec{x}) = 1$. Define the normalized states:

$$\begin{cases} |S\rangle = \frac{1}{\sqrt{M}} \sum\limits_{\vec{x} \text{ solution}} |\vec{x}\rangle \\ |n\rangle = \frac{1}{\sqrt{N-M}} \sum\limits_{x \text{ not a solution}} |x\rangle \end{cases}$$

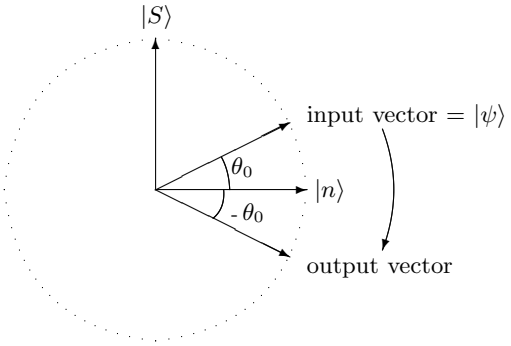
The input to the quantum oracle is

$$\frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle = \sqrt{\frac{M}{N}}|S\rangle + \sqrt{\frac{N-M}{N}}|n\rangle$$

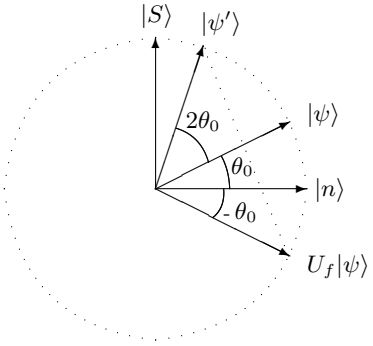
and the output is

$$\frac{1}{\sqrt{N}}\sum_{x=0}^{N-1}(-1)^{f(x)}|x\rangle = \sqrt{\frac{N-M}{N}}|n\rangle - \sqrt{\frac{M}{N}}|S\rangle$$

Setting $\sin\theta_0 = \sqrt{\frac{M}{N}}$ and $\cos\theta_0 = \sqrt{\frac{N-M}{N}}$ (note $\sin^2\theta_0 + \cos^2\theta_0 = 1$ is verified as it should) we see that the action of the oracle is the following reflection:



Let us now take the output vector and perform a reflection about the axis $\frac{1}{\sqrt{N}}\sum_{x=0}^{N-1}|x\rangle \equiv |\psi\rangle$:



This second reflection yields

$$|\psi'\rangle = (\cos 3\theta_0)|n\rangle + (\sin 3\theta_0)|S\rangle$$

Note that this vector is closer to $|S\rangle$ and this is the crucial point which makes Grover's algorithm work.

The second reflection can be performed even though its axis $|\psi\rangle$ is not known (indeed θ_0 is not known)! Let us explain this important point in more details.

Indeed a reflection about $|\psi\rangle$ is the unitary transformation:

$$|v\rangle \mapsto (2|\psi\rangle\langle\psi| - I)|v\rangle$$

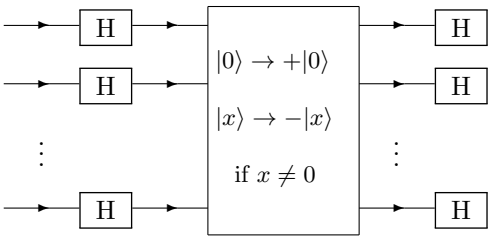
Now,

$$\begin{aligned} 2|\psi\rangle\langle\psi| - I &= 2H^{\otimes n}|0\dots0\rangle\langle0\dots0|H^{\otimes n} - I \\ &= H^{\otimes n}(2|0\dots0\rangle\langle0\dots0| - I)H^{\otimes n} \end{aligned}$$

and

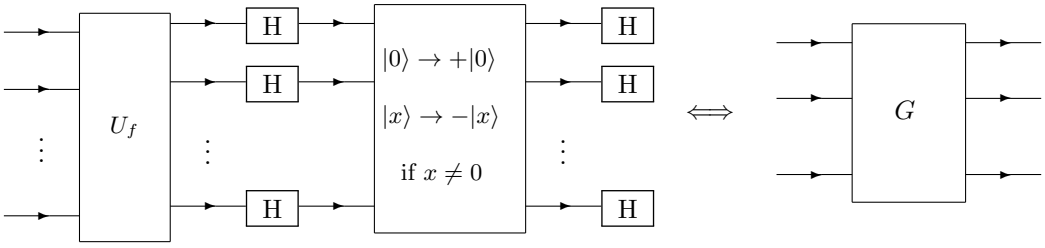
$$(2|0\dots0\rangle\langle0\dots0| - I)|v\rangle = \begin{cases} -|v\rangle \text{ if } v \neq 0\dots0 \\ +|0\dots0\rangle \text{ if } v = 0\dots0 \end{cases}$$

This shows that the circuit for the second reflection about $|\psi\rangle$ is:



The box can be represented by $O(n)$ elementary gates (exercise) [for this use a classical circuit, then make it reversible if needed and you have a quantum circuit].

The combination of the two reflections is defined as the "Grover operator" G



We have by the preceding discussion

$$G(\cos\theta_0|n\rangle + \sin\theta_0|S\rangle) = (\cos 3\theta_0)|N\rangle + (\sin 2\theta_0)|S\rangle$$

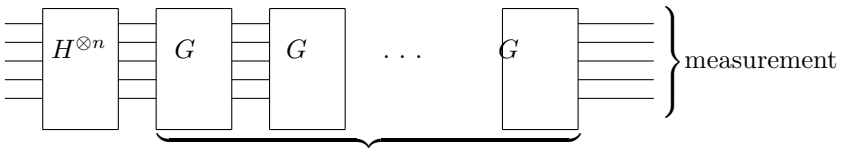
So the Grover operator is a rotation in the subspace $\{|N\rangle, |S\rangle\}$ by an angle $2\theta_0$.

The basic idea of the Grover algorithm is then to iterate this rotation

$$G^K(\cos\theta_0|N\rangle + \sin\theta_0|S\rangle) = (\cos(2K+1)\theta_0)|n\rangle + (\sin(2K+1)\theta_0)|S\rangle$$

in such a way that $\sin^2(2K+1)\theta_0 \approx 1$ and $\cos^2(2K+1)\theta_0 \approx 0$. Then a measurement will yield some state belonging to $|S\rangle = \frac{1}{\sqrt{M}} \sum_{x \text{ is a sol.}} |x\rangle$ with a prob equal to one.

The final quantum circuit for Grover's algorithm is



The size of this circuit is $O(n) + O(K) = (\log_2 N) + O(K)$. We will now see that in order to have $\sin^2(2K+1)\theta_0 = O(1)$ we must take $K = O(\sqrt{N})$.

12.2.2 Analysis of success probability: case where M assumed to be known.

- **Let first $M = 1$.** ("hardest case with one solution"). We have $\sin\theta_0 = \frac{1}{\sqrt{N}}$ so $\theta_0 \approx \frac{1}{\sqrt{N}}$. Since we want $\sin^2(2K+1)\theta_0 = O(1)$ we must iterate $K \approx \frac{\pi}{4}\sqrt{N}$ times. Thus $(2K+1)\theta_0 = (2[\frac{\pi}{4\theta_0}]+1)\theta_0 = \frac{\pi}{2} + 2\delta\theta_0$, where we used $[\frac{\pi}{4\theta_0}] = \frac{\pi}{4\theta_0} - \frac{1}{2} + \delta$ for $\delta < \frac{1}{2}$. Since $2\delta\theta_0 \approx \frac{2\delta}{\sqrt{N}} < \frac{2}{\sqrt{N}}$ we have $(2K+1)\delta_0 \approx \frac{\pi}{2} + O(\frac{1}{\sqrt{N}})$ The success probability is thus $\sin^2(2K+1)\theta_0 = \sin^2(\frac{\pi}{2} + O(\frac{1}{N})) = 1 - O(\frac{1}{N})$.

- **Let then $M = \frac{N}{4}$.** ("easiest" case quantum mechanically). We have $\sin\theta_0 = \sqrt{\frac{M}{N}} = \frac{1}{2} \Rightarrow \theta_0 = \frac{\pi}{6}$. Choose $K = 1$ one iteration. The success probability is $\sin^2\frac{3\pi}{6} = \sin^2\frac{\pi}{2} = 1$! With one iteration we find a solution! (remarkable).

- **Let M be general.**
  - If $M < \frac{3}{4}N$ then $\sin\theta_0 < \sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{2} \Rightarrow \theta_0 < \frac{\pi}{3}$. Iterate $K = [\frac{\pi}{4\theta_0}]$ times (which is at worst $O(\sqrt{N})$ if $M = 1$). Then $(2K+1)\theta_0 = (2[\frac{\pi}{4\theta_0}]+1)\theta_0 = \frac{\pi}{2} + 2\delta\theta_0$ where we used again $[\frac{\pi}{4\delta_0}] = \frac{\pi}{4\delta_0} - \frac{1}{2} + \delta$ for some $|\delta| < \frac{1}{2}$. Now $2|\delta|\theta_0 < \frac{\pi}{3}$ so the success probability is

$$\sin^2(2K+1)\theta_0 = \sin^2\left(\frac{\pi}{2} + 2\delta\theta_0\right) \geq \sin^2\left(\frac{\pi}{2} - \frac{\pi}{3}\right)$$
$$= \sin^2\frac{\pi}{6} = \frac{1}{4}$$

    So for $M < \frac{3}{4}N$ the success probability is $\frac{1}{4}$ which is enough because we can iterate the whole process to make it as close to 1 as we wish.
  - If $M \geq \frac{3}{4}N$ then forget about the quantum algorithm and pick $x \in$

$\{0 \dots M-1\}$ randomly uniformly. The success probability is at least $\frac{3}{4}$.

12.2.3 Case where M is unknown

In this case we can use the following version of Grover's algorithm.

1. Pick x at random. If $f(x) = 1$ output x and stop.
2. Otherwise let $\widetilde{M} = \sqrt{N}+1$. Choose $R \in \{0, 1, \dots, \widetilde{M}-1\}$ at random uniformly.
3. Apply Grover's algorithm with R iterations.
4. Measure the output to get some $x \in \{0, \dots, N-1\}$.

Let us prove that the success probability is at least $\frac{1}{4}$. Let M be the unknown number of solutions. If $M \geq \frac{3}{4}N$ we succeed with probability $\frac{3}{4}$ in the first step. If $M < \frac{3}{5}N$ we may not succeed and then we go to the second step. Then given R the success probability is $\sin^2(2R+1)\theta_0$ where $\sin\theta_0 = \sqrt{\frac{M}{N}}$. As before $\theta_0 < \frac{\pi}{3}$. Since R is chosen uniformly at random in $\{0 \dots \widetilde{M}-1\}$ the success probability is

$$\begin{aligned}
\text{prob. success} &= \frac{1}{R}\sum_{R=0}^{\widetilde{M}-1} \sin^2(2R+1)\theta_0 \\
&= \frac{1}{2R}\sum_{R=0}^{\widetilde{M}-1} (1-\cos((2R+1)2\theta_0)) \\
&= \frac{1}{2} - \frac{\sin 4\widetilde{M}\theta_0}{4\widetilde{M}\sin 2\theta_0}
\end{aligned}$$

Now we have $\sin 4\widetilde{M}\theta_0 < 1$ and

$$\begin{aligned}
\sin 2\theta_0 &= 2\sin\theta_0\cos\theta_0 \\
&= 2\sqrt{\frac{M}{N}}\sqrt{\frac{N-M}{M}} \\
&> \frac{1}{\sqrt{N}} > \frac{1}{\widetilde{M}}
\end{aligned}$$

Thus probability of success is greater than $\frac{1}{2} - \frac{1}{4} = \frac{1}{4}$.

12.3 Optimality of Grover's search algorithm.

Consider the example of database search. Suppose that we have some "marked" state x^* that we search for using Grover's algorithm. We need at most $O(\sqrt{N})$ steps to achieve success with finite probability. But can we do better? The answer is no! In other words we need at least $\Omega(\sqrt{N})$ steps!

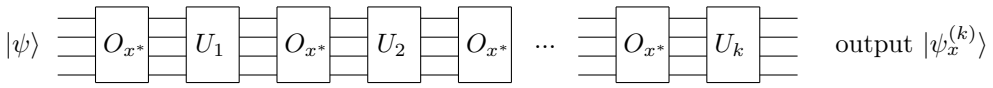
We suppose that we have oracles that recognize marked states x^*. Since they recognize these states the oracles can perform the unitary operations:

$$O_{x^*} = I - 2|x^*\rangle\langle x^*|$$

which acts as

$$O_{x^*}|v\rangle = \begin{cases} |v\rangle \text{ if } v \neq x^* \\ -|v\rangle \text{ if } v = x^* \end{cases}$$

Now, we suppose that we have an arbitrary initial state $|\psi\rangle$, and that we query the oracle O_{x^*}, K times as follows:



where $U_1 ... U_K$ are arbitrary unitary operators making up the general search algorithm. The output of this algorithm is

$$|\psi_{x^*}^{(K)}\rangle = U_K O_{x^*} U_{K-1} \ ... \ U_1 O_{x^*} |\psi\rangle$$

In order to recognize the marked state x^* with finite probability $\epsilon > 0$ we ask that this search procedure should satisfy

$$|\langle x^*|\psi_x^{(K)}\rangle|^2 \geq \epsilon > 0$$

We also ask that the choice of $U_1, \cdots, U_K$ is independent of the marked state x^*. In other words we want that the same algorithm works irrespective of the problem (otherwise this means that we use specific information about the database or the problem). Thus the last condition should be true for any x.

We will show that necessarily we must choose $K \geq c\sqrt{N}$ for some $c > 0$ (depending on ϵ).

Lemma 1:

Set $D_K = \sum_{x=o}^{N-1} ||\psi_x^{(K)} - \psi_K||^2$ where $|\psi_K\rangle = U_K U_{K-1} ... U_1 |\psi\rangle$. Then $D_K \leq 4K^2$.

Lemma 2:

We also have $D_K \geq cN$ for some c depending on ϵ.

Lemmas 1 and 2 imply $cN \leq 4K^2 \Rightarrow K = \Omega(\sqrt{N})$ so at least $\sqrt{N}$ queries of the oracle O_x are needed.

Proof of Lemma 2:

The condition $|\langle x|\psi_x^{(K)}\rangle|^2 \geq \epsilon > 0$ for all x will be used. We have

$$\begin{aligned}
D_K &= \sum_x ||(\psi_x^{(K)} - |x\rangle) - (\psi_K - |x\rangle)||^2 \\
&\geq \sum_x ||\psi_x^{(K)} - |x\rangle||^2 + \sum_x ||\psi_K - |x\rangle||^2 \\
&- 2\sum_x ||\psi_x^{(K)} - |x\rangle|| \cdot ||\psi_K - |x\rangle|| \\
&\geq \sum_x ||\psi_x^{(K)} - |x\rangle||^2 + \sum_x ||\psi_K - |x\rangle||^2 \\
&- 2\Big(\sum_x ||\psi_x^{(K)} - |x\rangle||^2\Big)^{\frac{1}{2}} \Big(\sum_x ||\psi_K - x||^2\Big)^{\frac{1}{2}} \\
&= \Big[\Big(\sum_x ||\psi_x^{(K)} - x||^2\Big)^{\frac{1}{2}} - \Big(\sum_x ||\psi_K - |x\rangle||^2\Big)^{\frac{1}{2}}\Big]^2
\end{aligned}$$

Now

$$\begin{aligned}
||\psi_x^{(K)} - x||^2 &= 2 - \langle x|\psi_x^{(K)}\rangle - \langle \psi_x^{(K)}|x\rangle \\
&\leq 2 - 2\sqrt{\epsilon}
\end{aligned}$$

which implies

$$(\sum_x ||\psi_x^{(K)} - x||)^{\frac{1}{2}} \leq \sqrt{2}\sqrt{N}(1 - \sqrt{\epsilon})^{\frac{1}{2}}$$

Moreover

$$||\psi_K - |x\rangle||^2 = 2 - \langle \psi_K|x\rangle - \langle x|\psi_K\rangle$$

which implies

$$\begin{aligned}
\sum_x ||\psi_K - x||^2 &= 2N - \sum_x \langle \psi_K|x\rangle - \sum_x \langle x|\psi_K\rangle \\
&\geq 2N - \Big(\sum_x \cdot 1\Big)^{\frac{1}{2}} \Big(\underbrace{\sum_x |\langle \psi_K|x\rangle|^2}\Big)^{\frac{1}{2}} \\
&= 2N - 2\sqrt{N}
\end{aligned}$$

In the first inequality we used Cauchy-Schwarz and in the last equality we used the fact that $|x\rangle$ is a complete set of states. Combining the above results yields

$$\begin{aligned}
D_K &\geq \Big[\sqrt{2N}(1 - \sqrt{\epsilon})^{\frac{1}{2}} - \sqrt{2N}(1 - \frac{1}{\sqrt{N}})^{\frac{1}{2}}\Big]^2 \\
&\geq 2N\Big[(1 - \frac{1}{\sqrt{N}})^{\frac{1}{2}} - (1 - \sqrt{\epsilon})^{\frac{1}{2}}\Big]^2
\end{aligned}$$

This completes the proof of Lemma 2.

Proof of Lemma 1:

This is a consequence of $O_x = I - 2|x\rangle\langle x|$. We proceed by induction. For $K =$

$0\ ;\ D_0 = 0$. Suppose $D_K \leq \psi K^2$. Compute D_{K+1}:

$$D_{K+1} = \sum_x ||O_x \psi_x^{(K)} - \psi_K||^2 \quad \text{because } U_{K+1} \text{ has norm 1.}$$

$$= \sum_x ||O_x(\psi_x^{(K)} - \psi_K) - (O_x - I)\psi_K||^2$$

$$\leq \sum_x ||\psi_x^{(K)} - \psi_K||^2 + \sum_x ||(O_x - I)\psi_K||^2$$

$$+2\sum_x ||\psi_x^{(K)} - \psi_x|| \cdot ||O_x - I|| \quad (*)$$

Now $(O_x - I)\psi_K = -2\langle x|\psi_x\rangle \cdot |x\rangle$

$\Rightarrow ||(O_x - I)\psi_K||^2 = \psi|\langle x|\psi_K\rangle|^2$

$\Rightarrow \sum_x ||(O_x - I)\psi_K||^2 = \psi \sum_x |\langle x|\psi_K\rangle|^2 = \psi \quad (**)$

From $(*), (**)$ and Cauchy-Schwarz on the last term we get

$$D_{K+1} \leq D_K + \psi + \psi\sqrt{D_K} \leq \psi K^2 + \psi + 8K = \psi(K+1)^2$$

This completes the proof of lemma 1.

12.4 Phase estimation and quantum counting

Suppose we want to count the number of solutions to a problem of the form $f(x_1 \ldots, x_N) = 1, N = 2^n, x_i \in \{0, 1\}$. This could be for example the 3-SAT problem. Classically in the worst case this may require $\Omega(N)$ queries of an oracle computing f. The "quantum counting algorithm" provides a way to count the number of solutions in $O(\sqrt{N})$ queries of the oracle. This algorithm is a combination of our two more basic quantum operations: the QFT and the Grover operator G. Here we wish to give a brief sketch of the main ideas involved and refer to the literature for a detailled analysis.

Looking back at Grover's algorithm we see that the number of solutions, say M, appears as an eigenvalue of the Grover rotation by an angle $2\theta_0$:

$$|\psi\rangle = \cos\theta_0|n\rangle + \sin\theta_0|S\rangle \to \cos 3\theta_0|n\rangle + \sin\theta_0|S\rangle$$

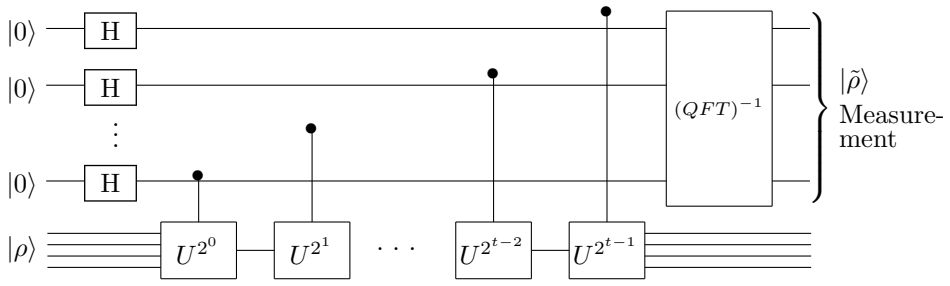
where $\sin\theta_0 = \sqrt{\frac{M}{N}}, \cos\theta_0 = \sqrt{\frac{N-M}{N}}$. This rotation is a unitary operator with eigenvalues $e^{\pm i2\theta_0}$ and corresponding eigenvectors $|2\theta_0\rangle, |-2\theta_0\rangle$:

$$G|\pm 2\theta_0\rangle = e^{\pm i2\theta_0}|\pm 2\theta_0\rangle$$

So counting the number of solutions M means estimating the phase of G. In the following paragraph we give a general algorithm for estimating the phase (eigenvalue) of a unitary when the corresponding eigenvector is known.

12.4.1 Phase estimation algorithm

Estimating the phase of a unitary operator U is almost like finding the period of the modular exponential function in Shor's algorithm. It should come as no surprise that we try out the following circuit (observe the ressemblance without Shor's factoring circuit):



Here $U|\rho\rangle = e^{i2\pi\rho}|\rho\rangle$ and we want to estimate the phase ρ. Let us show that the output $|\tilde{\rho}\rangle$ is a good approximation to $|\rho\rangle$.

For simplicity suppose $\rho = 2^{-t}\rho_t + ... + 2^{-2}\rho_2 + 2^{-1}\rho_1; \rho_i \in \{0,1\}$. [Note we define the phase as $0 < \rho < 1$]. The above circuit uses t Qbits $|0\rangle \otimes ... \otimes |0\rangle$ as input and t extra Qbits to input $|\rho\rangle$. The action of all controlled unitary gates yields (just before $(\text{QFT})^{-1}$)

$$\prod_{l=0}^{t-1}(|0\rangle + e^{(2\pi i 2^l \rho)}|1\rangle)$$

Since $\rho = 2^{-t}\rho_t + ... + 2^{-1}\rho_1 = 2^{-t}(\underbrace{2^{t-1}\rho_1 + ... + 2^0}\rho_t)$

$$= 2^{-t}\tilde{\rho}$$

we have just before the $(\text{QFT})^{-1}$ gate:

$$\prod_{l=0}^{t-1}(|0\rangle + e^{\frac{2\pi i 2^l}{2^t}\tilde{\rho}}|1\rangle) \qquad (*)$$

But looking back at the QFT we have that this state is precisely

$$\text{QFT } |\tilde{\rho}\rangle \text{ or QFT } |\rho_t ..., \rho_1\rangle$$

So when $(\text{QFT})^{-1}$ acts on $(*)$ we obtain $|\rho_1 ..., \rho_t\rangle = |\rho\rangle$. Thus a measurement will yield the phase ρ with probability equal to 1!

Of course in practice ρ has more than t bits, but it can be shown that this circuits enables to estimate ρ to t bits of accuracy with high probability. The error is basically

$$|\tilde{\rho} - \rho| < 2^{-t}$$

Moreover in practice the eigenstate $|\rho\rangle$ at the entry is unknown. So one prepares a suitable superposition of eigenstates of U namely $|v\rangle = \sum_u c_u|\rho_u\rangle$. Then at the output we will measure some ρ_u with prob $|c_u|^2$. When U has a small number of eigenstates and eigenvalues this works pretty well since $\sum_u |c_u|^2 = 1$ so some of them have to be finite $O(1)$.

12.4.2 Application to quantum counting

In place of U we put G, the Grover operator, in the previous circuit $(*)$. Thus we get an estimate of θ_0 to t bits of accuracy. We want to determine the number of bits t needed to get a reasonable estimate for M. We have

$$\theta_0 \approx \sqrt{\frac{M}{N}} \quad \text{for} \quad M << N.$$

Thus

$$\delta\theta_0 \approx \frac{\delta M}{\sqrt{MN}} \approx \frac{\sqrt{M}}{\sqrt{MN}} = \frac{1}{\sqrt{N}} \text{ if } \delta M \sim \sqrt{M}.$$

To get an estimate of M with an error of $\sqrt{M}$ we need that $2^{-t} = \frac{1}{\sqrt{N}}$ i.e $t = \log_2 \sqrt{N}$ bits.

How many times is the oracle queried? In each box G^{2^l} the Grover operator (and thus the oracle) is queried 2^l times. Thus the total number of queries is

$$2^0 + 2^1 + 2^2 + ... + 2^{t-1} = 2^t = 2^{\log_2 \sqrt{N}} = \sqrt{N}$$

We query the oracle $\sqrt{N}$ times to estimate the number of solutions M to an accuracy of order $\sqrt{M}$.

Finally, note that the eigenvector (input of the circuit) $|2\theta_0\rangle$ is not known but we can prepare the input

$$\frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle = \cos\theta_0|N\rangle + \sin\theta_0|S\rangle = c_{\theta_0}|2\theta_0\rangle + c_{-\theta_0}|-2\theta_0\rangle$$

At least one of the two coefficients is $> \frac{1}{\sqrt{2}}$. Thus we will obtain the corresponding pahse of the Grover rotation we finite probability.

13 Quantum Error Correction

Notes