Quantum information 1/2

Konrad Banaszek, Rafał Demkowicz-Dobrzański

June 1, 2012

Contents

Chapter 1

Qubit

1.1 Light polarization

The starting point of our discussion will be a plane electromagnetic wave with the frequency ω propagating along the z axis. Because the electromagnetic field is transverse, the electric field $\mathbf{E}(z,t)$ oscillates in the plane perpendicular to the propagation direction and can be written as a superposition of two components:

$$\mathbf{E}(z,t) = \mathbf{e}_x E_x(z,t) + \mathbf{e}_y E_y(z,t) \tag{1.1}$$

where $\mathbf{e}_x$ and $\mathbf{e}_y$ are unit vectors oriented along the x and the y axis respectively. The two components have in general the following form:

$$E_x(z,t) = E_{0x} \cos(kz - \omega t + \varphi_x) \tag{1.2}$$

$$E_y(z,t) = E_{0y} \cos(kz - \omega t + \varphi_y) \tag{1.3}$$

where E_{0x} and E_{0y} characterize the amplitudes of oscillations and φ_x and φ_y are respective phases. The wave vector k is given by the frequency ω divided by the speed of light.

It will be convenient to associate polarizations with shapes drawn by the tip of the electromagnetic field vector observed when facing the incident wave at a fixed point in space. For example, two rectilinear cases when only x or y components are non-zero correspond to the horizontal (↔) and the vertical (↕) polarization respectively. When $E_{x0} = E_{y0}$ there are four worthwhile cases. The electric field oscillates along diagonal directions when $\varphi_x = \varphi_y$ (⤢) or $\varphi_x = \varphi_y + \pi$ (⤡). If $\varphi_x = \varphi_y + \pi/2$ the length of the electric field vector is constant and rotates counterclockwise (↺), while for $\varphi_x = \varphi_y - \pi/2$

we have clockwise rotation (↻). In the general, the electric field vector draws an ellipse. These two polarizations are called circular. The general case is considered in Exercise 1.1.1.

It is very helpful to use a two-element complex vector, known in optics as *Jones vector*, constructed from the parameters characterizing the electric field:

$$\boldsymbol{\mathcal{E}} = \begin{pmatrix} \mathcal{E}_x \\ \mathcal{E}_y \end{pmatrix} = \begin{pmatrix} E_{0x}\mathrm{e}^{\mathrm{i}\varphi_x} \\ E_{0y}\mathrm{e}^{\mathrm{i}\varphi_y} \end{pmatrix} \tag{1.4}$$

The electric field can then be written simply as:

$$\mathbf{E}(z,t) = \mathrm{Re}(\boldsymbol{\mathcal{E}}\mathrm{e}^{\mathrm{i}kz-\mathrm{i}\omega t}). \tag{1.5}$$

where the z component is equal to zero by default. Let us note that multiplying the Jones vector by an overall complex phase is equivalent to shifting time and it does not change the figure drawn by the tip of the electric field.

The Jones vector is a convenient tool to describe transformations of the electromagnetic field by linear optical elements. One standard element is a polarizer shown in Fig. 1.1(a), which separates horizontal and vertical polarization components. The output beams are described by vectors with one of the components replaced by zero and can be obtained by the following linear transformations of the input Jones vector:

$$\begin{pmatrix} \mathcal{E}_x \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathcal{E}_x \\ \mathcal{E}_y \end{pmatrix}, \qquad \begin{pmatrix} 0 \\ \mathcal{E}_y \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathcal{E}_x \\ \mathcal{E}_y \end{pmatrix} \tag{1.6}$$

A matrix that describes a linear transformation of the Jones vector is called a *Jones matrices.*

A *wave plate* shown in Fig. 1.1(b) is an optical element made of birefringent material that has two different refractive indices for two orthogonal directions, called *principal axes.* If the axes are aligned with the coordinate systems, this means that phases φ_x and φ_y of the two components of the Jones vector are changed by different values α_x and α_y respectively. The transformation of the Jones vector can be written as

$$\begin{pmatrix} \mathrm{e}^{\mathrm{i}\alpha_x}\mathcal{E}_x \\ \mathrm{e}^{\mathrm{i}\alpha_y}\mathcal{E}_y \end{pmatrix} = \mathrm{e}^{\mathrm{i}(\alpha_x+\alpha_y)/2} \begin{pmatrix} \mathrm{e}^{\mathrm{i}\alpha/2} & 0 \\ 0 & \mathrm{e}^{-\mathrm{i}\alpha/2} \end{pmatrix} \begin{pmatrix} \mathcal{E}_x \\ \mathcal{E}_y \end{pmatrix} \tag{1.7}$$

where $\alpha = \alpha_x - \alpha_y$. Because we are not interested in the overall phase of the Jones vector, we will ignore in the following the overall phase factor

Figure 1.1: (a) Polarizer. (b) Wave plate.

$e^{i(\alpha_x+\alpha_y)/2}$ and characterize the action of a wave plate with α. Wave plates introducing $\alpha = \pi/2$ and $\alpha = \pi$ relative phase shifts are called respectively a *quarter-wave plate* and a *half-wave plate*.

Suppose now that a wave plate is oriented at an angle β with respect to our coordinate system. To calculate the corresponding Jones matrix, we need to switch to the coordinate system rotated by β about the z axis, apply the wave plate transformation and go back. This gives:

$$\begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix} \begin{pmatrix} e^{i\alpha/2} & 0 \\ 0 & e^{-i\alpha/2} \end{pmatrix} \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix}$$
$$= \begin{pmatrix} \cos\frac{\alpha}{2} + i\sin\frac{\alpha}{2}\cos 2\beta & i\sin\frac{\alpha}{2}\sin 2\beta \\ i\sin\frac{\alpha}{2}\sin 2\beta & \cos\frac{\alpha}{2} - i\sin\frac{\alpha}{2}\cos 2\beta \end{pmatrix} \tag{1.8}$$

Note that any matrix of this form is unitary and special. Some examples of wave plate transformations are analyzed in Exercise 1.1.2.

1.1.1 Calculate the orientation and length of principal axes for a plane electromagnetic wave whose electric field is characterized by parameters E_{0x}, E_{0y}, φ_y. For simplicity, assume that $\varphi_x = 0$.

1.1.2 What will happen to a linearly polarized send to a half-wave plate oriented at an angle ϑ with respect to input polarization? Show that a quarter wave plate at 45° transforms horizontal polarization into a circular one.

1.1.3 How circular polarization is changed when the coordinate system is rotated by an angle θ in the xy plane?

1.2 Polarization qubit

Let us consider an elementary experiment with light polarization. A light beam is sent to a polarizer whose output ports are monitored by photodetectors. The intensities measured by the detectors will be proportional to squared absolute values $|\mathcal{E}_x|^2$ and $|\mathcal{E}_y|^2$ of the elements of the Jones vector describing the input beam. Suppose now that we decrease the amplitude of the incident wave and detect light with very sensitive photodetectors, such as photomultipliers. For very low light levels the response of the detectors consists of "clicks" that herald generation of individual photoelectrons by the incident light. As suggested first by Einstein, the photoelectrons are generated by absorption of elementary quantum portions of the electromagnetic field called *photons*.

A meaningful question one may now ask is what happens if we send a single photon to the polarizer—which of the two detectors will register it? All experimental facts we know by now point to the conclusion that the outcome is probabilistic: everything that can be predicted is the chance that one or another detector will click. Therefore we need a theory that incorporates this randomness. The complete quantum theory of electromagnetic fields is rather complicated. But if we are interested only in a single degree of freedom, such as polarization, we may take a shortcut and introduce a simplified quantum description of a single photon. It turns out that the polarization of a single photon is described by an object analogous to the Jones vector. It has two complex components ψ_x and ψ_y, but their interpretation is now different: their squared absolute values $|\psi_x|^2$ and $|\psi_y|^2$ specify the *probabilities* that the photon will generate a click on one or another detector.

Because there is no other path for the photon to take at the exit, we require that the *normalization* condition $|\psi_x|^2 + |\psi_y|^2 = 1$ is satisfied. A macroscopic light beam can be thought of as composed of a large number of photons with the same polarization. Therefore it is natural to assume the polarization state of an individual photon is described by the Jones vector rescaled to satisfy the normalization condition. When many photons are sent to the polarizer, this will reproduce the division of classical intensities between the output port. For example, a photon polarized linearly at an angle θ will be described by a vector $\left(\begin{smallmatrix}\cos\theta\\ \sin\theta\end{smallmatrix}\right)$, and the probabilities of clicks are $\cos^2\theta$ and $\sin^2\theta$. This is the quantum analog of the Malus law.

In quantum theory, the object describing the state of a physical system is called a *state vector*. Dirac introduced a convenient notation in which a state vector is written as:

$$|\psi\rangle \equiv \begin{pmatrix}\psi_x\\ \psi_y\end{pmatrix} \tag{1.9}$$

The components of the state vector are called *probability amplitudes*. The column form of a state vector, denoted with a symbol closed with an angular bracket on the right hand side, is called a *ket* for a reason that will become clear in a moment.

It will be useful to denote horizontal and vertical polarization states of a single photon are:

$$|{\leftrightarrow}\rangle \equiv \begin{pmatrix}1\\ 0\end{pmatrix}, \quad |{\updownarrow}\rangle \equiv \begin{pmatrix}0\\ 1\end{pmatrix}. \tag{1.10}$$

The states $|{\leftrightarrow}\rangle$ and $|{\updownarrow}\rangle$ can be identified unambiguously using a polarizer. If we are tasked with encoding a classical message in the form of a string of bits into the polarization of a train of photons, the solution is straightforward: send the bit value 0 as $|{\leftrightarrow}\rangle$, the bit value 1 as $|{\updownarrow}\rangle$ and tell the receiving party to read out the message using a polarizer and two single-photon detectors. However, quantum mechanics offers us a possibility to prepare an arbitrary superposition state which can be seen most directly by rewriting Eq. (1.9) to the form

$$|\psi\rangle = \psi_x|{\leftrightarrow}\rangle + \psi_y|{\updownarrow}\rangle \tag{1.11}$$

where $|\psi_x|^2 + |\psi_y|^2 = 1$. This quantum mechanical generalization of the bit is called a *qubit* and it will be the main character of this story.

1.3 States and operators

It is useful to introduce special notation for frequently occuring polarization states. Diagonal polarization states at ±45° are denoted as:

$$|\text{⤢}\rangle \equiv \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}, \qquad |\text{⤡}\rangle \equiv \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix} \tag{1.12}$$

while it is natural to write the pair of circular polarization states as:

$$|\circlearrowleft\rangle \equiv \frac{1}{\sqrt{2}}\begin{pmatrix}1\\\mathrm{i}\end{pmatrix}, \qquad |\circlearrowright\rangle \equiv \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-\mathrm{i}\end{pmatrix}. \tag{1.13}$$

The algebraic procedure of hermitian conjugation transforms a ket $|\psi\rangle$ into a horizontal vector with complex-conjugated entries, which is called a *bra* and denoted as:

$$\langle\psi| = \big(|\psi\rangle\big)^\dagger \equiv (\psi_x^*, \psi_y^*)$$

A bra $\langle\psi|$ multiplying from the left side a ket $|\chi\rangle$ is simply the *scalar product* of two state vectors. It is customary to draw just a single vertical line between the bra and the ket when writing a scalar product:

$$\langle\psi|\chi\rangle \equiv (\psi_x^*, \psi_y^*)\begin{pmatrix}\chi_x\\\chi_y\end{pmatrix} = \sum_{j=x,y}\psi_j^*\chi_j$$

The object on the left-hand side has the form of a bracket which inspired Dirac to name the two halves of this expression a bra and a ket. The scalar product has the standard property $\langle\chi|\psi\rangle = \big(\langle\psi|\chi\rangle\big)^*$ for any pair of state vectors. The normalization condition for a state vector $|\psi\rangle$ can be written as $\langle\psi|\psi\rangle = 1$.

Mathematically, the state vectors belong to a two-dimensional complex vector space equipped with a scalar product. Let us write some basis algebraic facts using Dirac notation. The pairs of states defined in Eqs. (1.10), (1.12), and (1.13) are normalized and mutually orthogonal, i.e. the scalar product between the state vectors is zero. An arbitrary state vector can be represented as a linear combination of such a pair, which we will write in general as:

$$|\psi\rangle = \sum_k \psi_k|u_k\rangle \tag{1.14}$$

where the index k runs over the values 0 and 1, and $|u_0\rangle, |u_1\rangle$ stands for any of the pairs. The normalization and orthogonality conditions can be written jointly as:

$$\langle u_j|u_k\rangle = \delta_{jk}, \tag{1.15}$$

where δ_{jk} is the Kronecker delta. The coefficients ψ_k in the decomposition (1.14) can be found by projecting both sides of the above equation onto the bra $\langle u_j|$, which yields $\psi_j = \langle u_j|\psi\rangle$.

Optical elements discussed in Sec. 1.1 transform the state vector in a completely analogous way to the classical Jones vector. In the quantum context, such linear transformations are called *operators* and usually denoted by capital letters with carets. Let us consider a transformation $|\psi'\rangle = \hat{U}|\psi\rangle$. If we decompose $|\psi\rangle = \sum_k \psi_k|u_k\rangle$, then the coefficients for the transformed state $|\psi'\rangle$ can be written as:

$$\psi'_j = \langle u_j|\hat{U}|\psi\rangle = \sum_k \langle u_j|\hat{U}|u_k\rangle\psi_k \tag{1.16}$$

Thus when $|\psi\rangle$ and $|\psi'\rangle$ are written in the column vector form in the orthonormal basis $|u_0\rangle$, $|u_1\rangle$, the action of $\hat{U}$ is represented as multiplication by the matrix:

$$\hat{U} \equiv \begin{pmatrix} \langle u_0|\hat{U}|u_0\rangle & \langle u_0|\hat{U}|u_1\rangle \\ \langle u_1|\hat{U}|u_0\rangle & \langle u_1|\hat{U}|u_1\rangle \end{pmatrix}. \tag{1.17}$$

An important class of operators are those which preserve normalization of state vectors. For a qubit, this means that we need to satisfy the condition $\langle\psi|\hat{U}^\dagger\hat{U}|\psi\rangle = 1$ for any state vector. This in turn implies that $\hat{U}^\dagger\hat{U} = \mathbb{1}$. We will call these operators *unitary*.

There are three so-called *Pauli operators* that will appear frequently in our discussions. In the rectilinear basis they are given by matrices

$$\hat{\sigma}_1 \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{\sigma}_2 \equiv \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \hat{\sigma}_3 \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.18}$$

It is easy to see that the pairs of diagonal, circular, and rectilinear polarization states are respective eigenstates of these three operators with eigenvalues ±1.

1.3.1 Design a setup to distinguish two arbitrary orthogonal states of a single photon.

1.3.2 Show that a product of two Pauli matrices can be written as $\hat{\sigma}_k\hat{\sigma}_l = \delta_{kl}\hat{\mathbb{1}} + \mathrm{i}\epsilon_{klm}\hat{\sigma}_m$, where ϵ_{klm} is the Levi-Civita permutation symbol.

1.3.3 Show that any qubit operator $\hat{A}$ can be written as a decomposition $\hat{A} = a_0\hat{\mathbb{1}} + a_1\hat{\sigma}_1 + a_2\hat{\sigma}_2 + a_3\hat{\sigma}_3$, where $a_0 = \frac{1}{2}\mathrm{Tr}(\hat{A})$ and $a_k = \frac{1}{2}\mathrm{Tr}(\hat{A}\hat{\sigma}_k)$, $k = 1, 2, 3$.

1.4 Quantum random access codes

An obvious question is whether a qubit entity can be used to transmit more than one bit of classical information. For example, we could try to encode four classical messages (i.e. two bits) into states $|{\leftrightarrow}\rangle, |{\updownarrow}\rangle, |{\swarrow\!\!\!\!\nearrow}\rangle, |{\nwarrow\!\!\!\!\searrow}\rangle$. But these states would not give distinguishable outcomes at a polarizer: states $|{\swarrow\!\!\!\!\nearrow}\rangle$ and $|{\nwarrow\!\!\!\!\searrow}\rangle$ would give completely random clicks. One might suspect that this is because a polarizer is a too simple device for that task and we could devise a more complex measurement scheme that would enable as to discriminate these four states. It turns out that this is not possible, even in principle. The fundamental reason for that only states that are orthogonal can be distinguished unambiguously. Non-orthogonal states can be distinguished only with partial success, and this actually does not increase our capacity to trasfer classical information. We will discuss the problem of distinguishability in Ch. **??** and the actual capacity in Ch. **??**. Note that any pair of orthogonal states can be used to transmit classical information with the capacity of one bit, see Exercise 1.3.1.

There are however scenarios, in which sending a *qubit* gives us an advantage compared to a transmission of a classical bit. One of the simplest ones is the problem of *random access codes.* Consider two parties, Alice and Bob. In the simplest version of the protocol, Alice has been given two bits of information she is supposed to pass over to Bob. Bob will need only one of these bits, but it will be known which one of them is actually needed only *after* Alice and Bob have been given an opportunity to communicate. Furthermore, Alice is allowed to transmit only one bit of information to Bob. If the chance that Bob will need one or another bit of information is 50/50, then the optimal protocol is to transmit the first bit of information to Bob. That way if the first bit is needed he will know its value for sure, and if the second one is needed he chooses its value at random. The overall probability that Bob will have the correct value of the bit he needs is thus 75%.

What happens when Alice can send to Bob one qubit instead of one bit? It will be convenient to denote by $|\vartheta\rangle$ a linear polarization state at an angle $\frac{\vartheta}{2}$ with respect to the horizontal plane:

$$|\vartheta\rangle = \cos\vartheta|\leftrightarrow\rangle + \sin\vartheta|\updownarrow\rangle \tag{1.19}$$

Let Alice prepare the following four linear polarization states of the qubit depending on the pair of bits she would like to transmit to Bob:

$$00 \equiv |22.5^\circ\rangle, \quad 01 \equiv |-22.5^\circ\rangle, \quad 10 \equiv |67.5^\circ\rangle, \quad 11 \equiv |112.5^\circ\rangle. \tag{1.20}$$

Suppose now that Bob can wait with measuring the received qubit until he knows whether the first or the second qubit is needed. If he needs the value of the first bit, he measures sets his polarizer to distinguish horizontal and vertical polarization. The probability that he will obtain the correct bit value is $\cos^2 22.5^\circ = (1+1/\sqrt{2})/2 \approx 85\%$. If the value of the second bit is needed, Bob rotates the polarizers by 45°. It is easy to verify that the probability of success stays the same. Thus the average success rate exceeds that of the optimal classical protocol by approximately 10%.

1.4.1 Devise a quantum random access code for a generalized problem when the values of the first and the second qubit are needed with probabilities p and $1-p$.

1.5 Bloch sphere

There is a convenient way to visualize the state of a qubit in three-dimensional real space. It is based on the *Bloch vector*, which for a state vector $|\psi\rangle = \psi_x|\leftrightarrow\rangle + \psi_y|\updownarrow\rangle$ is defined as

$$\mathbf{s} = \begin{pmatrix} \langle\psi|\hat{\sigma}_1|\psi\rangle \\ \langle\psi|\hat{\sigma}_2|\psi\rangle \\ \langle\psi|\hat{\sigma}_1|\psi\rangle \end{pmatrix} = \begin{pmatrix} \psi_y^*\psi_x + \psi_x^*\psi_y \\ \mathrm{i}(\psi_y^*\psi_x - \psi_x^*\psi_y) \\ |\psi_x|^2 - |\psi_y|^2. \end{pmatrix} \tag{1.21}$$

It is straightforward to see that the three components of the Bloch vector are that real and that if the state $|\psi\rangle$ is normalized, then its length is equal to one, $|\mathbf{s}| = 1$. We denote here by a dot the standard scalar product in three-dimensional real space. The Bloch vector can be formally written as $\mathbf{s} = \langle\psi|\hat{\boldsymbol{\sigma}}|\psi\rangle$, where $\hat{\boldsymbol{\sigma}} = \begin{pmatrix}\hat{\sigma}_1\\ \hat{\sigma}_2\\ \hat{\sigma}_3\end{pmatrix}$ is a column vector with three Pauli operators as its components.

The Bloch vector contains all relevant information about the quantum state. In order to verify that, let us note that the normalization condition $\psi_x|^2+|\psi_y|^2$ allows us to write the absolute values of the probability amplitudes as $|\psi_x| = \cos\frac{\theta}{2}$ and $|\psi_y| = \sin\frac{\theta}{2}$, where $0 \leq \theta \leq \pi$. Furthermore, if the overall phase of the state vector does not matter, we can introduce only one phase factor $\mathrm{e}^{\mathrm{i}\phi}$ in the vertical probability amplitude, $\psi_y = \mathrm{e}^{\mathrm{i}\phi}\sin\frac{\theta}{2}$, where $0 \leq \phi < 2\pi$. It is easy to see that for this parametrization of $|\psi\rangle \equiv \begin{pmatrix}\cos\frac{\theta}{2}\\ \mathrm{e}^{\mathrm{i}\phi}\sin\frac{\theta}{2}\end{pmatrix}$, we have $\mathbf{s} = \begin{pmatrix}\sin\theta\cos\phi\\ \sin\theta\sin\phi\\ \cos\theta\end{pmatrix}$, i.e. θ and ϕ are respectively the inclination and the azimuth angles in the spherical coordinate system. Thus all qubit states form the Bloch sphere with unit radius.

It is easy to calculate that rectilinear, diagonal, and circular polarizations are symmetrically located on the Bloch sphere, as shown in Fig. ?. The squared absolute value of the scalar product between state vectors $|\psi\rangle$ and $|\psi'\rangle$ can be represented by the corresponding Bloch vectors $\mathbf{s}$ and $\mathbf{s}'$ as

$$|\langle\psi|\psi'\rangle|^2 = \frac{1}{2}(1 + \mathbf{s}\cdot\mathbf{s}'). \tag{1.22}$$

We leave the calculation as Exercise 1.5.1. Thus orthogonal states are located on their antipodes of the Bloch sphere.

Bloch sphere allows us to visualize various state transformations. A wave plate introducing a phase shift α between the horizontal and the vertical polarizations transforms the state, up to the overall phase factor, as $\begin{pmatrix}\cos\frac{\theta}{2}\\ \mathrm{e}^{\mathrm{i}\phi}\sin\frac{\theta}{2}\end{pmatrix} \mapsto \begin{pmatrix}\cos\frac{\theta}{2}\\ \mathrm{e}^{\mathrm{i}(\phi-\alpha)}\sin\frac{\theta}{2}\end{pmatrix}$ which implies that the azimuthal angle of the Bloch vector is changed by $\phi \mapsto \phi - \alpha$. Therefore the action of a wave plate with principal axes oriented in the rectilinear basis correponds to the rotation of the Bloch sphere by an angle α about the s_3 axis:

$$\mathbf{s} \mapsto \begin{pmatrix}\cos\alpha & -\sin\alpha & 0\\ \sin\alpha & \cos\alpha & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}s_1\\ s_2\\ s_3\end{pmatrix} \tag{1.23}$$

This immediately shows that a quarter-wave plate with $\alpha = \frac{\pi}{2}$ transforms diagonal polarizations onto circular ones.

Let us now consider physical rotation of the coordinate system by an angle β, given by an operator

$$\hat{B}(\beta) \equiv \begin{pmatrix} \cos\frac{\beta}{2} & \sin\frac{\beta}{2} \\ -\sin\frac{\beta}{2} & \cos\frac{\beta}{2} \end{pmatrix} \tag{1.24}$$

which induces a transformation of the state vector $|\psi\rangle \mapsto \hat{B}(\beta)|\psi\rangle$. The components of the Bloch vector for the transformed state will be given by expressions $\langle\psi|\hat{B}^\dagger(\beta)\hat{\sigma}_1\hat{B}(\beta)|\psi\rangle$, $i = 1, 2, 3$. It is easy to verify that

$$\begin{aligned} \hat{B}^\dagger(\beta)\hat{\sigma}_1\hat{B}(\beta) &= \hat{\sigma}_1\cos\beta - \hat{\sigma}_3\sin\beta \\ \hat{B}^\dagger(\beta)\hat{\sigma}_2\hat{B}(\beta) &= \hat{\sigma}_2 \\ \hat{B}^\dagger(\beta)\hat{\sigma}_1\hat{B}(\beta) &= \hat{\sigma}_1\sin\beta + \hat{\sigma}_3\cos\beta \end{aligned} \tag{1.25}$$

Therefore the transformed Bloch vector can be written as:

$$\mathbf{s} \mapsto \begin{pmatrix} \langle\psi|\hat{B}^\dagger(\beta)\hat{\sigma}_1\hat{B}(\beta)|\psi\rangle \\ \langle\psi|\hat{B}^\dagger(\beta)\hat{\sigma}_2\hat{B}(\beta)|\psi\rangle \\ \langle\psi|\hat{B}^\dagger(\beta)\hat{\sigma}_3\hat{B}(\beta)|\psi\rangle \end{pmatrix} = \begin{pmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} \tag{1.26}$$

i.e. it is a rotation about s_2 axis by the angle 2β. Makes sense: circular polarizations are invariant, $\beta = 45°$ maps rectilinear onto circular.

We have proven a stronger fact: any unitary transformation corresponds to a rotation of the Bloch sphere. This is because any unitary can be represented as composition of three transformations of the form considered above (see Exercise 1.5.2

1.5.1 Verify Eq. 1.22.

1.5.2 Show that any special unitary 2×2 matrix can be written as a product:

$$\begin{pmatrix} \mathrm{e}^{\mathrm{i}\alpha/2} & 0 \\ 0 & \mathrm{e}^{-\mathrm{i}\alpha/2} \end{pmatrix} \begin{pmatrix} \cos\frac{\beta}{2} & \sin\frac{\beta}{2} \\ -\sin\frac{\beta}{2}\beta & \cos\frac{\beta}{2} \end{pmatrix} \begin{pmatrix} \mathrm{e}^{\mathrm{i}\gamma/2} & 0 \\ 0 & \mathrm{e}^{-\mathrm{i}\gamma/2} \end{pmatrix} \tag{1.27}$$

1.5.3 For a real non-zero vector $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$, we can define an operator Pauli operator via a formal scalar product $\mathbf{a} \cdot \hat{\boldsymbol{\sigma}} = a_1\hat{\sigma}_1 + a_2\hat{\sigma}_2 + a_3\hat{\sigma}_3$. Show that eigenvalues of this operator are given by $\pm|\mathbf{a}|$ and its eigenvectors correspond to Bloch vectors $\pm\mathbf{a}/|\mathbf{a}|$.

1.5.4 Show that a unitary transformation $\exp(i\alpha\mathbf{n}\cdot\hat{\boldsymbol{\sigma}}/2)$, where $\mathbf{n}$ is a unit real vector, rotates the Bloch sphere by an angle α about the axis defined by $\mathbf{n}$.

Chapter 2

A more mystical face of the qubit

2.1 Beam splitter

The optical field analyzed so far consisted of two orthogonally polarized components traveling along the same path. These two components can be separated with a polarizing beam splitter into distinguishable spatial paths and made to have identical linear polarizations with the help of a halfwave plate. From the fundamental point of view, there is no conceptual difference between the field before and after this transformation. Before we needed two complex numbers to describe the horizontal and the vertical components of the electric field, now we also need two complex numbers to describe the amplitudes of the fields traveling along separate spatial paths. This leads us to the notion of a mode, i.e. a light beam with well defined characteristics, hose only tunable degree of freedom is the complex amplitude.

In the preceding chapter we assumed for simplicity that the modes are plane waves with horizontal and vertical polarizations. In laboratory, we are usually dealing with light beams that have finite both spatial extent and duration. Therefore it is more appropriate to think of modes as wave packets localized in space and in time. An elementary optical device that combines two spatially separate modes is a beam splitter which partly reflects and partly transmits each of the incident beams, see Fig. 2.1.

If a beam with an amplitude $\mathcal{E}_1$ enters through the upper port, a fraction $\mathcal{R}_1\mathcal{E}_1$ will get reflected into the upper output port, and a fraction $\mathcal{T}_1\mathcal{E}_1$ will get

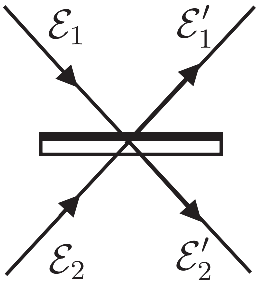


Figure 2.1: Beam-splitter

transmitted. Similarly, a beam with an amplitude $\mathcal{E}_2$ entering through the lower port will be split respectively into $\mathcal{T}_2\mathcal{E}_2$ into the upper output port and $\mathcal{R}_2\mathcal{E}_2$. We will assume that the alignment of the beams satisfies the condition of mode matching, which means that by looking at the characteristics of the outgoing beams in space and time one cannot say which direction they came from. Then with simultaneous illumination of both the input ports the output beams can be described with single amplitudes $\mathcal{E}_1'$ and $\mathcal{E}_2'$ which thanks to the superposition principle will be sums of contributions from the upper and lower input beams. We can describe the input modes entering the beam splitter with a two-element complex vector $\begin{pmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \end{pmatrix}$ which is transformed by the beamsplitter into $\begin{pmatrix} \mathcal{E}_1' \\ \mathcal{E}_2' \end{pmatrix}$. The dependence between the amplitudes of the incoming and outgoing modes is linear and can be written in the matrix form

$$\begin{pmatrix} \mathcal{E}_1' \\ \mathcal{E}_2' \end{pmatrix} = \mathcal{B} \begin{pmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \end{pmatrix}, \quad \mathcal{B} = \begin{pmatrix} \mathcal{R}_1 & \mathcal{T}_2 \\ \mathcal{T}_1 & \mathcal{R}_2 \end{pmatrix} \tag{2.1}$$

Matrix $\mathcal{B}$ is not arbitrary due to the energy conservation constraint. Since the intensity of the light beam is proportional to $|\mathcal{E}|^2$, the energy is conserved iff:

$$|\mathcal{E}_1'|^2 + |\mathcal{E}_2'|^2 = |\mathcal{E}_1|^2 + |\mathcal{E}_2|^2. \tag{2.2}$$

This equality should be satisfied for arbitrary input fields $\mathcal{E}_1$, $\mathcal{E}_2$ which leads to the following constraints on the entries of the $\mathcal{B}$ matrix:

$$|\mathcal{R}_1|^2 + |\mathcal{T}_1|^2 = |\mathcal{R}_2|^2 + |\mathcal{T}|_2^2 = 1 \tag{2.3}$$

$$\mathcal{R}_1\mathcal{T}_2^* + \mathcal{T}_1\mathcal{R}_2^* = \mathcal{R}_1^*\mathcal{T}_2 + \mathcal{T}_1^{*}\mathcal{R}_2 = 0 \tag{2.4}$$

Note that these conditions imply $|\mathcal{R}_1| = |\mathcal{R}_2|$, $|\mathcal{T}_1| = |\mathcal{T}_2|$, and hence one can introduce single power transmission and reflection coefficients $R = |\mathcal{R}_i|^2$, $T = |\mathcal{T}_i|^2$ (exercise 2.1.2). Constraints (2.3-2.4) are equivalent to the condition that $\mathcal{B}$ is a *unitary* matrix $\mathcal{B}^\dagger\mathcal{B} = \mathbb{1}$.

In what follows we will adopt a notation in which

$$\mathcal{B}(\theta) = \begin{pmatrix} \cos\theta/2 & \sin\theta/2 \\ -\sin\theta/2 & \cos\theta/2 \end{pmatrix} \tag{2.5}$$

is a *standard* beam splitter with power transmission $T = \sin^2\theta/2$. It is easy to convince oneself that the minus sign in the above definition is necessary to ensure unitarity of $\mathcal{B}(\theta)$. In particular the balanced beam-splitter with $T = R = 50\%$ corresponds to $\mathcal{B}(\pi/2)$.

2.1.1 Prove that energy conservation constraint leads to Eqs. (2.3-2.4).

2.1.2 Prove that Eqs. (2.3-2.4) imply $|\mathcal{R}_1| = |\mathcal{R}_2|$, $|\mathcal{T}_1| = |\mathcal{T}_2|$.

2.1.3 Write down the most general matrix $\mathcal{B}$ corresponding to a beam splitter with $T = R = 50\%$.

2.2 Mach-Zehnder interferometer

Consider now a Mach-Zehnder interferometer composed of two balanced beam-splitters and a relative phase delay φ between the two arms (see Fig. 2.2). The two components of the complex vector $\begin{pmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \end{pmatrix}$ correspond now to the amplitudes of the field in the upper and lower path of the interferometer. Each of the balanced beam splitters corresponds to the $\mathcal{B}(\pi/2) = \dfrac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ matrix, while the relative phase delay is represented by $F_1(\varphi) = \begin{pmatrix} e^{\mathrm{i}\varphi} & 0 \\ 0 & 1 \end{pmatrix}$. The overall action of the Mach-Zehnder interferometer is the result of the multiplication of the three matrices:

$$\mathrm{MZ}(\varphi) = B(\pi/2)F_1(\varphi)B(\varphi) = e^{\mathrm{i}\varphi}\begin{pmatrix} \mathrm{i}\sin\varphi/2 & \cos\varphi/2 \\ -\cos\varphi/2 & -\mathrm{i}\sin\varphi/2 \end{pmatrix} \tag{2.6}$$

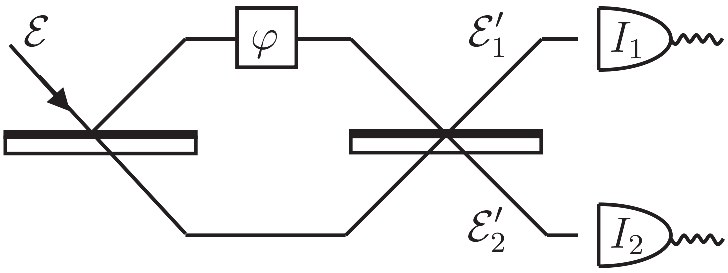


Figure 2.2: Mach-Zehnder interferometer

In particular, if a beam with amplitude $\mathcal{E}$ is send into the upper input arm the output amplitudes read

$$\begin{pmatrix} \mathcal{E}_1' \\ \mathcal{E}_2' \end{pmatrix} = \mathrm{MZ}(\varphi) \begin{pmatrix} \mathcal{E} \\ 0 \end{pmatrix} = \mathcal{E} e^{\mathrm{i}\varphi} \begin{pmatrix} \mathrm{i}\sin\varphi/2 \\ -\cos\varphi/2 \end{pmatrix}. \tag{2.7}$$

Light intensities I_i registered by the detectors placed at the outputs are proportional to $|\mathcal{E}_i|^2$, hence $I_1 = I\sin^2\varphi/2$, $I_2 = I\cos^2\varphi/2$, where I is the input beam intensity.

The above discussion implies that the Mach-Zehnder interferometer can be regarded as a beam-splitter with power transmission $T = \cos^2\varphi/2$. This makes the Mach-Zehnder setup a perfect design for making a tunable beam-splitter and in particular a fast optical switch as the phase delay may be changed quickly with the help of an electro-optic modulator.

The $e^{i\varphi}$ factor standing in front of the matrix in (2.6) applies identical phase shift to both of the amplitudes. If we restrict ourselves to the two-mode scenario this phase shift is never observed. All that we measure are intensities, and in an interference experiment the common phase shift will not yield any intensity change at the output ports. Therefore we may drop this term as unphysical. As a consequence we should have in mind the following rule: all the transformations or states that differ only by a phase factor are physically equivalent. This is again similar to what we have learned from the polarization properties of classical light: the polarization state of light does not change if the Jones vector is multiplied by $e^{\mathrm{i}\varphi}$.

2.2.1 Construct a $B(\theta)$ beam splitter, having only balanced $B(\pi/2)$ beam splitters and possibility of introducing arbitrary relative phase delays. Make sure that all the phase factors match.

2.3 Single photon interference

In the previous section we have analyzed propagation of the classical light through the Mach-Zehnder interferometer. The essential feature in the whole process was interference of the light beams. All that was said translates directly to the quantum description of a single photon. The state of a single photon is described is described by a pair of probability amplitudes corresponding to the photon traveling along the upper or lower arm. Analogously as in the case of polarization we may regard this as an implementation of a two-level quantum system—a qubit. To contrast it with the earlier *polarization qubit* we will refer to this implementation as a *dual-rail qubit* as the photon is in a superposition of two distinguishable spacial modes. A general state of the photon is a superposition:

$$|\psi\rangle = \psi_1|1\rangle + \psi_2|2\rangle, \tag{2.8}$$

where $|1\rangle$, $|2\rangle$ represent the photon traveling in the upper, lower arm respectively. and the probability of detecting a photon in the given arm is $|\psi_i|^2$. Hence, identically as for the polarization qubit the state of the photon is a normalized two component complex vector $|\psi\rangle = \left(\begin{array}{c}\psi_1\\ \psi_2\end{array}\right)$ and we treat as physically equivalent states differing only by a phase factor $|\psi\rangle \equiv e^{\mathrm{i}\varphi}|\psi\rangle$.

Even though the mathematics is the same, the dual-rail qubit may be conceptually more challenging than the polarization qubit. While it may be relatively easy to accept the fact that the diagonal polarization is a superposition of horizontal and vertical polarizations, it may be a bit harder to imagine a state $|+\rangle = (|1\rangle + |2\rangle)/\sqrt{2}$ which is an equal superposition of the photon traveling in the upper and the lower arm of an interferometer. We have to accept that the notion of superposition is not equivalent to a probabilistic mixture of two different states and that the state $|+\rangle$ corresponds to the situation in which the photon *is simultaneously* in the upper and the lower arm of the interferometer.

Let us go step by step through a process in which a single photon is sent into the upper arm of the Mach-Zehnder interferometer (Fig. 2.2). After the first beam splitter the state of the photon becomes $(|1\rangle + |2\rangle)/\sqrt{2}$, representing the photon traveling simultaneously in both the upper and lower

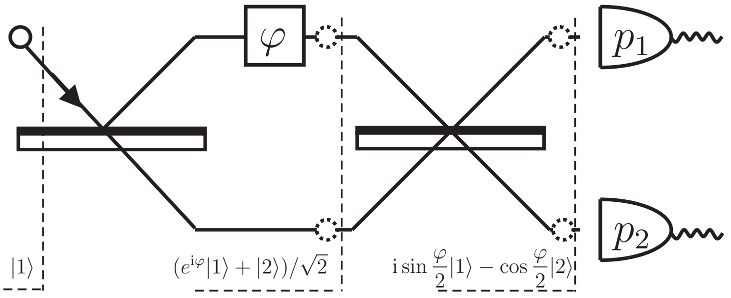

φ
p_1
p_2
$|1\rangle$
$(e^{i\varphi}|1\rangle + |2\rangle)/\sqrt{2}$
$i\sin\frac{\varphi}{2}|1\rangle - \cos\frac{\varphi}{2}|2\rangle$

Figure 2.3: Propagation of a single photon through a Mach-Zehnder interferometer

arm. If at this point the photon was measured the probability of measuring the photon in each of the arms would be 1/2. If, however, the photon is let to travel further through the interferometer it will experience the relative phase delay, $(e^{i\varphi}|1\rangle + |2\rangle)/\sqrt{2}$, and finally the *two paths* will interfere at the final beam splitter yielding $|\psi'\rangle = i\sin(\varphi/2)|1\rangle - \cos(\varphi/2)|2\rangle$. The detectors placed at the output ports will measure the photon with respective probabilities: $p_1 = \sin^2\varphi/2$, $p_2 = \cos^2\varphi/2$.

Consider for the moment the $\varphi = 0$ case. In this case the photon will certainly go to the lower arm as $p_2 = 1$. This is an example where we most clearly see that we need to accept the fact that the superposition is something totally different than the probabilistic mixture. In particular, if someone insisted that the state $(|1\rangle + |2\rangle)/\sqrt{2}$ inside the interferometer corresponds simply to a photon traveling the upper arm with probability 1/2 or the photon traveling the lower arm with probability 1/2, he would no be able to explain this observation that for $\varphi = 0$ the only possible event is the clicking of the lower detector. This is because if photon indeed traveled one particular arm, but it would be merely for our ignorance that we do know which one, once it hit the final beam splitter it would have 50% chance to go to either of the output ports. Without invoking the interference effect ,for which the simultaneous propagation of the photon in both arms is necessary, we are not able to explain the clicks at the output port of the Mach-Zehnder interferometer.

2.4 Polarization vs dual-rail qubit

Since we have a mathematical *isomorphism* between the polarization and the dual-rail qubit we may translate all the states, operations and measurements from one implementation to another.

If we identify $|\leftrightarrow\rangle$, $|\updownarrow\rangle$ polarization states with $|1\rangle$, $|2\rangle$ dual-rail states, we see that the action of the balanced beam splitter that transforms, $|1\rangle \rightarrow (|1\rangle - |2\rangle)/\sqrt{2}$, $|2\rangle \rightarrow |+\rangle = (|1\rangle + |2\rangle)/\sqrt{2}$ is analogous to the placing a half-wave plate that transform $|\leftrightarrow\rangle$, $|\updownarrow\rangle$ into diagonal, anti-diagonal polarizations $|⤡\rangle, |⤢\rangle$. Along the same lines, one can convince oneself that placing detectors directly in the upper and lower arms correspond to measuring photon polarization in $|\leftrightarrow\rangle$, $|\updownarrow\rangle$ basis, while placing them *after* the balanced beam splitter correspond to measuring the photon polarization in $|⤢\rangle$, $|⤡\rangle$ basis.

2.4.1 What measurement setup in the dual-rail implementation corresponds to measuring polarization in the circular polarization basis.

2.4.2 Design a dual-rail setup which corresponds in polarization implementation to to a φ-waveplate rotated by an angle θ.

2.4.3 Design a polarization analogue of the Mach-Zehnder interferometer

---

2.5 Surprising applications

In order to feel how unintuitive quantum mechanics can be, consider the following two thought experiments.

2.5.1 Quantum bomb detection

Imagine a bomb that explodes at the tiniest possible interaction i.e. even when it is hit by a single photon. Your goal is to the detect the presence of a bomb in certain place without making it explode. Clearly, this is an impossible task when you approach it classically. You need to interact with the object in some way and this causes the bomb to explode.

Let us tackle this problem from the quantum perspective. We build a Mach-Zehnder interferometer with one arm passing through a place were the suspected bomb might had been placed. We built the interferometer in a way that the relative phase delay $\varphi = 0$. As discussed earlier, in ideal situation this makes the detection of the photon at the lower output port certain and in the upper arm impossible. Consider know what happens if the bomb is present in the upper arm. Since the bomb explodes once it is hit by a photon this situation is equivalent performing a measurement an a photon asking "which path has the photon traveled". If the bomb is present and explodes this clearly implies the photon traveled the upper arm. If the bomb is present and the *bomb has not exploded* this implies that the photon has traveled the lower arm. Is there a chance to detect a bomb without making it explode. Yes! If the bomb was not there we would only have the lower detector click. If on the other hand we measure a click in the upper arm, we know that there was something measuring the photon and destroying the interference effect. If the bomb was present, then if we are lucky the photon had 50% chance to go the lower arm, and additionally another 50% chance to go to the upper detector. This makes a 25% chance that we detect a bomb without making it explode. Maybe it is not much but still it is much more than we could do classically. Actually a more sophisticated scheme could boost the success probability arbitrarily close to 100%.

2.5.2 Shaping the history of the universe billions years back

Consider a Mach-Zehnder interference setup in a cosmological scale. A star emits a photon in superposition of paths separate by some small angle. The two paths go on two opposite sites of a massive body (black hole?) that curves them so that they finally both hit the earth. An observer on earth may place two telescopes each facing one of the direction from which a photon can come. After registering the photon in one od the telescopes, the observer may say: "the photon traveled along the path i".

Instead of measuring from which direction the photon has come, the observer may on the other hand place a balanced beam-splitter in the place where two paths cross and put the telescopes only after the beam-splitter. In this way detection of a photon in one of the output ports tells him nothing on the direction from which the photon has come but merely about the relative phase factor between the terms representing the photon going either path. In a sense, this measurement project the photon state on two basis states $(|1\rangle + |2\rangle)/\sqrt{2}$, $(|1\rangle - |2\rangle)/\sqrt{2}$. Depending on the measurement results the observer can say: "the photon traveled along both paths simultaneously and the relative phase delay was 0 (or π)".

The intriguing thing is the fact that the choice of the measurement: "path" vs "phase" measurements is done billions of years after the photon emission, but the choice of measurement determines how we will think about the photon's past. Whether we will ascribe it a definite path, or we will claim it traveled both ways simultaneously and only determine its relative phase delay. Putting this reasoning to extreme we may claim that by choosing one or the other measurement we are shaping the history of the universe billions years back . . .

Chapter 3

Distinguishability

3.1 Quantum measurement

Let us revisit the polarizer with two detectors monitoring the output ports. The probabilities of a click on the two detectors are given respectively by

$$|\psi_x|^2 = \left|\langle\leftrightarrow|\psi\rangle\right|^2 = \langle\psi|\leftrightarrow\rangle\langle\leftrightarrow|\psi\rangle \tag{3.1}$$

$$|\psi_y|^2 = \left|\langle\updownarrow|\psi\rangle\right|^2 = \langle\psi|\updownarrow\rangle\langle\updownarrow|\psi\rangle \tag{3.2}$$

The expression $|\leftrightarrow\rangle\langle\leftrightarrow|$ can be viewed as a linear operator acting on the state vectors. We will denote it as $\hat{P}_{\leftrightarrow} = |\leftrightarrow\rangle\langle\leftrightarrow|$. The result of its action on an arbitrary state $|\psi\rangle$ is the state $|\leftrightarrow\rangle$ multiplied by $\langle\leftrightarrow|\psi\rangle$. Alternatively, for the state $|\psi\rangle$ written as a column vector we can represent $\hat{P}_{\leftrightarrow}$ as a 2×2 matrix

$$\hat{P}_{\leftrightarrow} = |\leftrightarrow\rangle\langle\leftrightarrow| \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix}(1,0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{3.3}$$

Analogously, we will denote

$$\hat{P}_{\updownarrow} = |\updownarrow\rangle\langle\updownarrow| \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix}(0,1) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.4}$$

The probabilities of detection events can be consequently written as:

$$|\psi_x|^2 = \langle\psi|\hat{P}_{\leftrightarrow}|\psi\rangle, \qquad |\psi_y|^2 = \langle\psi|\hat{P}_{\updownarrow}|\psi\rangle.$$

An expression that consists of a linear operator $\hat{A}$ sandwiched between a bra and a ket corresponding to the same state vector, which yields a single (in general complex) number $\langle\psi|\hat{A}|\psi\rangle$, appears very often in quantum mechanics and is called an *expectation value.*

Using Eqs. (**??**) and (1.11), the representation of a ket $|\psi\rangle$ as a superposition of $|{\leftrightarrow}\rangle$ and $|{\updownarrow}\rangle$ can be written as:

$$|\psi\rangle = \psi_x|{\leftrightarrow}\rangle + \psi_y|{\updownarrow}\rangle = |{\leftrightarrow}\rangle\langle{\leftrightarrow}|\psi\rangle + |{\updownarrow}\rangle\langle{\updownarrow}|\psi\rangle = \big(\hat{P}_{\leftrightarrow} + \hat{P}_{\updownarrow}\big)|\psi\rangle$$

This is valid for any state $|\psi\rangle$, which means that the sum of two operator appearing in the parentheses must be equal to the identity operator:

$$|{\leftrightarrow}\rangle\langle{\leftrightarrow}| + |{\updownarrow}\rangle\langle{\updownarrow}| = \hat{\mathbb{1}}. \tag{3.5}$$

This property is also obvious in the matrix representation. Physically, this means that the probabilities of all the possible outcomes add up to one. Generally, an operator $\hat{P}$ that satisfies $\hat{P}^2 = \hat{P}$ is called a *projector.* Any operator of the form $|\chi\rangle\langle\chi|$ where $|\chi\rangle$ is a normalized state vector is a projector, in particular both $\hat{P}_{\leftrightarrow}$ and $\hat{P}_{\updownarrow}$ are of this form. A measurement for which probabilities of all possible outcomes are represented by expectation values of projectors is called a *projective measurement.*

It is easy to devise an example of a measurement that is not projective. Consider a plate oriented at the Brewster angle. The entire horizontal component gets transmitted, while the vertical component is split. Let us denote by $\mathscr{T}$ the fraction of the incident vertical amplitude that gets through and by $\mathscr{R}$ the reflected fraction. For a lossless element, we will have $|\mathscr{T}|^2 + |\mathscr{R}|^2 = 1$. The probabilities of outcomes will be given by $p_1 = |\psi_x|^2 + |\mathscr{T}\psi_y|^2$ and $p_2 = |\mathscr{R}\psi_y|^2$. These can be written as expectation values $p_1 = \langle\psi|\hat{M}_1|\psi\rangle$ and $p_2 = \langle\psi|\hat{M}_2|\psi\rangle$, where:

$$\hat{M}_1 = \begin{pmatrix} 1 & 0 \\ 0 & |\mathscr{T}|^2 \end{pmatrix}, \qquad \hat{M}_2 = \begin{pmatrix} 0 & 0 \\ 0 & |\mathscr{R}|^2 \end{pmatrix} \tag{3.6}$$

Easy to check that these operators are not projectors, but they add up to one.

How to describe the most general measurement on a photon allowed by quantum mechanics? Consider a measuring device which fed with a photon yields one of outcomes labelled with an index r. All that quantum theory can predict is the probability that a photon prepared in a state $|\psi\rangle$ will produce a specific outcome. We will postulate that each outcome r is associated with a certain linear operator $\hat{M}_r$ and that the probability of obtaining that outcome is given by the expectation value $\langle\psi|\hat{M}_r|\psi\rangle$. This set of operators provides full quantum mechanical description our measuring apparatus. What conditions must this set satisfy? First, for any state $|\psi\rangle$ the expectation value $\langle\psi|\hat{M}_r|\psi\rangle$ needs to be greater or equal to zero, otherwise it could not be interpreted as a probability. This means that each $\hat{M}_r$ has to be positive. Secondly, the sum of all probabilities must be equal to one, which we can write as: $\sum_r \langle\psi|\hat{M}_r|\psi\rangle = 1 = \langle\psi|\hat{\mathbb{1}}|\psi\rangle$. As this equation is valid for any $|\psi\rangle$, we have:

$$\sum_r \hat{M}_r = \hat{\mathbb{1}}. \tag{3.7}$$

This is a generalization of Eq. (3.5). We will call a set of positive definite operators $\hat{M}_r$ that satisfy Eq. (3.7) simply a *measurement*.

Two useful facts about expectation values. Suppose that for an operator $\hat{A}$ we can find an orthonormal basis $|a_1\rangle, |a_2\rangle$ composed of eigenstates, i.e. $\hat{A}|a_j\rangle = \alpha_j|a_j\rangle$, where α_j are corresponding eigenvalues. The operator $\hat{A}$ can be written as a sum of projectors onto the eigenstates multiplied by respective eigenvalues:

$$\hat{A} = \sum_j \alpha_j |a_j\rangle\langle a_j|. \tag{3.8}$$

The expectation value can be written as:

$$\langle\psi|\hat{A}|\psi\rangle = \sum_j \alpha_j \big|\langle a_j|\psi\rangle\big|^2.$$

Thus the expectation value is a weighted sum of eigenvalues with weights $\big|\langle u_j|\psi\rangle\big|^2$ that add up to one. If all the eigenvalues are real, the expectation value for any state lies always between the minimum and the maximum eigenvalues. Obviously, eigenvalues of operators that form a measurement must be real and bounded between zero and one.

An expectation value $\langle\psi|\hat{A}|\psi\rangle$ can be written in an alternative form which we will use frequently in the future:

$$\langle\psi|\hat{A}|\psi\rangle = \langle\psi|\hat{A}\hat{\mathbb{1}}|\psi\rangle = \sum_{j=\leftrightarrow,\updownarrow} \langle\psi|\hat{A}|j\rangle\langle j|\psi\rangle = \sum_{j=\leftrightarrow,\updownarrow} \langle j|\psi\rangle\langle\psi|\hat{A}|j\rangle \tag{3.9}$$

where in the second step we used Eq. (3.5). The last expression under the sum can be interpreted as a diagonal element of a matrix representing the product of a projector $|\psi\rangle\langle\psi|$ and an operator $\hat{A}$. Thus sum of all diagonal matrix elements of an operator $\hat{A}$ is the trace of the product:

$$\langle\psi|\hat{A}|\psi\rangle = \mathrm{Tr}\big(|\psi\rangle\langle\psi|\hat{A}\big). \tag{3.10}$$

Trace operation is linear with respect to its argument.

3.1.1 Show that the eigenvalues of a projector must be either 0 or 1.

3.1.2 Verify that $\mathrm{Tr}(\hat{A}\hat{B}) = \mathrm{Tr}(\hat{B}\hat{A})$.

3.1.3 Show that if the expectation value of an operator on any state vector is real, then the operator is hermitian.

3.1.4 Show that if $\mathbf{s}$ is the Bloch vector corresponding to a state $|\psi\rangle$, then

$$|\psi\rangle\langle\psi| = \frac{1}{2}(\hat{\mathbb{1}} + \mathbf{s}\cdot\hat{\boldsymbol{\sigma}}), \tag{3.11}$$

where $\mathbf{s}\cdot\hat{\boldsymbol{\sigma}} = s_1\hat{\sigma}_1 + s_2\hat{\sigma}_2 + s_3\hat{\sigma}_3$.

3.2 Minimum-error discrimination

Suppose that we are given a qubit prepared in one of two states $|\psi\rangle$ or $|\chi\rangle$ and our task is to find out which one of these two states it is. If the two states correspond to orthogonal linear polarizations, then a properly oriented polarizer will do the job. If general two mutually orthogonal states, then a quarter wave plate and a suitably oriented polarizer as discussed in Exercise 1.3.1.

What if the two states are not orthogonal? Let us start from a simple example:

$$|\psi\rangle = \begin{pmatrix} \sin\frac{\theta}{2} \\ \cos\frac{\theta}{2} \end{pmatrix}, \quad |\chi\rangle = \begin{pmatrix} -\sin\frac{\theta}{2} \\ \cos\frac{\theta}{2} \end{pmatrix} \tag{3.12}$$

The scalar product is $\langle\chi|\psi\rangle = \cos\theta$. Choosing θ from the range $0, \pi/2$ gives us the full range of the absolute value of the scalar product between two normalized state vectors.

It will be helpful to think in terms of money. Suppose that both the states are equiprobable. We gain €1 if we guess correctly, if we are wrong we need to pay €1. What is the average pay-off if we play the game many times. Of course each time we receive a new qubit prepared freshly in randomly chosen one of two states. Suppose that that we receive €1 for correct identification, while in the case of a mistake we need to pay back €1. What is our average pay-off if the game is repeated many times?

The basis to make a guess must be a certain a measurement. Let the measuring apparatus have two possible outcomes 'ψ' and 'χ' meaning respectively 'I think it was state $|\psi\rangle$' and 'I think it was state $|\chi\rangle$'. These results correspond to a pair of positive definite operators $\hat{M}_\psi$ and $\hat{M}_\chi$ that sum up to the identity operator, $\hat{M}_\psi + \hat{M}_\chi = \hat{\mathbb{1}}$. The average pay-off P will be:

$$\mathsf{P} = \frac{1}{2}\langle\psi|\hat{M}_\psi|\psi\rangle - \frac{1}{2}\langle\psi|\hat{M}_\chi|\psi\rangle + \frac{1}{2}\langle\chi|\hat{M}_\chi|\chi\rangle - \frac{1}{2}\langle\chi|\hat{M}_\psi|\chi\rangle.$$

Using Eq. (3.10) this expression can be transformed to:

$$\mathsf{P} = \frac{1}{2}\mathrm{Tr}\big[\big(|\psi\rangle\langle\psi| - |\chi\rangle\langle\chi|\big)\big(\hat{M}_\psi - \hat{M}_\chi\big)\big]$$

Using the relation $\hat{M}_\chi = \hat{\mathbb{1}} - \hat{M}_\psi$ simplifies the expression to:

$$\mathsf{P} = \mathrm{Tr}\big[\big(|\psi\rangle\langle\psi| - |\chi\rangle\langle\chi|\big)\hat{M}_\psi\big] \tag{3.13}$$

where we used the fact that $\mathrm{Tr}\big(|\psi\rangle\langle\psi|\big) = \mathrm{Tr}\big(|\chi\rangle\langle\chi|\big) = 1$. Let us now calculate the matrix representation of the operator:

$$|\psi\rangle\langle\psi| - |\chi\rangle\langle\chi| \equiv \begin{pmatrix} 0 & \sin\theta \\ \sin\theta & 0 \end{pmatrix} \equiv \hat{\sigma}_1 \sin\theta, \tag{3.14}$$

where we denoted by $\hat{\sigma}_1$ the operator:

$$\hat{\sigma}_1 \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \equiv |\swarrow\!\!\!\!\!\nearrow\rangle\langle\swarrow\!\!\!\!\!\nearrow| - |\nwarrow\!\!\!\!\!\searrow\rangle\langle\nwarrow\!\!\!\!\!\searrow| \tag{3.15}$$

where the second form can be viewed as its spectral decomposition. Inserting the final form back into Eq. (3.13) yields:

$$\mathsf{P} = \sin\theta\big(\langle\swarrow\!\!\!\!\!\nearrow|\hat{M}_\psi|\swarrow\!\!\!\!\!\nearrow\rangle - \langle\nwarrow\!\!\!\!\!\searrow|\hat{M}_\psi|\nwarrow\!\!\!\!\!\searrow\rangle\big)$$

It is easy to see that P reaches the maximum allowed value when $\langle\swarrow\!\!\!\!\!\nearrow|\hat{M}_\psi|\swarrow\!\!\!\!\!\nearrow\rangle = 1$ and $\langle\nwarrow\!\!\!\!\!\searrow|\hat{M}_\psi|\nwarrow\!\!\!\!\!\searrow\rangle = 0$. This can be achieved with the projective operator $\hat{M}_\psi = |\swarrow\!\!\!\!\!\nearrow\rangle\langle\swarrow\!\!\!\!\!\nearrow|$, which implies that $\hat{M}_\chi = |\nwarrow\!\!\!\!\!\searrow\rangle\langle\nwarrow\!\!\!\!\!\searrow|$.

The maximum can be easily attained, for our specific example just take let us just take $\hat{M}_\psi = |\swarrow\!\!\!\nearrow\rangle\langle\swarrow\!\!\!\nearrow|$, which means that $\hat{M}_\chi = |\nwarrow\!\!\!\searrow\rangle\langle\nwarrow\!\!\!\searrow|$.

Thus the maximum pay-off equals to:

$$\mathsf{P} = \sin\theta = \sqrt{1-\cos^2\theta} = \sqrt{1-\big|\langle\chi|\psi\rangle\big|^2}.$$

The second expression of the pay-off in terms of the absolute value of the scalar product $\langle\chi|\psi\rangle$, holds for an arbitrary pair of states, whose scalar product is not necessarily real. The probability of correct identification equals to:

$$\langle\psi|\hat{M}_\psi|\psi\rangle = \langle\chi|\hat{M}_\chi|\chi\rangle = \frac{1}{2}\left(1+\sqrt{1-\big|\langle\chi|\psi\rangle\big|^2}\right).$$

of wrong indentification:

$$\langle\chi|\hat{M}_\psi|\chi\rangle = \langle\psi|\hat{M}_\chi|\psi\rangle = \frac{1}{2}\left(1-\sqrt{1-\big|\langle\chi|\psi\rangle\big|^2}\right).$$

3.2.1 Find a unitary transformation which maps two qubit states $|\psi'\rangle$ and $|\chi'\rangle$ onto the states $|\psi\rangle$ and $|\chi\rangle$ defined in Eq. (3.12) under a constraint $\big|\langle\chi'|psi'\rangle\big| = \cos\theta$.

3.2.2 What is the pay-off when the two states are given with probabilities p_ψ and p_χ?

3.3 Unambiguous discrimination

Let us now consider the task of discriminating between two non-orthogonal states $|\psi\rangle$ and $|\chi\rangle$, but with different rules. This time we do not want to make a mistake. Because we already know that discrimination is not possible in 100% of cases, we will allow the measuring apparatus to have three outcomes: 'ψ': that was for sure $|\psi\rangle$, 'χ': that was for sure $|\chi\rangle$, and '?': the identification failed in this instance, and corresponding operators $\hat{L}_\psi, \hat{L}_\chi$, and $\hat{L}_?$.

Another game: our task is to identify the state, but we do not want to make mistakes. We allow for a possibility that we did not manage. Three outcomes: $\hat{L}_\psi, \hat{L}_\chi, \hat{L}_?$. The condition of no mistakes means that:

$$\langle\chi|\hat{L}_\psi|\chi\rangle = \langle\psi|\hat{L}_\chi|\psi\rangle = 0. \tag{3.16}$$

Let us first consider $\hat{L}_\psi$. We can introduce two orthogonal eigenvectors $|l_1\rangle, |l_2\rangle$ and corresponding eigenvalues λ_1, λ_2. The expectation value can be written as:

$$\langle\chi|\hat{L}_\psi|\chi\rangle = \lambda_1\big|\langle l_1|\chi\rangle\big|^2 + \lambda_2\big|\langle l_2|\chi\rangle\big|^2 \tag{3.17}$$

If an eigenvalue, say λ_1 is nonzero, then the scalar product must vanish, $\langle u_1|\chi\rangle = 0$. This means that $|u_1\rangle$ must be orthogonal to $|\chi\rangle$. We can write it as $|u_1\rangle = |\chi^\perp\rangle$. For a qubit, this state is defined unambiguously up to an overall phase which does not play any role. The second eigenstate is orthogonal to the first one and therefore it is simply $|u_2\rangle = |\chi\rangle$. But this means in turn that $\lambda_2 = 0$. Thus according to Eq. (3.8) we can write $\hat{L}_\psi = \lambda_1|\chi^\perp\rangle\langle\chi^\perp|$. We can carry out a similar reasoning for $\hat{L}_\psi$, which needs to be proportional to the projector $|\psi^\perp\rangle\langle\psi^\perp|$. In principle the nonzero eigenvalue can be different, but symmetry suggests that the eigenvalues are equal, we will denote them just by λ. The general case is a subject of Problem **??**. Thus we assume that $\hat{L}_\psi = \lambda|\chi^\perp\rangle\langle\chi^\perp|$ and $\hat{L}_\chi = \lambda|\psi^\perp\rangle\langle\psi^\perp|$, where explicitly:

$$|\psi^\perp\rangle \equiv \begin{pmatrix} -\cos\frac{\theta}{2} \\ \sin\frac{\theta}{2} \end{pmatrix}, \qquad |\chi^\perp\rangle \equiv \begin{pmatrix} \cos\frac{\theta}{2} \\ \sin\frac{\theta}{2} \end{pmatrix}$$

The probability of correct identification is proportional to λ, therefore it should be as large as possible. What limits this factor? Let us calculate the matrix representation of $\hat{L}_?$, which is now defined uniquely:

$$\hat{L}_? = \hat{\mathbb{1}} - \hat{L}_\psi - \hat{L}_\chi \equiv \begin{pmatrix} 1-\lambda(1+\cos\theta) & 0 \\ 0 & 1-\lambda(1-\cos\theta) \end{pmatrix}$$

Because the matrix is diagonal, it gives us the eigenvalues, equal to $1 - \lambda(1 \pm \cos\theta)$. The condition that the eigenvalues are nonnegative gives $\lambda \leq 1/(1+|\cos\theta|)$, and of course we should choose the maximum value. Thus the probability of correct identification is:

$$\langle\psi|\hat{L}_\psi|\psi\rangle = \lambda\big|\langle\chi^\perp|\psi\rangle\big|^2 = \frac{\sin^2\theta}{1+|\cos\theta|} = 1 - |\cos\theta|.$$

and similarly for $|\chi\rangle$, while the probability of an inconclusive result is

$$\langle\psi|\hat{L}_?|\psi\rangle = \langle\chi|\hat{L}_?|\chi\rangle = |\cos\theta| = \big|\langle\chi|\psi\rangle\big|$$

where the last expression is valid for arbitrary two states.

3.3.1 Consider the general case of $\lambda_\psi \neq \lambda_\chi$.

3.3.2 Generalize the unambiguous measurement to the case when the two states are given with unequal probabilities p_ψ and p_χ, and our aim is to minimize the average probability of the '?' outcome.

3.3.3 Verify that the unambiguous measurement gives a lower pay-off in the game considered in Sec. 3.2.

3.4 Optical realisation

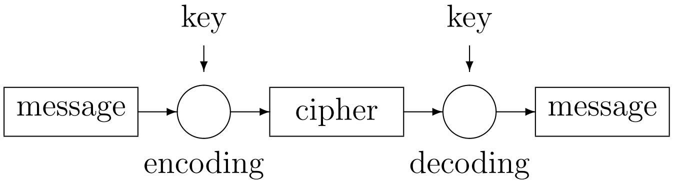
Chapter 4

Quantum cryptography

4.1 Codemakers vs. codebreakers

The science of cryptography is about transmitting a messages in the way that no illegitimate party can learn its meaning. One of the earliest cryptographic method was *Ceasar cipher* in which a letter in a message was replaced by a letter k places further in the alphabet. If we took $k = 3$ then CEASAR would be encoded as FADVDU. Such a code can be broken easily once one knows that the message was encoded using Ceasar cipher. One simply has to check all possible values of k, which is the number of letters in the alphabet – 26, which is not a great amount of work.

The general scheme in cryptography can be depicted as follows:



In case of the Ceasar cipher the message is CEASAR, the cipher is FADVDU and the key is $k = 3$.

A more general cipher is the substitution cipher, where each letter in mapped onto another letter. The Ceasar cipher is an example of substitution cipher. In a general substitution cipher we have 26! possibilities. Hence, such a cipher can not be broken by checking all possible letter substitutions, and thus is more secure than the Ceasar cipher. Nevertheless, it can be broken by letter frequency analysis, since each language has its particular letter

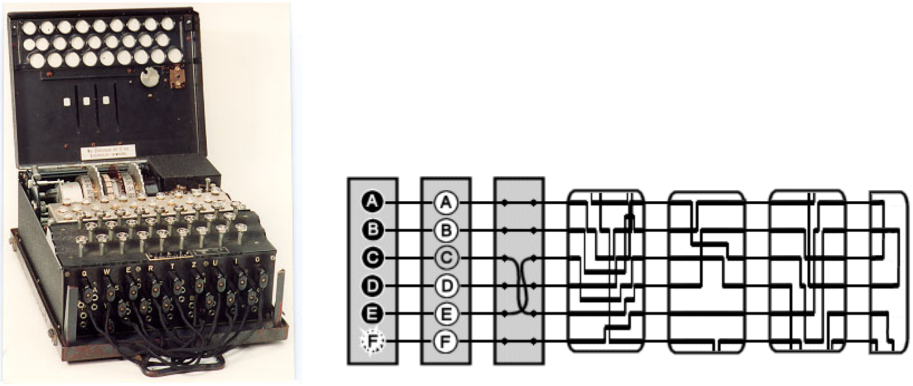


Figure 4.1: ENIGMA machine and the schematic representation of the encryption mechanism. On the scheme from the left: input keyboard, output lamps, plugboard, three scrambling discs, reflector.

frequency pattern, and one can quickly find out which letter was substituted to which one by investigating frequencies in which they appear.

There are a number of ways in which the substitution cipher may be immunized to frequency analysis attacks with ENIGMA being the most famous example. The basic idea is to change continuously the substitution cipher while encoding the consecutive letters. In this way a letter "A" at one place in the message is encoded to e.g "S" while the same letter "A" in in another place may be encoded to e.g. "D". Clearly for this cipher to work there needs to be a *rule* of changing the substitution cipher known both to the sender and the legitimate receiver. The ENIGMA looks like a sophisticated electrical typewriter with the keyboard for entering the letters of the message and the lamps lighting up the corresponding letter of the cipher. At the heart of the ENIGMA laid a scrambling disc which provided a non trivial electrical connection between 26 input and output letters. The scrambling disc could rotate while the message was being entered and take 26 different positions, each of which could be regarded as a realization of a different substitution cipher. To make things even worse for a potential codebreaker the machine was equipped with three such scrambling discs, each successive one taking as the input the output letter of the preceding one. With each letter entered the first disc moved to the next position, and each time it mad a full round trip the second disc moved one step. Analogously, with a complete rotation of the second disc the third disc moved one step. In this way the machine could be in one of 26 × 26 × 26 = 17576 different scrambling discs positions, each corresponding to a different substitution cipher. It should not be surprising that such a system is practically immune to any kind of frequency analysis attack as on average each letter is encoded with equal frequencies to all other letters.

Even though the number of discs positions combination seems large it is not large enough to make ENIGMA secure. Recall that when analyzing security of a given encryption system we assume the codebreaker to *know the encryption system* and the only thing he does not know is the *key*. In the ENIGMA case one assumes that the codebreaker *has* the ENIGMA machine. What he does not have is the key, which in the case of ENIGMA would be the initial setting of the three scrambling discs. Looking from this perspective, one could brake the system by setting the discs to all 17576 combinations and looking for a meaningful output. This look like a hard work but is in principle possible. That is way the ENIGMA was additionally equipped with a plugboard allowing 6 arbitrary chosen pairs of letter to be swapped. Taking into account that the scrambling disc could also be permuted this made together total of $10^1$6 different keys much to much for a brute force attack. The key that needed to be shared by the sender and the receiver consisted now of: discs permutation, discs positions, 6 pairs of letters to be swapped. In short, the role of the plugboard was simply to increase the number of possible keys while the role of the scramblers was to make the system immune to frequency analysis attacks. Was it possible to decode a message without knowing the key? The system seems so complicated that it is hard to imagine how it could have been broken. It had been broken though, and the first one who did it was Marian Reyewski . . .

This should make us realize that almost all ciphers can theoretically be broken, and they strength stems from the practical difficulties of doing so Nevertheless, there is one exception to this rule. There is a cipher which is *proven* to be secure! The cipher is *the one time pad* and it works as follows. Write your message in the binary form, take the key which is a completely random sequence of 0 and 1 of the same length as the message and perform bitwise XOR operations to obtain the cipher

message	0	1	1	0	1	1	0
key	0	0	0	1	0	1	1
cipher = message $\oplus$ key	0	1	1	1	1	0	1

Notice that since the key is completely random so is the cipher. More formally let K be the random variable associated with the key. Let the key have length n. Complete randomness means that all binary sequences are equally probable: $p(K) = 1/2^n$, which implies that the cipher carries no information about the message for someone who does not know the key. In order to decode the message the receiver has to apply exactly the same operation as the sender. Application of the XOR operation to the received cipher and the key bits yields the message.

The one time pad has one drawback which makes it impractical for real life communication: it must be the same length as the transmitted message, otherwise it is no longer secure. The main obstacle is thus the effective distribution of the random key to legitimate parties. Once we know how to distribute a long secret key securely we are done.

There are ways to avoid the issue of distributing the secret key. The ones that are used today are based on a *belief* that some mathematical operations are easy to perform in one direction but are very difficult to reverse. A prominent example is multiplication vs. factoring on which the RSA protocol is based. The problem with RSA and other similar protocols is that they security is based not even on a mathematical proof but on a mathematical belief. Nobody has proven that factoring is indeed hard, i.e. it cannot be done in time which grows polynomially with number of digits of the number to be factored. It is even worse. Somebody has found an algorithm which actually does it, but it requires a quantum computer to be run on (see Chapter ?)

The future does not look bleak for the codemakers though. Leaving aside the security which is based on complexity (or a belief in complexity) of some mathematical problems one can base the security of a secret key distribution on the laws of quantum physics. This promising alternative is the *quantum key distribution* (QKD).

4.2 BB84 quantum key distribution protocol

We have observed in previous chapters that nonorthogonal quantum states cannot be distinguished perfectly. This *inaccessibility* of quantum states which seems only a nuisance at a first glance proves to be the key to the secure information transmission.

Let us describe here the most famous protocol proposed by Bennet and Brassard in 1984 (BB84). Consider two parties A, B, which are connected by a quantum channel allowing for transmission od qubits (e.g. an optical fiber in through which single photons are sent), and a classical public channel (e.g. telephone). We assume that noth channels are insecure and can be subjected to eavesdropping. We only assume that classical channel is authenticated i.e. A and B know that they talk to each other and their classical messages although potentially tapped cannot be altered.

A and B will use photon polarization for qubits transmitted via the quantum channel. A will send to B one of four states: $|\leftrightarrow\rangle, |\updownarrow\rangle, |\swarrow\!\!\!\!\!\nearrow\rangle, |\nwarrow\!\!\!\!\!\searrow\rangle$ randomly with equal probabilities. We will say that the first two states form basis 1, and the last two basis 2. A and B assign logical values to this states as follows:

basis 1		basis 2	
$\|\leftrightarrow\rangle$	$\|\updownarrow\rangle$	$\|\swarrow\!\!\!\!\!\nearrow\rangle$	$\|\nwarrow\!\!\!\!\!\searrow\rangle$
0	1	0	1

(4.1)

B measures the polarization state of an incoming photon randomly in one of two basis. If he measures in the correct basis his results should be perfectly correlated with bits sent, whereas when he measures in the incorrect basis his results will be completely uncorrelated with that of A. After the transmission took place B communicates to A via the classical channel in which basis he performed the measurement in a given run. He does not reveal, however, the actual results obtained. After this communication A i B keep only bits measured in compatible basis (approximately half). We call this a sifting stage. In ideal situation A and B should have perfectly correlated bits.

A	$\leftrightarrow$	$\swarrow\!\!\!\!\!\nearrow$	$\swarrow\!\!\!\!\!\nearrow$	$\nwarrow\!\!\!\!\!\searrow$	$\updownarrow$	$\updownarrow$	$\leftrightarrow$	$\swarrow\!\!\!\!\!\nearrow$
B	$+$	$\times$	$+$	$+$	$+$	$\times$	$\times$	$\times$
compatible?	✓	✓			✓			✓
key	0	0			1			0

Now the quantum features enter the game. How A and B can be sure that they share bits that nobody else know about – i.e. that they have a one time pad. Put in one sentence this can be stated as follows:

> You can not distinguish perfectly between 4 states used in BB84, and moreover you cannot learn anything about their identity without introducing disturbance.

Hence, A and B can make themselves sure that nobody have eavesdropped on their communication, by revealing part of their bits on classical channel (e.g. 100 bits), and checking whether they all agree. If there are no errors they can be sure with high degree of confidence (the higher the more bits they have revealed) that nobody had eavesdropped. If all bits agree, the revealed bits are of course discarded, while the remaining ones are kept and constitute the one time pad.

To get a feeling why lack of errors proves no eavesdropping, note first that the messages exchanged by A and B over the public channel carry no information about the generated key. Therefore Eve needs to gain information about the state of the quantum systems passed from Alice to Bob. However, we have seen in the preceding chapters that the task of discriminating between non orthogonal quantum states is highly nontrivial. Therefore any measurement performed by Eve on photons sent by Alice will bring only partial information about their states. But if Eve wants to conceal her intervention, she needs to resend a photon to Bob in every case. Consequently, the states of some of the photons received by Bob will be modified. This will be seen as errors in the key.

If there are errors in the bits revealed, one cannot exclude the presence of an eavesdropper E. Naively, A and B should abort communication and try again using e.g. different channel. This, however, is impractical. In reality there will always be errors in communication even if there is no eavesdropper. Error may result from noise in the channel, imperfect detectors etc. In what follows we will denote QBER to be the average bit error rate in communication form A to B. The question we now ask is: What is the tolerable QBER below which we can still in some way distill a one time pad that will have no errors and will be secure i.e. no third party will have any information on it. This can be done using classical methods of error correction and privacy amplification provided the eavesdropper information on the key is less than the information of the legitimate receiver B. We will discuss error-correction and privacy amplification ideas later on after introducing proper information theoretic concepts.

For the time being, we will take a simplified approach and investigate the security of the QKD analyzing just the QBER and the bit error rate of E which we denote as e_E. We adopt the following simplified security criterion:

> If under a given attack $e_E \leq QBER$ we claim that the protocol is *not secure*. Otherwise, we will say that the protocol is *secure* with respect to this attack.

Clearly, to prove full security, one needs to analyze the most general attacks obeying the laws of quantum mechanics. This in itself is a highly non-trivial task and we will discuss it in Chapter ???.

4.2.1 Intercept and resend attacks on BB84

At the moment we will restrict ourselves to a simple class of attacks called intercept and resend attacks, which are not optimal, and hence considering only them does not guarantee full security, but are often considered since they are the only realistic attacks under present technology.

In general, in intercept and resend attacks (IRA), E first measures incoming qubit in some basis and after learning result of the measurement prepares a corresponding state which she sends to B. Ideally (for E of course) she would like to learn what state was sent and resend exactly the same state to B in order not to be detected.

In BB84, two basis are used for communication, basis 1:$|\leftrightarrow\rangle, |\updownarrow\rangle$, and basis 2: $|⤢\rangle, |⤡\rangle$. During transmission E does not know which basis she should measure in since this is revealed only after all qubits has been sent. Consider two strategies she may choose:

1. Measurement in a randomly chosen basis – with probability 1/2, E measures either in $|\leftrightarrow\rangle, |\updownarrow\rangle$ or in a $|⤢\rangle, |⤡\rangle$ basis

2. Measurement in an intermediate basis – every time E measures in $|22.5°\rangle, |112.5°\rangle$, which is an basis "in between" two basis used in BB84

Let us calculate what is the E error rate in each of this attacks and what QBER this attacks cause in the A to B communication.

Random basis In half of the cases E will measure in correct basis, hence will learn the state and transmit the state without any disturbance. In the second half, she will measure in the wrong basis. Since $|\langle\leftrightarrow|⤢\rangle^2 = |\langle\leftrightarrow|⤡\rangle^2 = 1/2$ and $|\langle\updownarrow|⤢\rangle^2 = |\langle\updownarrow|⤡\rangle^2 = 1/2$, she will obtain a correct measurement result with probability 1/2. She will resend, a state in the wrong basis, however, and consequently B has 1/2 probability of registering an error in communication even though his basis is set according with that of A. Summarizing B on average will observe qubit error rate (QBER) $QBER = 1/4$. Probability that E will measure an incorrect bit sent by A is $e_E = 1/2 \cdot 1/2 = 1/4$, hence errors will be the same as between A and B. Analyzing the above attack we can conclude that once A and B measure

$QBER \geq 1/4$ they should abort their communication since E in principle could have the same all smaller bit error rate as they have.

Intermediate basis attack Using intermediate basis, probability that E measures a wrong bit value is

$$e_E = q = |\langle 22.5^\circ|\updownarrow\rangle|^2 = 1/4(2-\sqrt{2}) \simeq 0.146. \tag{4.2}$$

Notice that this error is smaller than average error in random basis attack. Such an attack induces again $QBER = 2q(1-q) = 1/4$. We see that the situation is even worse from the point of view of A and B than in the random basis attack. Analyzing random basis attack we have concluded that once QBER $\geq$ 25% the protocol is not secure. Can we now get a tighter estimation of the QBER borderline above which the protocol is insecure?

It is possible if we generalize E attack in a way that E attack only a r fraction of the flying qubits. If this is the case only this fraction will get disturbed so the QBER will be lower. Additionally E will simply have to guess the values of the bits she had not measured. The attack results in:

$$\text{QBER} = r/4, \quad e_E = rq + (1-r)/2 \tag{4.3}$$

Asking when $e_E \leq$ QBER we find that this corresponds to $r \geq 2/(1+\sqrt{2})$ and QBER $\geq 1/[2(1+\sqrt{2})] \approx 20.7\%$. Hence whenever A and B find the QBER $\geq$ 20.7% their protocol is certainly not secure.

4.2.2 General attacks

As mentioned before, proving security requires analysis of the most general attacks allowed by quantum mechanics and identifying bounds on the amount of information obtainable by E for a given level of disturbance introduced in A and B communication. We will discuss this issue in more detail in Chapter ???. For the moment we just mention the final result which states that the protocol *is secure* if $QBER \leq 11\%$. We therefore see that the requirement is realistic and can be met in practical implementation of QKD, which is the reason why quantum key distribution is becoming more and more a commercial product than just an academic topic.

4.3 B92 protocol

A natural question arises: if two nonorthogonal states cannot be perfectly distinguished, then maybe one can construct a QKD protocol using only two states instead of four as used in BB84. Amazingly, this is indeed possible and was realized by Bennett in 1992.

A sends either $|\leftrightarrow\rangle$, or $|\nearrow\!\!\!\!\swarrow\rangle$. B measures either in $|\leftrightarrow\rangle, |\updownarrow\rangle$ or in $|\nearrow\!\!\!\!\swarrow\rangle, |\nwarrow\!\!\!\!\searrow\rangle$ basis. Unlike in BB84 he does not communicate the basis he used, but rather informs A about the cases in which he measured $|\updownarrow\rangle$ or $|\nwarrow\!\!\!\!\searrow\rangle$ (without specifying which of them). This is an information that tells A that in this run B had a basis incompatible with the one she used. Hence if she denotes by 0 and 1 the cases when she sends $|\leftrightarrow\rangle$ and $|\nearrow\!\!\!\!\swarrow\rangle$ respectively, and B denotes by 0 and 1 the cases when he used basis $|\nearrow\!\!\!\!\swarrow\rangle, |\nwarrow\!\!\!\!\searrow\rangle$ and $|\leftrightarrow\rangle, |\updownarrow\rangle$, their bits will be perfectly correlated. Other events when B measured $|\leftrightarrow\rangle$ or $|\nearrow\!\!\!\!\swarrow\rangle$ are discarded.

Notice also that there was no information revealed to an eavesdropper when B informed A about positions at which he measured $|\updownarrow\rangle$ or $|\nwarrow\!\!\!\!\searrow\rangle$. Moreover non-perfect distinguishability of nonorthogonal states forces E to induce errors whenever she wants to learn something and thus makes the protocol secure.

4.3.1 Investigate the security of B92 protocol under two different attacks: (i) E perfroms measures randomly in one of the basis $|\leftrightarrow\rangle, |\updownarrow\rangle$ or $|\nearrow\!\!\!\!\swarrow\rangle, |\nwarrow\!\!\!\!\searrow\rangle$, (ii) E performs minimal-error discrimination measurement (see Chapter ???). In each case try to get the tightest condition on QBER above which the protocol is not secure against a given attack (take into account the possibility that only a fraction of qubits is being attacked).

4.3.2 Six states protocol (6S). BB84 protocol may be naturally generalized to a six-state protocol where we introduce a third basis $|\circlearrowright\rangle, |\circlearrowleft\rangle$. A sends each of six states randomly with equal probabilities, while B measures randomly in one of the three basis. B again announces the basis he measured in and if the basis is not compatible with that of A the data is discarded. Analyze the security of the protocol under (i) random basis attack, (ii) intermediate basis attack — note that an intermediate basis attack should be an attack in which E measures in a fixed basis that has no bias towards any of the three basis used in the protocol. Does such a basis exist? — think of the states used in 6S using the Bloch sphere picture.

4.3.3 In B92 protocol we can in principle use two arbitrary non-orthogonal states, e.g. $|\leftrightarrow\rangle$, $|\alpha\rangle$, where $|\alpha\rangle$ is the linear polarization at an angle α with respect to the horizontal direction. How the usefulness of the protocol will change with the change of α from 0° to 90°. In what situations one should choose α closer to 0° and in which situation α closer to 90°.

4.3.4 SARG04 protocol. A protocol propose in 2004 is a seemingly innocent variation to BB84 protocol, but has some advantages in realistic implementation in QKD. A again sends randomly one of four states used in BB84, and B measures randomly in either horizontal-vertical or diagonal basis. However, instead of revealing the basis at the sifting stage, A announces publicly one of four pairs $\{|\leftrightarrow\rangle, |\nearrow\!\!\!\!\!\swarrow\rangle\}$, $\{|\leftrightarrow\rangle, |\nwarrow\!\!\!\!\!\searrow\rangle\}$, $\{|\updownarrow\rangle, |\nearrow\!\!\!\!\!\swarrow\rangle\}$, $\{|\updownarrow\rangle, |\nwarrow\!\!\!\!\!\searrow\rangle\}$. The announced pair contains a state send by A but it is not revealed which one. The convention is that $|\nearrow\!\!\!\!\!\swarrow\rangle, |\nwarrow\!\!\!\!\!\searrow\rangle$ are assigned logical value 0 while $|PolH\rangle, |PolV\rangle$ logical value 1. To understand how secret key can be generated consider situation in which A sends $|\leftrightarrow\rangle$ and announces the pair $\{|\leftrightarrow\rangle, |\nearrow\!\!\!\!\!\swarrow\rangle\}$, with probability 1/2 B measures in horizontal-vertical basis and he gets $|\leftrightarrow\rangle$. He is not sure, however which state from the announced pair caused this results so he discard it. With probability 1/2 he measures in the diagonal basis in which case half of the times he gets $|\nearrow\!\!\!\!\!\swarrow\rangle$ and half of the time he gets $|\nwarrow\!\!\!\!\!\searrow\rangle$. Only in this last case he is sure that the state send by A was $|\leftrightarrow\rangle$ and he writes down the bit value 0.

What portion of the bits is discarded. Analyze the security of the protocol under random basis attacks. If the pairs of states had been announced before sending the qubit, could E perform a more powerful intercept and resend attack?

Chapter 5

Practical quantum cryptography

Even though the security of a *theoretical quantum cryptographic protocol* is guaranteed by the laws of quantum mechanics it is far from obvious whether a particular implementations of the protocol is secure. For an optical implementation of a standard quantum key distribution (QKD) protocol such as BB84, we need three main components:

1. photon sources
2. optical channels
3. photon detectors

Ideally, we would like to have true single photon sources on demand with high repetition rates (around few GHz), optical channels transmitting photon state faithfully with 100% efficiency and perfect detectors which click if and only if they encounter a photon. None of these ideal elements exists. Moreover, what we have at our disposal with current technology is still far from that.

5.1 Optical components

5.1.1 Photon sources

Although there are single photon sources being intensely developed using different physical hardware: nonlinear crystals, quantum dots, single atoms, ions, etc..., non if this technologies provides an efficient single photon source easy to use for practical purposes. That is why most of QKD realization are based on the use of faint laser pulses. The advantage is that faint laser pulses are easy to prepare as one only needs to attenuate on ordinary laser pulse to the desired level. The drawback is that these are not strictly speaking single photon pulses. Photon number distribution in a laser pulse is governed by the Poissonian statistics, and the probability of n photons present in a pulse reads:

$$p_n = \frac{\bar{n}^n}{n!} e^{-\bar{n}}, \tag{5.1}$$

where $\bar{n}$ is the mean number of photons in a pulse. There will always be a non-zero probability of generating multi-photon pulses. This significantly undermines QKD security since an eavesdropper E can in theory take one photon from a multi-photon pulse, store it, send the remaining photons undisturbed to B and then wait until the measurement basis are announced by A and B. Knowing the proper measurement basis E can measure the photon stored and learn the bit of the key without introducing any disturbance. We should mention at this point that the E cannot apply this strategy to single photon pulses as this would require producing a faithful copy of an unknown quantum state — an operation forbidden by the so called quantum no-cloning theorem, see Section.???.

This clearly shows, that A and B need to make sure that multi-photon events are rare and take them into account in the security analysis. Consider an attenuated laser pulse with $\bar{n} \ll 1$. In this case

$$p_1 \approx \bar{n}, \; p_{n\geq 2} \approx \bar{n}^2/2, \; p_{n\geq 2}/p_1 \approx \bar{n}/2. \tag{5.2}$$

In other words if we want the ratio of multi photon event to single photon events to be low we need to assure that $\bar{n}$ is very small as well, which means that very often there will be no photon in a pulse send by the laser and therefore the efficiency of QKD will drop. We will discuss this issue more quantitatively in Section.???

5.1.2 Optical channels

The main difficulty in transmitting photons from A to B using single photons or faint laser pulses is light attenuation. In standard optical fiber telecommunication this problem is solved by setting up a network of optical amplifiers compensating for the loss of the signal amplitude along the way. As already mentioned this is not possible in the quantum regime due to the no-cloning theorem, since amplification is then equivalent to producing many copies of an unknown quantum state, see Section.???.

The intensity of light that traveled a distance l through a medium with the attenuation coefficient α drops exponentially according to:

$$I(l) = I(0)10^{-\alpha l} \tag{5.3}$$

where α conventionally is written using units of dB/km. $\alpha = 10$ dB/km corresponds to intensity dropping 10 times over the length of 1km, $\alpha = 20$ dB/km corresponds to intensity dropping 100 times, etc.

The two most popular choices of optical channels for QKD are free space communication and optical fibers.

Free space Assuming the weather is good, free space provides relatively low loss transmission with $\alpha < 0.1$dB/km for wavelengths in $780nm - 850nm$ and $1520nm - 1600nm$ windows. The drawback is that one needs to *see* the target and make assure that the beam does not broaden to much due to diffraction and atmospheric fluctuations. Nowadays, one of the hottest topics along these lines is to make the first earth to satellite QKD.

Optical fibers Optical fibers technology is the base of modern telecommunication which is the reason that a lot of high quality optical components are available commercially. It is therefore tempting to make use of the existing telecom infrastructure to perform practical QKD. There are two difficulties on the way to realize a large scale QKD network. One of them is loss, which in commercially available fibers is 0.34 dB/km for 1330nm and 0.2 dB/km for 1550nm. Other wavelengths suffer significantly larger losses and are practically never used for optical fiber communication. The other issue is related with the fact the silica based materials which the fibers are made of are birefringent. More importantly the birefringence is sensitive to environmental temperature fluctuations and fiber stress. That is why the birefringence fluctuates along the fiber and if not controlled or compensated results typically in a completely random light polarization at the output. Encoding information in polarization degrees of freedom of a photon is therefore highly non-trivial and alternative approaches based on time-bin encoding have become more popular, see Section.???.

5.1.3 Single photon detectors

The most popular devices able to detect single photons are the *avalanche photo diodes* (APD). A photon hits a semiconductor and via the photoelectric effect causes an electron to escape. The electron is then subject to external electric field which makes it accelerate and allows is to to knock other electrons free. This process continues causing a macroscopic current flow in the end. As such APDs are not able to distinguish single photon from multi-photon events. The basic parameters of APDs are their: *quantum efficiency*, η , which is the probability of detecting a photon that enters the device; *dark count rate*, δ, is the number of false events per second, i.e. clicks that are not caused by an incoming photon, but rather by thermal fluctuations or background noise; *count rate*; the maximal number of events registered per second. This in particular tells us how fast the detector resets itself to the ground state after a photon detection.

Si APD Silicon based APDs are designed to detect photons in the wavelength range 400nm − 1000nm. Typically, their quantum efficiency $\eta \geq 60\%$, while the dark count rate at −30°C is $\delta \approx 100$ Hz. The maximal count rates are around $15MHz$.

InGaAS APD Silicon APDs are not suitable for detecting telecom wavelength photons ($1330nm$, $1550nm$), hence other devices need to be used for this purpose. Most of them requires much stronger cooling and are usually not as efficient as Si APDs. A typical InGaAs APD has $\eta \approx 10\%$. It does not operate continuously and requires gating. Operating at −100°C with a count rate of 0.1MHz, it suffers the dark count rate at the level of 100 Hz.

5.2 Time-bin phase encoding

Due to birefringence fluctuations, polarization encoding is not the easiest one to implement when performing QKD via optical fibers. Fortunately, we know from Chapter.??? that one can make use of the dual-rail implementation of a qubit. We can rephrase the BB84 protocol in the language of the dual-rail implementation.

Consider the setup depicted in Fig. **??**. We denote by $|0\rangle$, $|1\rangle$ the states corresponding to the photon traveling in the upper, lower arm respectively. *A*

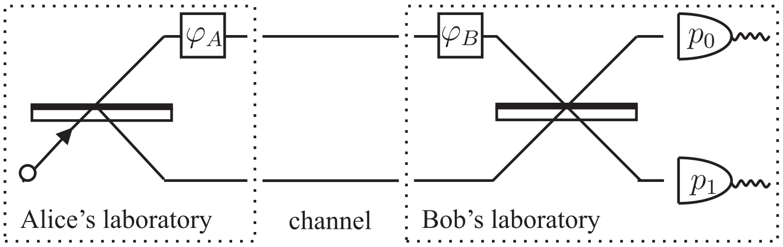


Figure 5.1: The simplest but impractical implementation of the BB84 protocol in the dual-rail implementation.

prepares one of the four states $(e^{i\varphi)_A}|0\rangle + |1\rangle)/\sqrt{2}$ by randomly choosing the phase delay φ_A to be $\{0, \pi, \pi/2, 3/2\pi\}$. As in the polarization BB84 protocol, A chooses between the two basis: $\{0, \pi\}$ or $\{\pi/2, 3/2\pi\}$, where each of the states in the basis encodes 0 and 1 respectively. The four states send by A lie on the equator of the Bloch sphere and in the polarization encoding would correspond to $|\leftrightarrow\rangle, |\updownarrow\rangle, |\circlearrowleft\rangle, |\circlearrowright\rangle$.

By setting φ_B to 0 or $-\pi/2$, B chooses to measure the photon in $\{0, \pi\}$ or $\{\pi/2, 3\pi/2\}$ basis. B ascribes the value of the registered bit according to the measurement results, 0 or 1. Since the probabilities of the detector clicks read

$$p_0 = \sin^2\left(\frac{\varphi_A + \varphi_B}{2}\right), \quad p_1 = \cos^2\left(\frac{\varphi_A + \varphi_B}{2}\right) \tag{5.4}$$

this protocol is equivalent to the BB84.

Nevertheless, this implementation is not practical. It requires keeping the optical length difference between the arms of a many km long interferometer stable up to a fraction of the wavelength—less the micron. This is hardly possible as temperature and stress fluctuations in the fiber will cause the relative phase delay to vary in an uncontrollable manner on a much larger scale. One of the solutions to the problem is to make sure that the two "paths" of the photon experience the same phase fluctuations and thus the relative phase remains well defined.

Consider a modified setup depicted in Fig. **??**. Compared with the previous implementation, now A introduces additionally controlled large delay τ in the upper arm (larger than the length of the pulse). Delay τ introduces a new degree of freedom for the state of the photon, time of arrival. We will adopt the notation where $|i, t\rangle$ describes a photon traveling through the i-th

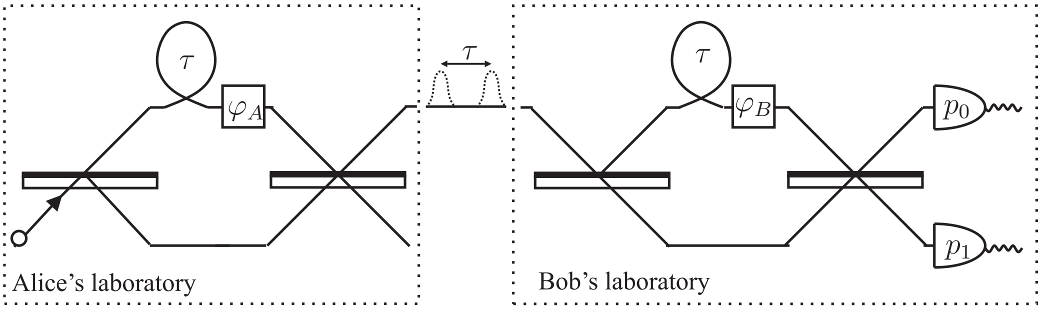


Figure 5.2: A practical implementation of the BB84 protocol using the time-bin phase encoding.

arm of the interferometer in the time slot t. Subsequent state preparation steps at the A laboratory transform the photon state as follows:

$$\begin{aligned}|1,t_0\rangle &\to \frac{1}{\sqrt{2}}\left(|0,t_0\rangle + |1,t_0\rangle\right) \to \frac{1}{\sqrt{2}}\left(e^{\mathrm{i}\varphi_A}|0,t_0+\tau\rangle + |1,t_0\rangle\right) \to \\ &\to \frac{1}{2}\left(e^{\mathrm{i}\varphi_A}|0,t_0+\tau\rangle + |0,t_0\rangle + e^{\mathrm{i}\varphi_A}|1,t_0+\tau\rangle - |1,t_0\rangle\right)\end{aligned} \tag{5.5}$$

Only the upper output path of the A laboratory is connected via the optical channel to B. Two last terms in Eq. (5.5) correspond to cases when the photon goes to the lower output. These events are discarded. Therefore, the preparation process succeeds with probability $1/2$ and the conditional output state is:

$$|\psi\rangle_A = \frac{1}{\sqrt{2}}\left(e^{\mathrm{i}\varphi_A}|0,t_0+\tau\rangle + |0,t_0\rangle\right). \tag{5.6}$$

What we have achieved with this scheme, is the encoding of the BB84 states into the *time-bin* qubit, where the two distinguishable states correspond to the photon traveling in the earlier or in the later time slot.

At the receiving station B performs analogous transformation of the state, with the same large delay loop τ in order to overlap the two time-bin components on the beam-splitter and make the relative phase between the time bins translate to photon detection probabilities. Going through the B setup, the final photon state just before entering the detectors reads:

$$|\psi\rangle_B = \frac{1}{2\sqrt{2}}\left[e^{\mathrm{i}(\varphi)A+\varphi_B)}\left(|0,t_0+2\tau\rangle - |1,t_0+2\tau\rangle\right) - \left(|0,t_0\rangle + |1,t_0\rangle\right)+ \right. \\ \left. +|1,t_0+\tau\rangle\left(e^{\mathrm{i}\varphi_B} - e^{\mathrm{i}\varphi_A}\right) - |2,t_0+\tau\rangle\left(e^{\mathrm{i}\varphi_A} + e^{\mathrm{i}\varphi_B}\right)\right] \tag{5.7}$$

Terms corresponding to the photon arriving at times t_0 or $t_0 + 2\tau$ carry no information on the phase φ_A. Hence, are useless for B and should be discarded. This happens with probability 1/2. The remaining terms cause the detectors to click at time $t + \tau$ with probabilities:

$$p_0 = \sin^2\left(\frac{\varphi_A - \varphi_B}{2}\right), \quad p_1 = \cos^2\left(\frac{\varphi_A - \varphi_B}{2}\right). \tag{5.8}$$

Comparing the above formulas with Eq. (5.4) we see that we retrieve the original BB84 protocol provided we replace $\varphi_B \to -\varphi_B$.

From the implementation point of view this proposal requires only stabilization of interferometers in the LABs which is feasible. Thanks to the fact that the later and earlier pulses travel through the same optical channel at the time difference (ns) much smaller than the characteristic time phase delay fluctuations in the optical channel (s) the relative phase formation is preserved. The drawback of this proposal is the 25% drop in the transmission rate due to the conditional preparation and rejecting the measurement results that happened in t_0 and $t_0 + 2\tau$ time slots.

5.3 Multi-photon pulses and the BB84 security

We analyze here how the QBER threshold for secure key distribution via BB84 is reduced due to the presence of multiphoton pulses at the A side.

Chapter 6

Composite systems

6.1 Two qubits

Suppose that we are dealing with two distinguishable qubits — these can be for example polarizations of two photons travelling in opposite directions — and we would like to characterize their joint state. For convenience, let us denote the qubits with indices A and B. If the photon A is prepared in a polarization state $|\psi\rangle_A = \psi_0|\leftrightarrow\rangle + \psi_1|\updownarrow\rangle$ and the photon B in a state $|\chi\rangle_A = \chi_0|\leftrightarrow\rangle + \chi_1|\updownarrow\rangle$, their combined state is obtained by an operation called by mathematicians the *tensor product* and denoted with a symbol $\otimes$ that is linear with respect to each of its arguments:

$$|\psi\rangle_A \otimes |\chi\rangle_B = \psi_0\chi_0|\leftrightarrow\leftrightarrow\rangle + \psi_0\chi_1|\leftrightarrow\updownarrow\rangle + \psi_1\chi_0|\updownarrow\leftrightarrow\rangle + \psi_1\chi_1|\updownarrow\updownarrow\rangle. \tag{6.1}$$

Here on the right hand side we used a shorthand notation writing $|\leftrightarrow\leftrightarrow\rangle$ instead of $|\leftrightarrow\rangle_A \otimes |\leftrightarrow\rangle_B$, etc. Note that the four states appearing on the right hand side are fully distinguishable: it is sufficient to measure each photon in the rectilinear basis. Therefore we can think of these four states as the orthonormal basis for the composite system of two qubits. In a fixed basis, we can think of the tensor product as an operation producing one four-component vector from a pair of two-component vectors according to the recipe:

$$|\psi\rangle_A \otimes |\chi\rangle_B \equiv \begin{pmatrix}\psi_0\\ \psi_1\end{pmatrix} \otimes \begin{pmatrix}\chi_0\\ \chi_1\end{pmatrix} = \begin{pmatrix}\psi_0 \begin{pmatrix}\chi_0\\ \chi_1\end{pmatrix}\\ \psi_1 \begin{pmatrix}\chi_0\\ \chi_1\end{pmatrix}\end{pmatrix} = \begin{pmatrix}\psi_0\chi_0\\ \psi_0\chi_1\\ \psi_1\chi_0\\ \psi_1\chi_1\end{pmatrix} \tag{6.2}$$

Of course, the order of the vectors in the tensor product matters — the two systems can be physically completely different. This definition has straightforward generalization to arbitrary dimensions of systems A and B, but we will stay with qubits for the sake of simplicity.

The scalar product between two product state vectors is calculated systemwise according to:

$$\big({}_A\langle\psi| \otimes {}_B\langle\chi|\big)\big(|\psi'\rangle_A \otimes |\chi'\rangle_B\big) = \langle\psi|\psi'\rangle\langle\chi|\chi'\rangle \tag{6.3}$$

For example, suppose that two photons are prepared in a state $|\psi\rangle_A \otimes |\chi\rangle_B$ are measured in the bases $|\pm\mathbf{a}\rangle_A$ and $|\pm\mathbf{b}\rangle_A$ respectively. The probability of detecting the first photon in the state $|\mathbf{a}\rangle_A$ and the second one in the state $|\mathbf{b}\rangle_B$ is be given by:

$$\big|\big({}_A\langle\mathbf{a}| \otimes {}_B\langle\mathbf{b}|\big)\big(|\psi\rangle_A \otimes |\chi\rangle_B\big)\big|^2 = \big|\langle\mathbf{a}|\psi\rangle\big|^2\big|\langle\mathbf{b}|\chi\rangle\big|^2 \tag{6.4}$$

and it factorizes into a product of probabilities for individual subsystems. Therefore the measurement outcomes are totally uncorrelated.

What is really exciting about composite systems is that the complete class of all superposition states

$$|\Psi\rangle_{AB} = \sum_{i,j=\leftrightarrow,\updownarrow} \Psi_{ij}|ij\rangle_{AB} \tag{6.5}$$

is much broader than product states of the form given in Eq. (6.1). As an example, let us consider a two-component superposition called the *singlet state*:

$$|\Psi_-\rangle = \frac{1}{\sqrt{2}}\big(|{\leftrightarrow}{\updownarrow}\rangle - |{\updownarrow}{\leftrightarrow}\rangle\big) \tag{6.6}$$

For this state the probability of detecting the first photon in the state $|\mathbf{a}\rangle$ and the second photon in the state $|\mathbf{b}\rangle$ reads:

$$p(\mathbf{a},\mathbf{b}) = \big|\big({}_A\langle\mathbf{a}| \otimes {}_B\langle\mathbf{b}|\big)|\Psi_-\rangle_{AB}\big|^2 = \frac{1}{4}(1 - \mathbf{a}\cdot\mathbf{b}). \tag{6.7}$$

We leave this result as an exercise. Note that $p(\mathbf{a},\mathbf{b}) + p(-\mathbf{a},\mathbf{b}) + p(\mathbf{a},-\mathbf{b}) + p(-\mathbf{a},-\mathbf{b}) = 1$ as it should be. Clearly, the expression in Eq. (6.7) does not factorize. The probability of detecting both the photons in the identical polarizations is zero. What if we are interested in the polarization of only one of the photons, say A? We should sum over possible results for the second photon. This yields $p(\mathbf{a}) = p(\mathbf{a}, \mathbf{b}) + p(\mathbf{a}, -\mathbf{b}) = \frac{1}{2}$. This is reasonable: results of the measurement of the photon A do not depend on the polarizer settings used to measure photon B. Finally, suppose that the first photon was observed in the polarization $\mathbf{a}$. What is the conditional probability that the second photon will be detected in the state $\mathbf{b}$? This can be written formally as:

$$p(\mathbf{b}|\mathbf{a}) = \frac{1}{2}(1 - \mathbf{a} \cdot \mathbf{b}) = \left|\langle \mathbf{b}| - \mathbf{a}\rangle\right|^2, \tag{6.8}$$

i.e. photon B behaves as if it was conditionally prepared in the state $|-\mathbf{a}\rangle$. We will see in the next section that this leads to one of most striking results in quantum mechanics.

6.2 Bell's inequalities

A new game, played by two remote parties, Alice and Bob. At a given instance of time, Alice and Bob are asked one of two questions: X or X' for Alice, Y or Y' for Bob. Questions are equiprobable and uncorrelated between Alice and Bob. Alice and Bob have to answer yes or no, but they do not have time to find out which question the other party was asked. Payoff: +€4 when Alice and Bob give identical answers for combinations of questions XY, XY', $X'Y$ and −€4 for $X'Y'$, and otherwise for opposite answers. What is the maximum average payoff per a single round of the game? An obvious strategy: always say yes. Average payoff is +€2. Can they do better? Not if we remain on the ground of classical physics.

Let us introduce a variable $A = \pm 1$ that specifies the yes/no answer for question X, etc. The payoff averaged over many rounds of the game is given by:

$$\mathsf{W} = \langle AB\rangle + \langle AB'\rangle + \langle A'B\rangle - \langle A'B'\rangle = \langle A(B+B') + A'(B-B')\rangle \tag{6.9}$$

The second form gives as an answer: for every realization, one of the expressions in round parentheses is zero, and the other one is ± 2. Because $|A| = |A'| = 1$, this implies that $|\mathsf{W}| \leq 2$.

Suppose now that Alice and Bob use the following strategy. For every round of the game they perform polarization measurements on a pair of qubits prepared in a singlet state $|\Psi_-\rangle$. Of course, for every round they use a fresh pair. The pay-off for the pair XY will be given by €4×$C(\mathbf{a}, \mathbf{b})$, where

$$C(\mathbf{a}, \mathbf{b}) = p(\mathbf{a}, \mathbf{b}) - p(-\mathbf{a}, \mathbf{b}) - p(\mathbf{a}, -\mathbf{b}) + p(-\mathbf{a}, -\mathbf{b}) = -\mathbf{a} \cdot \mathbf{b}, \tag{6.10}$$

where we explicitly used Eq. (6.7). Thus the pay-off averaged over all four pair of questions, expressed in euros, is given by:

$$\mathsf{W} = C(\mathbf{a},\mathbf{b}) + C(\mathbf{a},\mathbf{b}') + C(\mathbf{a}',\mathbf{b}) - C(\mathbf{a}',\mathbf{b}') \tag{6.11}$$

Let us now take the orientations of the polarizers defined by the following Bloch vectors:

$$X : \mathbf{a} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad X' : \mathbf{a}' = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad Y : \mathbf{b} = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 0 \\ -1 \end{pmatrix}, \quad Y' : \mathbf{b} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}. \tag{6.12}$$

It is easy to calculate that for this choice of measurements $\mathsf{W} = 2\sqrt{2}$.

This results carries a very profound message about the quantum world. In the derivation of the bound $|\mathsf{W}| \leq 2$ being an example of Bell's inequality, we made two assumptions: (i) answers to both the questions are well defined in every round, even though only one of them is asked; (ii) the two parties were not able to find out what question the other party was asked. Applied to polarization measurements, the first means that results of measurements are elements of *objective physical reality*. The second assumption is about *locality*, i.e. impossibility of superluminal communication. Theories that satisfy both these assumptions are said to satisfy *local realism*. Clearly, quantum mechanics does not belong to this class. What does it mean? One interpretion is there is some sort of instantaneous action at arbitrary large distances, i.e. outcome of one measurement depends on the setting of the other apparatus. Note however that we cannot read out this setting. Just looking at the results of our apparatus, we get completely random answers. This "nonlocality" reveals itself only when comparing measurement results at the two locations. The second one is that measurement results were not defined at all before photons entered the respective apparatuses. Operationally, the question about measurement outcomes for measurements $\mathbf{a}$ and $\mathbf{a}'$ is meaningless, as there is no way to perform both of them on the same qubit at once.

This tells us that the statistical character of the quantum theory. We have seen randomness in measurement results. One might suspect that there exist some variables that actually strictly define all the results, but we cannot access them. The correlation function such as $C(\mathbf{a},\mathbf{b})$

$$C(\mathbf{a},\mathbf{b}) = \int \mathrm{d}\lambda\, P(\lambda) A(\mathbf{a};\lambda) B(\mathbf{b};\lambda) \tag{6.13}$$

where $P(\lambda)$ is a proper, positive definite probability distribution The locality is ensured by the fact that A depends only on $\mathbf{a}$ and B only on $\mathbf{b}$. In such a scenario, λ are called *local hidden variables.* Bell's inequalities show that this model cannot explain correlations observed for the singlet state.

One final remark: observing correlations is nothing strange and it does not imply superluminal propagation (shoes from one pair send to different galaxies). What is crucial is that the randomness cannot be modelled by expression 6.13. Sometimes this incompatibility is called quantum nonlocality. This phrase is common in the literature, but it is somewhat misleading, because it does not imply that quantum mechanics is in any way nonlocal.

6.2.1 Show that Bell's inequalities are valid also if local realities satisfy less stringent conditions $-1 \leq A(\mathbf{a};\lambda) \leq 1$ and $-1 \leq B(\mathbf{a};\lambda) \leq 1$. *Hint:* use the identity

$$\langle A'B\rangle - \langle A'B'\rangle = \int \mathrm{d}\lambda P(\lambda)[A'(\lambda)B(\lambda) - A'(\lambda)B'(\lambda)]$$

$$= \int \mathrm{d}\lambda P(\lambda)A'(\lambda)B(\lambda)[1 \pm A(\lambda)B'(\lambda)] - \int \mathrm{d}\lambda P(\lambda)A'(\lambda)B'(\lambda)[1 \pm A(\lambda)B(\Lambda)]$$

where for simplicity we denoted $A(\lambda) = A(\mathbf{a};\lambda)$, $A'(\lambda) = A(\mathbf{a}';\lambda)$ and similarly for Bob's measurements.

6.3 Correlations

Although we have calculated directly the joint probability $p(\mathbf{a},\mathbf{b})$ defined in Eq. Eq:SingletProb, it is instructive to repeat the calculation using operator identities. This is a good opportunity to review properties of the tensor product applied to operators. First, let us write $p(\mathbf{a},\mathbf{b})$ as:

$$p(\mathbf{a},\mathbf{b}) = \langle\Psi_-|\big(|\mathbf{a}\rangle\otimes|\mathbf{b}\rangle\big)\big(\langle\mathbf{a}|\otimes\langle\mathbf{b}|\big)|\Psi_-\rangle \tag{6.14}$$

The projection onto a product vector can be equivalently written as a tensor product of two projectors:

$$\big(|\mathbf{a}\rangle\otimes|\mathbf{b}\rangle\big)\big(\langle\mathbf{a}|\otimes\langle\mathbf{b}|\big) = |\mathbf{a}\rangle\langle\mathbf{a}|\otimes|\mathbf{b}\rangle\langle\mathbf{b}| \tag{6.15}$$

where the tensor product for two general operators $\hat{A}$ and $\hat{B}$ is defined as: Tensor product of operators:

$$(\hat{A} \otimes \hat{B})|\Psi\rangle = \sum_{i,j=0,1} \Psi_{ij} \big(\hat{A}|i\rangle_A\big) \otimes \big(\hat{B}|j\rangle_B\big) \tag{6.16}$$

Of course, we also have $(\hat{A} \otimes \hat{B})(\hat{A}' \otimes \hat{B}') = (\hat{A}\hat{A}') \otimes (\hat{B}\hat{B}')$. If we write operators as matrices, it is easy to see that in the representation used in Eq. (6.2) we have:

$$\hat{A} \otimes \hat{B} \equiv \begin{pmatrix} A_{00}\begin{pmatrix} B_{00} & B_{01} \\ B_{10} & B_{11} \end{pmatrix} & A_{01}\begin{pmatrix} B_{00} & B_{01} \\ B_{10} & B_{11} \end{pmatrix} \\ A_{10}\begin{pmatrix} B_{00} & B_{01} \\ B_{10} & B_{11} \end{pmatrix} & A_{11}\begin{pmatrix} B_{00} & B_{01} \\ B_{10} & B_{11} \end{pmatrix} \end{pmatrix} \tag{6.17}$$

Notationally, this means that each entry in the matrix representing the operator $\hat{A}$ is multiplied by a replica of the matrix corresponding to the opearator $\hat{B}$. Correlation function $C(\mathbf{a}, \mathbf{b})$ calculated for the state $|\Psi_-\rangle$ can be written as:

$$\begin{aligned} C(\mathbf{a}, \mathbf{b}) &= \langle\Psi_-|\big(|\mathbf{a}\rangle\langle\mathbf{a}| - |-\mathbf{a}\rangle\langle-\mathbf{a}|\big) \otimes \big(|\mathbf{b}\rangle\langle\mathbf{b}| - |-\mathbf{b}\rangle\langle-\mathbf{b}|\big)|\Psi\rangle \\ &= \langle\Psi_-|(\mathbf{a}\cdot\hat{\boldsymbol{\sigma}}) \otimes (\mathbf{b}\cdot\hat{\boldsymbol{\sigma}})|\Psi_-\rangle \end{aligned} \tag{6.18}$$

This can be written as a trace over the composite system (which we will denote by the subscript AB):

$$C(\mathbf{a}, \mathbf{b}) = \mathrm{Tr}_{AB}\{[(\mathbf{a}\cdot\hat{\boldsymbol{\sigma}}) \otimes (\mathbf{b}\cdot\hat{\boldsymbol{\sigma}})]|\Psi_-\rangle\langle\Psi_-|\} \tag{6.19}$$

where trace is summation over four terms.

An interesting expression (direct calculation):

$$|\Psi_-\rangle\langle\Psi_-| = \frac{1}{4}(\hat{\mathbb{1}} \otimes \hat{\mathbb{1}} - \hat{\sigma}_1 \otimes \hat{\sigma}_1 - \hat{\sigma}_2 \otimes \hat{\sigma}_2 - \hat{\sigma}_3 \otimes \hat{\sigma}_3) \tag{6.20}$$

The trace of a tensor product over the composite system can be written as: $\mathrm{Tr}_{AB}(\hat{A} \otimes \hat{B}) = (\mathrm{Tr}_A\hat{A})(\mathrm{Tr}_B\hat{B})$.

Now using trace properties of density matrix yields Eq. (6.10).

6.3.1 The Bell's combination can be written as an expectation value of an operator $\hat{W} =$. Find the maximum and the minimum eigenvalue of this operator. *Hint:* calculate $\hat{W}^2$.

6.3.2 The *swap operator* $\hat{S}$ acting on two systems of equal dimension is defined as $\hat{S}(|\psi\rangle \otimes |\chi\rangle) = |\chi\rangle \otimes |\psi\rangle$ for any two state vectors $|\psi\rangle$ and $|\chi\rangle$. Verify that for a two-qubit system $\hat{S} = \hat{\mathbb{1}} \otimes \hat{\mathbb{1}} - 2|\Psi_-\rangle\langle\Psi_-|$.

6.3.3 Show that $\mathrm{Tr}[\hat{S}(\hat{A} \otimes \hat{B})] = \mathrm{Tr}(\hat{A}\hat{B})$.

6.4 Mixed states

Quantum mechanics can predict only probabilities of different measurement outcomes to occur. In order to compare these predictions with experiment, we need to perform measurements on identically prepared systems. So far we assumed that all the systems have been prepared in the same state $|\psi\rangle$. But this does not have to be the case. The preparing party may prepare the system in a state $|\psi_1\rangle$ with probability w_1, etc. We assume that the preparations are random and independent. If we are now interested in the expectation value of a certain operator, say $\hat{A}$, we should average it over possible preparations. Such an average can be written as:

$$\sum_n w_n \langle\psi_n|\hat{A}|\psi_n\rangle = \sum_n w_n \mathrm{Tr}\big(\hat{A}|\psi_n\rangle\langle\psi_n|\big) = \mathrm{Tr}\big(\hat{A}\hat{\varrho}\big) \tag{6.21}$$

where

$$\hat{\varrho} = \sum_n w_n |\psi_n\rangle\langle\psi_n| \tag{6.22}$$

is called the *density operator*. This object contains all the information necessary to calculate expectation values for a statistical ensemble and it is much simpler than specifying all the state vectors and corresponding probabilities. States that are described a state vector, i.e. for which $\hat{\varrho} = |\psi\rangle\langle\psi|$ are called *pure states*. A density operator that cannot be represented in this form of a rank-one projector corresponds to a *mixed state*.

Properties: hermitian, $\mathrm{Tr}\hat{\varrho} = 1$, $\hat{\varrho} \geq 0$. Corresponds to a pure state iff $\hat{\varrho} = \hat{\varrho}^2$.

The set of all density operators can be very nicely visualized for a single qubit. We have seen that all pure states can be represented as a Bloch sphere of unit radius. Consider now a statistical ensemble with states $|\psi_n\rangle$ characterized by Bloch vectors $\mathbf{s}_n$. Using the result of Exercise 3.1.4, the density operator is then given by:

$$\hat{\varrho} = \frac{1}{2}\sum_n w_n(\hat{\mathbb{1}} + \mathbf{s}_n \cdot \hat{\boldsymbol{\sigma}}) = \frac{1}{2}(\hat{\mathbb{1}} + \mathbf{s} \cdot \hat{\boldsymbol{\sigma}}) \tag{6.23}$$

where $\mathbf{s} = \sum_n w_n \mathbf{s}_n$ is a weighted sum of Bloch vectors in the ensemble. Thus the interior of the Bloch sphere corresponds to mixed states. The Bloch sphere for a density operator $\hat{\varrho}$ can be calculated as $\mathbf{s} = \mathrm{Tr}(\hat{\varrho}\hat{\boldsymbol{\sigma}})$.

A completely mixed qubit state:

$$\frac{1}{2}(|{\leftrightarrow}\rangle\langle{\leftrightarrow}| + |{\updownarrow}\rangle\langle{\updownarrow}|) = \frac{1}{2}(|⤢\rangle\langle ⤢| + |⤡\rangle\langle ⤡|) = \frac{1}{2}\hat{\mathbb{1}}. \tag{6.24}$$

Note that we cannot distinguish between these two ensembles.

6.4.1 Show that a qubit density matrix characterized with a Bloch vector $\mathbf{s}$ has eigenvectors given by $|\pm\mathbf{s}/|\mathbf{s}|\rangle$.

6.5 Separability

What is so special about the state $|\Psi_-\rangle$? It is easier to say what states of two subsystems will not be very interesting. First, any product state $|\psi\rangle_A \otimes |\chi\rangle_B$. We will call them product states. More generally, any statistical mixture of product states:

$$\hat{\varrho} = \sum_n w_n |\psi_n\rangle_A\langle\psi_n| \otimes |\chi_n\rangle_B\langle\chi_n|. \tag{6.25}$$

In the case of two qubits, for states of this form the correlation function for joint polarization measurement C can be written as:

$$C(\mathbf{a},\mathbf{b}) = \sum_n w_n \langle\psi_n|\hat{\sigma}_{\mathbf{a}}|\psi_n\rangle\langle\chi_n|\hat{\sigma}_{\mathbf{b}}|\chi_n\rangle \tag{6.26}$$

which satisfies all the assumptions made in the derivation of Bell's inequalities. States of the form (6.25) are called *separable*. States that are not separable are *entangled*. We will see in the next chapters that they are behind effects such as dense coding, teleportation, etc.

How to find out whether a given state $\hat{\varrho}$ is separable or not? A general density operator can be written in a specific basis as:

$$\hat{\varrho} = \sum_{i,j,k,l=0,1} \varrho_{ik,jl} |ik\rangle\langle jl| = \sum_{i,j,k,l=0,1} \varrho_{ik,jl} |i\rangle_A\langle j| \otimes |k\rangle_B\langle l| \tag{6.27}$$

where $\varrho_{ik,jl} = \langle ik|\hat{\varrho}|jl\rangle$. Let us consider partial transposition with the subsystem B in this basis, which changes $|k\rangle_B\langle l|$ onto $|l\rangle_B\langle k|$. This operation will be denoted by $^\Gamma$ (which looks like half of ^{T}). Explicitly,

$$\hat{\varrho}^\Gamma = \sum_{i,j,k,l=0,1} \varrho_{ik,jl} |i\rangle_A\langle j| \otimes |l\rangle_B\langle k| = \sum_{i,j,k,l=0,1} \varrho_{il,jk} |i\rangle_A\langle j| \otimes |k\rangle_B\langle l| \tag{6.28}$$

Obviously, if $\mathrm{Tr}\hat{\varrho} = 1$ then also $\mathrm{Tr}(\hat{\varrho}^\Gamma) = 1$. It is easy to see from Eq. (6.25) that if $\hat{\varrho}$ is separable, then $\hat{\varrho}^\Gamma$ is positive definite, and therefore is also a valid density operator. The reason for that is that if $|\chi_n\rangle_B\langle\chi_n|$ is a single-qubit density matrix, then of course $\left(|\chi_n\rangle_B\langle\chi_n|\right)^T$ also is. Thus we have a necessary condition for separability. It turns out that for two qubits this condition is also sufficient. But the proof is much less straightforward and we will skip it here.

6.5.1 Consider a family of states $\hat{\varrho}_p = p|\Psi_-\rangle\langle\Psi_-| + \frac{1-p}{4}\hat{\mathbb{1}} \otimes \hat{\mathbb{1}}$, where $0 \leq p \leq 1$. Find the ranges of p for which the Bell's inequality is violated and the PPT criterion is not satisfied. Are they identical?

6.5.2 What transformation of the Bloch sphere corresponds to the transposition of a qubit density matrix?

6.5.3 Calculate explicitly the statistical mixture $\int \mathrm{d}\mathbf{a}\, |\mathbf{a}\rangle\langle\mathbf{a}| \otimes |\mathbf{a}\rangle\langle\mathbf{a}|$, where

$$\int \mathrm{d}\mathbf{a} = \frac{1}{4\pi}\int_0^\pi \mathrm{d}\theta \sin\theta \int_0^{2\pi} \mathrm{d}\phi. \tag{6.29}$$

Chapter 7

Entanglement

7.1 Dense coding

Consider four maximally entangled states:

$$|\Psi_\pm\rangle = \frac{1}{\sqrt{2}}\big(|01\rangle \pm |10\rangle\big), \qquad |\Phi_\pm\rangle = \frac{1}{\sqrt{2}}\big(|00\rangle \pm |11\rangle\big) \tag{7.1}$$

Easy to verify that these four states are mutually orthogonal. But we can transform them by operations on just a single qubit:

$$\begin{aligned}(\hat{\sigma}_1 \otimes \hat{\mathbb{1}})|\Phi_+\rangle &= |\Psi_+\rangle \\ (\hat{\sigma}_2 \otimes \hat{\mathbb{1}})|\Phi_+\rangle &= \mathrm{i}|\Psi_-\rangle \\ (\hat{\sigma}_3 \otimes \hat{\mathbb{1}})|\Phi_+\rangle &= |\Phi_-\rangle\end{aligned} \tag{7.2}$$

Protocol: suppose that Alice and Bob share a pair of qubits prepared initially in the state $|\Phi_+\rangle$. Alice applies one of four transformations $\hat{\mathbb{1}}$, $\hat{\sigma}_1$, $\hat{\sigma}_2$, and $\hat{\sigma}_3$ and sends her qubit to Bob. Clearly, this way she can transmit two bits of information. It seems that we have found a way to encode two bits of classical information into one qubit. But we should keep in mind that these two qubits must have been prepared in a joint state that did not have the product form. Therefore qubits must have had a common origin and one qubit must have been exchanged between Alice and Bob. Interestingly, it could have been sent from Bob to Alice, or from Alice to Bob — in both cases, even before Alice has learnt the message she wanted to communicate to Bob. This does not have any counterpart in classical communication.

Dense coding would not be possible with a product state, and by a straightforward extension with separable states. There is something special in non-separable states that enables us to play tricks such as violation of Bell's inequalities or dense coding. This feature can be thought of as a new resource in quantum information processing — namely, *entanglement*. The phenomenon of entanglement can be most conveniently discussed within the paradigm of distant laboratories. Suppose that Alice and Bob have qubits, each one of them prepared separately in well defined state. They can use local operations and classical communication (LOCC) — clearly, all they can produce by these means is a separable state. To share an entangled state, Alice's and Bob's systems must have been prepared jointly and then distributed to the parties. One distributed, entanglement cannot be increased by LOCC.

We have seen that any of the Bell states is equivalent when it comes to applications - the reason for this is that one can transform between them by LOCC (which happens to be just a unitary on one of the subsytems). How to characterize entanglement? Procedure is easy for pure states. Let us think about probability amplitudes as a square matrix. Such a matrix can be subjected to *singular value decomposition*:

$$\Psi_{ij} = \sum_k U^*_{ik} \lambda_k V_{kj} \tag{7.3}$$

where U_{ik} and V_{kj} form unitary matrices and λ_k are real and non-negative. Let us introduce:

$$|u_k\rangle = \sum_i U^*_{ik} |i\rangle = \hat{U}^\dagger |k\rangle, \qquad |v_k\rangle = \sum_j V_{kj} |j\rangle = \hat{V} |k\rangle \tag{7.4}$$

This allows us to write:

$$|\Psi\rangle = \sum_k \lambda_k |u_k\rangle \otimes |v_k\rangle = \left(\hat{U}^\dagger \otimes \hat{V}\right) \left(\sum_k \lambda_k |k\rangle \otimes |k\rangle \right). \tag{7.5}$$

Thus any pure state can be brought by LOCC to the form $\sum_k \lambda_k |k\rangle \otimes |k\rangle$, with nonnegative λ_k. This is called Schmidt decomposition and works in any dimension. For a pair of qubits we have $\lambda_0|00\rangle + \lambda_1|11\rangle$, and because of the normalization constraint $\lambda_0^2 + \lambda_1$ we have only one free parameter. When $\lambda_0 = \lambda_1 = 1/\sqrt{2}$ we've got a maximally entangled state. When either λ_0 or

λ_1 is zero, a product state. What inbetween? Partial entanglement. We will discuss later what can we do with it.

7.1.1 Try to violate Bell's inequalities with a partly entangled two-qubit pure state.

7.2 Remote state preparation

Let us go back to the scalar product which appeared in Eq. (6.7), but we will consider here a general pure bipartite state $|\Psi\rangle_{AB} = \sum_{ij} \Psi_{ij} |i\rangle_A \otimes |j\rangle_B$. It will be convenient to write here

$$|\mathbf{a}\rangle_A = \alpha_0|0\rangle_A + \alpha_1|1\rangle_A, \qquad |\mathbf{b}\rangle_B = \beta_0|0\rangle_B + \beta_1|1\rangle_B$$

We can split the calculation of the scalar product ${}_A\langle\mathbf{a}| \otimes {}_B\langle\mathbf{b}||\Psi\rangle_{AB}$ into two steps. A partial scalar product ${}_A\langle\mathbf{a}|\Psi\rangle_{AB}$ can be thought of as certain state vector for the subsytem B. We will denote it as $|\tilde{\psi}(\mathbf{a})\rangle_B$. Easy to find an explicit expression:

$$|\tilde{\psi}(\mathbf{a})\rangle_B = {}_A\langle\mathbf{a}|\Psi\rangle_{AB} = \sum_{ij} \Psi_{ij}\ {}_A\langle\mathbf{a}|i\rangle_A|j\rangle_B = \sum_j \left(\sum_i \alpha_i^* \Psi_{ij}\right)|j\rangle_B \tag{7.6}$$

The joint probability can now be written as:

$$p(\mathbf{a},\mathbf{b}) = \left|\langle\mathbf{b}|\tilde{\psi}(\mathbf{a})\rangle\right|^2 = \langle\tilde{\psi}(\mathbf{a})|\big(|\mathbf{b}\rangle\langle\mathbf{b}|\big)|\tilde{\psi}(\mathbf{a})\rangle \tag{7.7}$$

The marginal probability of detecting qubit A in the state $|\mathbf{a}\rangle$ is:

$$p(\mathbf{a}) = p(\mathbf{a},\mathbf{b})+p(\mathbf{a},-\mathbf{b}) = \langle\tilde{\psi}(\mathbf{a})|\big(|\mathbf{b}\rangle\langle\mathbf{b}|+|-\mathbf{b}\rangle\langle-\mathbf{b}|\big)|\tilde{\psi}(\mathbf{a})\rangle = \langle\tilde{\psi}(\mathbf{a})|\tilde{\psi}(\mathbf{a})\rangle \tag{7.8}$$

Thus the state $|\tilde{\psi}(\mathbf{a})\rangle_B$ is not normalized, but its squared norm has a well defined meaning: it is the probability that the result **a** has been obtained. This result is valid for an arbitrary measurement applied by Bob, which we leave as Exercise 7.2.2, which makes sense: Alice's probabilities should not depend on Bob's action (if we do not know their result). The conditional probability obtained from Eq. (7.7) reads:

$$p(\mathbf{b}|\mathbf{a}) = \frac{p(\mathbf{a},\mathbf{b})}{p(\mathbf{a})} = \frac{\langle\tilde{\psi}(\mathbf{a})|\big(|\mathbf{b}\rangle\langle\mathbf{b}|\big)|\tilde{\psi}(\mathbf{a})\rangle}{\langle\tilde{\psi}(\mathbf{a})|\tilde{\psi}(\mathbf{a})\rangle} \tag{7.9}$$

Thus if Bob's is interested in the statistics only for the cases for which Alice obtained the result $\mathbf{a}$, he can calculate these taking the state $|\tilde{\psi}(\mathbf{a})\rangle/\sqrt{p(\mathbf{a})}$. In particular for the singlet state:

$$|\tilde{\psi}_{-}(\mathbf{a})\rangle_B = \frac{1}{\sqrt{2}}\big(\alpha_0^*|1\rangle_B - \alpha_1^*|0\rangle_B\big) \tag{7.10}$$

Its squared norm is $1/2$. Interesting: this state is orthogonal to $|\mathbf{a}\rangle$.

Conditional states are a source of misundestanding — they suggest that a measurement at one location instantaneously changes the state of the other particle. But this is an overinterpretation of the mathematical formalism! What matters are results of measurements, and here causality is not violated.

Suppose that two systems have been prepared in an arbitrary joint state $\hat{\varrho}$:

$$\hat{\varrho} = \sum_{ijkl} \varrho_{ik,jl}|i\rangle_A\langle j| \otimes |k\rangle_B\langle l|. \tag{7.11}$$

Alice applies a measurement $\hat{A}_r$ while Bob applies measurement $\hat{B}_s$. The joint probability is given by $p_{r,s} = \mathrm{Tr}(\hat{\varrho}\hat{A}_r \otimes \hat{B}_s)$. If we have access only to Alice's results then the marginal probability $p_r = \sum_s p_{r,s}$ reads:

$$p_r = \mathrm{Tr}_{AB}[\hat{\varrho}(\hat{A}_r \otimes \hat{\mathbb{1}}_B)] = \sum_{ijkl} \varrho_{ik,jl}\mathrm{Tr}_A\big(|i\rangle_A\langle j|\hat{A}_r\big)\mathrm{Tr}_B\big(|k\rangle_B\langle l|\big). \tag{7.12}$$

where we have used $\sum_s \hat{B}_s = \mathbb{1}_B$. Notice that the resulting formula does not depend on the type of measurement performed by Bob. Obviously, $Tr_B\big(|k\rangle_B\langle l|\big) = \delta_{kl}$. This enables us to write $p_r = \mathrm{Tr}_A\big(\hat{\varrho}_A\hat{A}_r\big)$, where

$$\hat{\varrho}_A = \sum_{ijk} \varrho_{ik,jk}|i\rangle_A\langle j| \tag{7.13}$$

is called the reduced density matrix of the subsystem A. Summation over the index k in the above formula is simply the partial trace over the subsystem B, therefore we can also write $\hat{\varrho}_A = \mathrm{Tr}_B\hat{\varrho}$.

Analogously, we will also have $\hat{\varrho}_B = \mathrm{Tr}_A\hat{\varrho}$. Let us consider two qubits prepared in a state $|\Psi\rangle$ and perform the trace of the subsystem A in the basis $|\pm\mathbf{a}\rangle_A$. We see immediately that:

$$\hat{\varrho}_B = |\tilde{\psi}(\mathbf{a})\rangle_B\langle\tilde{\psi}(\mathbf{a})| + |\tilde{\psi}(-\mathbf{a})\rangle_B\langle\tilde{\psi}(-\mathbf{a})| \tag{7.14}$$

Interpretation: if Alice's outcome is unknown, then Bob's subsystem is described by $\hat{\varrho}_B$. If we know the result, then the normalized state is $|\tilde{\psi}(\pm\mathbf{a})\rangle/\sqrt{p(\pm\mathbf{a})}$. Statistical mixture of these states with weights $p(\pm\mathbf{a})$ reproduces Eq. (7.14).

Example: the singlet state Among two-qubit entangled states the singlet state

$$|\Psi_-\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle - |10\rangle\right) \tag{7.15}$$

is the most intriguing of all. It has a unique property that it remains invariant under the application of the same unitary operation to both qubits: $U \otimes U|\Psi_-\rangle = |\Psi_-\rangle$. This implies that we may write the singlet state in the same form but using different basis vectors. If we return to the polarization notation, this implies in particular that the three apparently different formulas

$$\frac{1}{\sqrt{2}}\left(|{\leftrightarrow}{\updownarrow}\rangle - |{\updownarrow}{\leftrightarrow}\rangle\right) = \frac{1}{\sqrt{2}}\left(|{⤢}{⤡}\rangle - |{⤡}{⤢}\rangle\right) = \frac{1}{\sqrt{2}}\left(|{\circlearrowright}{\circlearrowleft}\rangle - |{\circlearrowleft}{\circlearrowright}\rangle\right) \tag{7.16}$$

describe the same state $|\Psi_-\rangle$.

Imagine now that $|\Psi_-\rangle$ state is shared by Alice and Bob. Alice may decide to measure her qubit in either $\{|{\leftrightarrow}\rangle, |{\updownarrow}\rangle\}$, $\{|{⤢}\rangle, |{⤡}\rangle\}$ or $\{|{\circlearrowright}\rangle, |{\circlearrowleft}\rangle\}$ basis. Due to the properties of the singlet state, the conditional state of Bob will always be the complementary state to the one measured by Alice, hinting us to call it an example of *remote state preparation*. By choosing to measure her qubit in e.g. $\{|{\leftrightarrow}\rangle, |{\updownarrow}\rangle\}$ basis, Alice causes the state at the Bob's site to be either $\{|{\leftrightarrow}\rangle$ or $|{\updownarrow}\rangle\}$, etc. In this sense Alice may decide what kind of polarization state she wishes Bob should have, whether it be one of $\{|{\leftrightarrow}\rangle$ or $|{\updownarrow}\rangle\}$, one of $\{|{⤢}\rangle, |{⤡}\rangle\}$ or one of $\{|{\circlearrowright}\rangle, |{\circlearrowleft}\rangle\}$. If Bob was able to distinguish between these three cases without communication from Alice, we would have to admit that instantaneous communication at a distance is possible. Fortunately (or sadly. . .) this is not the case. Even though Alice *chooses* the measurement basis, she *does not choose* the measurement outcome. For the singlet state the two possible outcomes are equally probable for every basis. Therefore irrespectively on the measurement basis choice, the state on the Bob's site on average is the maximally mixed state, since:

$$\mathbb{1} = \left(|{\leftrightarrow}\rangle\langle{\leftrightarrow}| + |{\updownarrow}\rangle|{\updownarrow}\rangle\right) = \left(|{⤢}\rangle\langle{⤢}| + |{⤡}\rangle|{⤡}\rangle\right) = \left(|{\circlearrowright}\rangle\langle{\circlearrowright}| + |{\circlearrowleft}\rangle|{\circlearrowleft}\rangle\right) \tag{7.17}$$

and as such Bob is not able to tell which measurement basis Alice had chosen.

Even though entanglement properties of the singlet state do not allow for instantaneous communication, they may be used in practical applications such as e.g. *quantum teleportation* described in the next section.

7.2.1 Consider the most general two-qubit density matrix $\rho_{AB} = \sum_{ik,jl=0}^{1} \rho_{jl}^{ik} |i\rangle_A \langle j| \otimes |k\rangle_B \langle l|$ which is written explicitly as:

$$\rho_{AB} = \begin{pmatrix} \rho_{00}^{00}, \rho_{01}^{00}, \rho_{10}^{00}, \rho_{11}^{00} \\ \rho_{00}^{01}, \rho_{01}^{01}, \rho_{10}^{01}, \rho_{11}^{01} \\ \rho_{00}^{10}, \rho_{01}^{10}, \rho_{10}^{10}, \rho_{11}^{10} \\ \rho_{00}^{11}, \rho_{01}^{11}, \rho_{10}^{11}, \rho_{11}^{11} \end{pmatrix} \tag{7.18}$$

Calculate the reduce density matrix $\rho_A = \mathrm{Tr}_B(\rho_{AB})$, and formulate a simple operational method to calculate it given the full matrix explicitly.

7.2.2 Suppose that Alice projects onto $|\pm \mathbf{a}\rangle$, while Bob performs a measurement $\hat{B}_r$. Verify that the probability of Alice obtaining the result $+\mathbf{a}$ while Bob's outcome can be any is $\sum_r p(\mathbf{a}, r) = \langle\tilde{\psi}(\mathbf{a})|\tilde{\psi}(\mathbf{a})\rangle$. Further, the conditional probability for result r on Bob's side given that Alice has obtained $+\mathbf{a}$ can be expressed as, using Eq. (6.27):

$$p(r|\mathbf{a}) = \frac{\langle\tilde{\psi}(\mathbf{a})|\hat{B}_r|\tilde{\psi}(\mathbf{a})\rangle}{\langle\tilde{\psi}(\mathbf{a})|\tilde{\psi}(\mathbf{a})\rangle} \tag{7.19}$$

7.2.3 Show that indeed the singlet state is invariant under application of the same unitary operation to both systems, i.e.: $U \otimes U|\Psi_-\rangle = |\Psi_-\rangle$.

7.3 Teleportation

Imagine that you want to teleport an unknown quantum state of a physical system to a distant place without physically sending the actual system to this place. Recall the Star-Trek teleporter device, where an object to be teleported is transformed to pure energy (light?), then beamed to a distant placed and than rematerialize at the final spot. Authors as well as fans of the series had a serious puzzle how to reconcile faithful teleportation with fuzzy nature of the quantum states, and in particular their non perfect distinguishability. One could think that the teleporter simply scans all the atoms of the object in order to read out its quantum state and then uses this information to recreate the object at the final stage of the process. We have learned however, that non-orthogonal quantum states are fundamentally non-distinguishable and as such the read-out process will never be perfect. Therefore, we cannot really be sure in what quantum state the object to be teleported was initially and as such the final state will in general differ from the original. It seems that the there is no way out, and this is probably what authors of the series though when they referred to enigmatic “Heisenberg compensators” allowing for temporal violation of the rules of quantum mechanics in order to make the teleportation process possible. A few years have passed, and in 1993 physicists realized that the quantum state teleportation is possible within the laws of quantum mechanics and the necessary resource for this is entanglement.

To understand how this is possible, let us consider now a more modest task of teleporting an unknown state of a single qubit Imagine you are Alice and you receive a single qubit in a unknown state $|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$. Your goal is to reproduce its state in a distant Bob's laboratory without physically sending the qubit there. Clearly, due to impossibility of discriminating non-orthogonal states you cannot simply measure the qubit and give Bob an instruction on how to reproduce the state in his laboratory. However, assume that you and Bob share additionally two qubits in a maximally entangled state, e.g. $|\Psi_-\rangle$. Therefore, the initial state of the three qubits can be written as $|\psi\rangle_1 \otimes |\Psi_-\rangle_{23}$, where qubits $1, 2$ are in the Alice's laboratory while the qubit 3 is in possession of Bob.

The key step to understand how the teleportation is possible is to notice that the input state

$$\begin{aligned} |\psi\rangle_1 \otimes |\Psi_-\rangle_{23} = (\alpha_0|0\rangle + \alpha_1|1\rangle) \otimes \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) = \\ = \frac{1}{\sqrt{2}}(\alpha_0|001\rangle + \alpha_1|101\rangle - \alpha_0|010\rangle - \alpha_1|110\rangle) \end{aligned} \tag{7.20}$$

can be rewritten in an equivalent form:

$$\begin{aligned} \frac{1}{2}[-|\Psi_-\rangle \otimes (\alpha_0|0\rangle + \alpha_1|1\rangle) - |\Psi_+\rangle \otimes (\alpha_0|0\rangle - \alpha_1|1\rangle) + \\ +|\Phi_-\rangle \otimes (\alpha_0|1\rangle + \alpha_1|0\rangle) + |\Phi_+\rangle \otimes (\alpha_0|1\rangle - \alpha_1|0\rangle)]. \end{aligned} \tag{7.21}$$

In order to realize the teleportation, Alice measures her two qubit in the Bell basis: $\{|\Psi_i\rangle\}$ where $|\Psi_1\rangle = |\Psi_-\rangle$, $|\Psi_2\rangle = |\Psi_+\rangle$, $|\Psi_3\rangle = |\Phi_-\rangle$, $|\Psi_4\rangle = |\Phi_+\rangle$. From Eq. (7.21) it is clear that all four measurement outcomes are equally probable and the conditional states of the third qubit are respectively:

$$\begin{aligned} |\tilde{\psi}(1)\rangle = \frac{1}{2}(\alpha_0|0\rangle + \alpha_1|1\rangle), \ |\tilde{\psi}(2)\rangle = \frac{1}{2}(\alpha_0|0\rangle - \alpha_1|1\rangle), \\ |\tilde{\psi}(3)\rangle = \frac{1}{2}(\alpha_0|1\rangle + \alpha_1|0\rangle), \ |\tilde{\psi}(4)\rangle = \frac{1}{2}(\alpha_0|1\rangle - \alpha_1|0\rangle). \end{aligned} \tag{7.22}$$

Only the first of these four states is the original state to be teleported. Nevertheless, once Bob learns Alice's measurement outcome i, he can apply a local operation U_i to his qubit in order to recover the original state. Depending on the measurement outcome the required operation U_i is:

$$U_1 = \mathbb{1},\ U_2 = \sigma_z,\ U_3 = \sigma_x,\ U_4 = \mathrm{i}\sigma_y. \tag{7.23}$$

After this the teleportation protocol is complete and the state of the qubit in Bob's laboratory is $|\psi\rangle$. We may write the teleportation protocol in a consise mathematical form as:

$$\sum_{i=1}^{4} U_i|\tilde{\psi}(i)\rangle_3 = |\psi\rangle_3, \quad |\tilde{\psi}(i)\rangle_3 = {}_{12}\langle\Psi_i|\left(|\psi\rangle_1 \otimes |\Psi_-\rangle_{23}\right), \tag{7.24}$$

where $|\tilde{\psi}(i)\rangle_3$ is a state of qubit 3 conditioned on Alice obtaining the measurement outcome i in her Bell measurement on qubits 1 and 2.

The essential part of the protocol is communicating the measurement result from Alice to Bob. Until Bob learns the measurement result, he cannot apply his local operations and the teleportation process cannot be regarded as complete. Notice, that in the case Bob is ignorant about the result of Alices measurement the average state of his qubit is:

$$\sum_{i=1}^{4} |\tilde{\psi}(i)\rangle\langle\tilde{\psi}(i)| = \frac{1}{2}\mathbb{1} \tag{7.25}$$

and carries no information on the teleported state. Only when the measurement results is learned by Bob and the correcting operations U_i applied we can claim that the state of the qubit has been teleported to Bob. In this sense teleportation process is not instantaneous, but is limited by the speed of classical communication of the result i hence the speed of light.

It is also worth pointing out that neither Alice nor Bob learns anything about the teleported state during the teleportation process. Moreover, the information carried by the original state is completely destroyed due to the Bell measurement by Alice, and hence the state is indeed teleported and not cloned. This is in accordance with the no-cloning theorem forbidding cloning of unknown quantum states that we will discuss in Sec.???.

7.3.1 Show that $(\alpha_0|0\rangle + \alpha_1|1\rangle) \otimes |\Psi_-\rangle$ can be rewritten as in Eq.(7.21)

7.3.2 Formulate the teleportation protocol, but this time assuming Alice and Bob share $|\Phi_+\rangle$, rather then $|\Psi_-\rangle$ Bell state.

7.4 No-cloning and the impossibility of superluminal communication

Event though quantum mechanics allow for a remote state preparation using entangled states, the stochastic nature of the quantum measurement does not allow to exploit this feature for instantaneous communication. When analyzing the properties of the singlet state in Sec.??? we have observed that by choosing her measurement basis Alice chooses the basis from which a state is randomly chosen at Bob's site. If Bob was able to determine what measurement basis Alice had chosen the instantaneous communication would be possible. In our example, Bob would need to be able to distinguish whether he got of $|\leftrightarrow\rangle, |\updownarrow\rangle$ or one of $|\swarrow\!\!\!\!\!\!\nearrow\rangle, |\nwarrow\!\!\!\!\!\!\searrow\rangle$. Since both states in each basis are equally probable the average density matrix is $\mathbb{1}/2$ irrespectively of the measurement basis chosen by Alice.

Nevertheless, it instructive to push further and consider at least in our imagination some strategies that could in principle distinguish between these two cases. One of such strategies is to use a cloning machine. Imagine that we have a cloning machine that is capable of produce two copies out of a single copy of an unknown quantum state. In order to describe a quantum cloning machine in a rigorous way consider the Hilbert space $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_X$, where $\mathcal{H}_1$ is the space supporting the state of a system to be copied, $\mathcal{H}_2$ supports the states of the system which is our "blank page", and $\mathcal{H}_X$ supports all other degrees of freedom including the copying machine and the rest of the universe. We say that the operation U_{cloning} (according to quantum theory should be unitary) performs cloning of a state $|\psi\rangle$ iff:

$$U_{\text{cloning}}|\psi\rangle \otimes |0\rangle \otimes |X_0\rangle = |\psi\rangle \otimes |\psi\rangle \otimes |X_\psi\rangle. \qquad (7.26)$$

In other words should produce a state $|\psi\rangle$ in both systems 1 and 2 while the remaining degrees of freedom can change depending on the cloned state. Notice that the output state is a product state – there is no entanglement between subsystems. It has to be so, otherwise clones when inspected independently would be in mixed states.

If such a machine existed then Bob could apply it to his qubit and produce two, or even arbitrary many copies of the state. Assume that Bob has produced N perfect copies of his state, so he has $|\psi\rangle^{\otimes N}$ at his disposal. If so, then returning to our hypothetical superluminal communication protocol his goal would amount to distinguish whether he has one of $\{|\leftrightarrow\rangle^{\otimes N}, |\updownarrow\rangle^{\otimes N}\}$ or one of $\{|\nearrow\rangle^{\otimes N}, |\nwarrow\rangle^{\otimes N}\}$. But now, his average density matrices are different whenever N¿1:

$$\frac{1}{2}\left(|\leftrightarrow\rangle\langle\leftrightarrow|^{\otimes N} + |\leftrightarrow\rangle\langle\leftrightarrow|^{\otimes N}\right) \neq \frac{1}{2}\left(|\nearrow\rangle\langle\nearrow|^{\otimes N} + |\nwarrow\rangle\langle\nwarrow|^{\otimes N}\right). \qquad (7.27)$$

which means that Bob will see the difference between events in which Alice had chosen one or the other measurement basis. In particular if $N \to \infty$ the four states at Bobs site become orthogonal ($\lim_{N\to\infty} \langle\leftrightarrow|\nearrow\rangle^N = 0$) and hence could be discriminated perfectly. This observation is another argument that quantum cloning must be forbidden by the laws of quantum mechanics, since we have proven in Sec. that perfect discrimination of nonorthogonal states. If cloning was possible we could produce arbitrary number of copies of non-orthogonal states and then would be able to discriminate them perfectly, violating the fundamental bounds derived in Sec.???. The following no-cloning theorem shows explicitly that the cloning task is impossible.

No cloning theorem There is no deterministic cloning transformation (unitary) performing cloning for two nonorthogonal state.

Proof. Let $|\psi_1\rangle$, $|\psi_2\rangle$ be two different nonorthogonal states: $0 < |\langle\psi_1|\psi_2\rangle| < 1$. Assume cloning is possible:

$$\begin{aligned} U|\psi_1\rangle \otimes |0\rangle \otimes |X_0\rangle &= |\psi_1\rangle \otimes |\psi_1\rangle \otimes |X_{\psi_1}\rangle && (7.28)\\ U|\psi_2\rangle \otimes |0\rangle \otimes |X_0\rangle &= |\psi_2\rangle \otimes |\psi_2\rangle \otimes |X_{\psi_2}\rangle. && (7.29)\end{aligned}$$

Thanks to unitarity the scalar product of the input states should be equal to the scalar product of the output states:

$$\langle\psi_1|\psi_2\rangle\langle 0|0\rangle\langle X_0|X_0\rangle = \langle\psi_1|\psi_2\rangle\langle\psi_1|\psi_2\rangle\langle X_{\psi_1}|X_{\psi_2}\rangle \qquad (7.30)$$

this leads to:

$$\langle\psi_1|\psi_2\rangle(1 - \langle\psi_1|\psi_2\rangle\langle X_{\psi_1}|X_{\psi_2}\rangle) = 0 \qquad (7.31)$$

which is only possible for $\langle\psi_1|\psi_2\rangle = 0$ or $\langle\psi_1|\psi_2\rangle = 0$. Hence, we arrive at a contradiction and conclude that cloning of nonorthogonal states is impossible. ■

Even though perfect cloning of non-orthogonal states is impossible, approximate cloning operations find its application in analyzing the security of quantum key distribution.

7.5 Entanglement based quantum cryptography

While the standard BB84 protocol does not make use of entangled states it is possible to formulate an equivalent protocol that employs entanglement. Imagine that instead of sending one of four states form Alice to Bob, we assume that Alice and Bob share N pairs of qubits each in the maximally entangled state $|\Phi_+\rangle$. Alice randomly measures her qubit in either $\{|\leftrightarrow\rangle, |\updownarrow\rangle\}$ or $\{|\swarrow\!\!\!\nearrow\rangle, |\nwarrow\!\!\!\searrow\rangle\}$. It is easy to see that the conditional states of Bob will correspond to the state measured by Alice and as such Bob will obtain each of the four states with probability 1/4 just as in the original BB84 protocol. As in the original BB84 protocol Bob measures his qubit randomly in one of the two basis, and all the subsequent steps of basis reconciliation, error correction and privacy amplification remain unchanged. The main conceptual difference here is the symmetry between Alice and Bob, you can no longer tell who is the sender and who is the receiver, and in particular who generated the random key. The random key is generated as a result of measurements by Alice and Bob, and it does not matter who measures his/her qubit first.

Entanglement based cryptography has an appealing interpretation if we recall the violation of Bell inequalities by two-qubit entangled states. Since we know that e.g. $|\Phi_+\rangle$ violates Bell inequalities this implies that there no hidden parameters that predetermine the measurement outcomes of Alice and Bob. In this sense we can say that the bits of the secret key that Alice and Bob register did not exist before their measurements. Hence the knowledge on them is not available to other parties. We can even allow the potential eavesdropper to produce an entangled state for us. This is still secure provide we make sure that we indeed have a state which is sufficiently strongly entangled as e.g. $|\Phi_+\rangle$.

7.5.1 Imagine that instead of an idea state $|\Phi_+\rangle$, Alice and Bob share a noisy state of the form:

$$\rho_{AB} = (1-\eta)|\Phi+\rangle\langle\Phi_+| + \eta\mathbb{1}/4 \tag{7.32}$$

where η is the noise parameter.

- Calculate QBER as a function of η
- Above what η Bell inequalitites are no longer violated.
- The corresponding value of QBER is a good estimate of the QBER threshold above which security of quantum key distribution is not guaranteed. Compare it with the values we have obtained using simple intercept and resend strategies.

Chapter 8

Channels

8.1 Which way?

Let us go back to a single photon in Mach-Zehnder discussed in Sec. 2.3. Let us denote the state of the photon in the upper path as $|0\rangle_A$, in the lower path as $|1\rangle_A$. After the first beam splitter, the state of the photon is $|+\rangle_A = (|0\rangle_A + |1\rangle_A)/\sqrt{2}$. Suppose that we would like to find out which path the photon has taken, but without destroying it. Here is an idea: let us take a second qubit, labeled with a subscript E, and introduce an interaction by the following unitary transformation:

$$\hat{U}|i\rangle_A|j\rangle_E = (-1)^{ij}|i\rangle_A|j\rangle_E, \qquad i,j = 0,1. \tag{8.1}$$

It is seen that if the qubit A is the state $|0\rangle_A$ nothing changes, while if the qubit is in the state $|1\rangle_A$ the phase of the ancillary qubit state $|1\rangle_E$ is flipped.

If the qubit E was initially prepared in the state $|0\rangle_E$ nothing happes. But if we take $|+\rangle_E$ as the initial state then:

$$|\Psi\rangle_{AE} = \hat{U}\big(|+\rangle_A|+\rangle_E\big) = \frac{1}{\sqrt{2}}\big(|0\rangle_A|+\rangle_E + |1\rangle_A|-\rangle_E\big) \tag{8.2}$$

It is seen that measuring the qubit E in the basis $|+\rangle$, $|-\rangle$ tells us which path the photon has taken. Did we have to give up anything? The interaction has changed the state of the qubit A. It is now described by a density matrix

$$\hat{\varrho}'_A = \mathrm{Tr}_E\big(|\Psi\rangle_{AE}\langle\Psi|\big) = \frac{1}{2}\big(|0\rangle_A\langle 0| + |1\rangle_A\langle 1|\big). \tag{8.3}$$

This state is invariant with respect to the phase shift and the second beam splitter: the probabilities of detecting the photon at one or another port of the interferometer are now 1/2 and are independent of ϕ! By gaining which-path ingormation we erased interference fringes.

In contrast, if the qubit E was intially prepared in the state $|0\rangle_E$, then nothing would have changed. This means that we would still see interference fringes, but the state of the qubit E remains independent whether the photon took upper or lower path in the interferometer. Let us take more generally $|e_{\rm ini}\rangle_E = |\vartheta\rangle$, where $|\vartheta\rangle = \cos\vartheta|0\rangle + \sin\vartheta|1\rangle$. Evolution transforms into:

$$\hat{U}\big(|+\rangle_A|e_{\rm ini}\rangle_E\big) = \frac{1}{\sqrt{2}}\big(|0\rangle_A|e_0\rangle_E + |1\rangle_A|e_1\rangle_E\big). \tag{8.4}$$

where $|e_0\rangle = |\vartheta\rangle$ and $|e_1\rangle = |-\vartheta\rangle$. Let us quantify distinguishability of the ancilla qubit in terms of the pay-off, which we not must be less or equal than

$$\mathsf{P} \leq \sqrt{1 - |\langle e_0|e_1\rangle|^2} \tag{8.5}$$

What about fringes? The reduced density matrix written in the $|0\rangle$, $|1\rangle$ basis is now:

$$\hat{\varrho}'_A = \frac{1}{2}\begin{pmatrix} 1 & \langle e_1|e_0\rangle \\ \langle e_0|e_1\rangle & 1 \end{pmatrix} \tag{8.6}$$

Modulation is customarily characterized with the help of visibility:

$$V = \frac{p_{\max} - p_{\min}}{p_{\max} + p_{\min}} = |\langle e_0|e_1\rangle| \tag{8.7}$$

We immediately see that

$$\mathsf{P}^2 + V^2 \leq 1. \tag{8.8}$$

[Phys. Rev. Lett. by Englert in 1996].

8.1.1 Derive transformation of an arbitrary density matrix under the operation described above. How is the Bloch sphere transformed?

8.2 Quantum operations

The scenario considered in the preceding section is much more general. Suppose that a system A is prepared initially in a state $\hat{\varrho}_A$, and the auxiliary system in $|e_{\text{ini}}\rangle_E$, and then are subjected to a joint unitary evolution $\hat{U}$. Properties of the subsystem A after the evolution are described by a reduced density matrix:

$$\hat{\varrho}'_A = \text{Tr}_E\big[\hat{U}\big(\hat{\varrho}_A \otimes |e_{\text{ini}}\rangle_E\langle e_{\text{ini}}|\big)\hat{U}^\dagger\big] = \sum_i \big({}_E\langle e_i|\hat{U}|e_{\text{ini}}\rangle_E\big)\hat{\varrho}_A\big({}_E\langle e_{\text{ini}}|\hat{U}^\dagger|e_i\rangle_E\big) \tag{8.9}$$

where in the second form we introduced an orthonormal basis $|e_i\rangle_E$ over which the trace operation is performed. An operator acting on a joint system AE sandwiched between a bra and a ket corresponding to one of the subsystems yields an operator acting in on states of the remaining subsystem — this can be seen using matrix representation and we leave it as an exercise. Let us denote:

$$\hat{K}_i = {}_E\langle e_i|\hat{U}|e_{\text{ini}}\rangle_E \tag{8.10}$$

which allows us to write $\hat{\varrho}'_A = \sum_i \hat{K}_i\hat{\varrho}_A\hat{K}_i^\dagger$. These are called *Kraus operators.* Individual terms in this sum also have physical meaning. Suppose that after the interaction we perform a measurement $\hat{A}_r$ on the subsystem A and a projective measurement in the basis $|e_i\rangle$ on the subsystem E. The joint probability can be written as:

$$p_{ri} = \text{Tr}\big[\big(\hat{A}_r \otimes |e_i\rangle_E\langle e_i|\big)\hat{U}\big(\hat{\varrho}_A \otimes |e_{\text{ini}}\rangle_E\langle e_{\text{ini}}|\big)\hat{U}^\dagger\big] = \text{Tr}_A\big[\hat{A}_r\big(\hat{K}_i\hat{\varrho}_A\hat{K}_i^\dagger\big)\big] \tag{8.11}$$

Here $\hat{K}_i\hat{\varrho}_A\hat{K}_i^\dagger$ can be thought as the conditional density matrix describing A provided that an outcome i was measured on E. The trace of this operator is equal the probability of obtaining i, $p_i = \text{Tr}_A\big(\hat{K}_i\hat{\varrho}_A\hat{K}_i^\dagger\big)$. This can be seen immediately from the above equation by taking $\hat{\mathbb{1}}_A$ in place of $\hat{A}_r$. The conditional probability distributions for a measurement r is:

$$p(r|i) = \frac{\text{Tr}_A\big[\hat{A}_r\big(\hat{K}_i\hat{\varrho}_A\hat{K}_i^\dagger\big)\big]}{\text{Tr}_A\big(\hat{K}_i\hat{\varrho}_A\hat{K}_i^\dagger\big)} \tag{8.12}$$

The family $\hat{K}_i$ defines a transformation of the subsystem A. Secondly, it defines a measurement on the subsystem A: $p_i = \text{Tr}(\hat{\varrho}_A\hat{K}_i^\dagger\hat{K}_i)$. But it also tells us what happens to the system after a measurement: obtaining the outcome i leaves the subsystem A in the (unnormalized) state $\hat{K}_i\hat{\varrho}_A\hat{K}_i^\dagger$. It may happen that the measuring apparatus "glues" together several outcomes. Then the conditional density matrix would be a sum of $\hat{K}_i\hat{\varrho}_A\hat{K}_i^\dagger$ over a certain subset of is.

Lastly: do we need to define the initial and final states of E and the interaction $\hat{U}$? What conditions must be satisfied by $\hat{K}_i$s? We would like trace to be preserved. This leads to a condition [exercise]:

$$\sum_i \hat{K}_i^\dagger \hat{K}_i = \hat{\mathbb{1}}, \tag{8.13}$$

and this is the only requirement!

Several observations: a given transformation of the density matrix can be represented by different families of operators $\hat{K}_i$. For example, they depend on the measurement performed on the ancilla subsystem. Different measurements can yield different amounts of information, but they will introduce the same disturbance.

Given a measurement $\hat{M}_r$, we can associate an operation $\hat{K}_r = \sqrt{\hat{M}_r}$ (well defined, as $\hat{M}_r$ are positive). In particular, if we take rank-one projectors, $\hat{M}_i = |i\rangle\langle i|$, then also $\hat{K}_r = |i\rangle\langle i|$. This is sometimes called *collapse* of the state vector: after obtaining outcome i the system ends in the state $|i\rangle$. But it does not necessarily have to be that way. Kraus operators $\hat{K}_r = \hat{U}_i|i\rangle\langle i|$ where $\hat{U}_i$ can be chosen independently for each i would also induce the same projective measurement. What happens to the system after measurement depends on the actual specific interaction!

8.2.1 Calculate Kraus operators for the which-way experiment discussed in the preceding section for measurements in the 0/1 basis and the ± basis.

8.3 Complete positivity

The problem of physical transformations of a quantum state can be approached from a more abstract perspective. In general such transformations will be described by a certain map $\Lambda(\hat{\varrho})$. It is natural to require linearity which warrants proper action on statistical ensembles. Further, we impose trace preservation: if $\mathrm{Tr}\hat{\varrho} = 1$, then also $\mathrm{Tr}[\Lambda(\hat{\varrho})] = 1$. Finally, we require that if $\hat{\varrho}$ is positive, then also $\Lambda(\hat{\varrho})$ is positive. This condition is called *positivity* of the map Λ. Is this a complete set of conditions which guarantees that Λ describes a transformation that can be realized physically?

Let us consider the case of a single qubit. The action of a map Λ on a general density matrix can be written using the Bloch vector as:

$$\Lambda(\hat{\varrho}) = \frac{1}{2}[\Lambda(\hat{\mathbb{1}}) + s_1\Lambda(\hat{\sigma}_1) + s_2\Lambda(\hat{\sigma}_2) + s_3\Lambda(\hat{\sigma}_3)]. \tag{8.14}$$

Thus, owing to the linearity of Λ it is sufficient to know its action on $\hat{\mathbb{1}}$ and Pauli matrices. Let us start by considering a simple case:

$$\Lambda(\hat{\mathbb{1}}) = \hat{\mathbb{1}}, \qquad \Lambda(\hat{\sigma}_i) = \eta_i\hat{\sigma}_i, \quad i = 1, 2, 3. \tag{8.15}$$

The trace preserving property is satisfied. To make sure that hermitian matrices are mapped onto hermitian, we require that all three η_i's are real. Finally, to guarantee positivity, if $\mathbf{s}$ has norm less or equal to one, then also the vector $\begin{pmatrix}\eta_1 s_1\\ \eta_2 s_2\\ \eta_3 s_3\end{pmatrix}$ should have norm not exceeding one. This is satisfied if

$$-1 \le \eta_i \le 1, \quad i = 1, 2, 3. \tag{8.16}$$

Thus the set of all vectors $\boldsymbol{\eta}$ that guarantees positivity of Λ forms a cube. This is quite puzzling. For example $\eta_1 = \eta_2 = \eta_3 = -1$ would transform any pure state onto an orthogonal one. This is not a unitary transformation, we have seen that unitaries correspond to proper rotations of the Bloch sphere.

It turns out that the positivity condition is not sufficient for a map to be physical. Suppose the qubit under consideration is a part of a bigger system, and we subject the qubit A to a map Λ, while we leave the subsystem B intact. Such a procedure is described by a trivial extension to a map $\Lambda \otimes I$, where I denotes the identity map on the subsystem B. The catch is that positivity of Λ does not automatically imply positivity of $\Lambda \otimes I$! As an example, let us take B to be also a qubit and consider action of Λ on the singlet state $|\Psi_-\rangle_{AB}\langle\Psi_-|$. The result is:

$$(\Lambda\otimes I)(|\Psi_-\rangle_{AB}\langle\Psi_-|) = \frac{1}{4}\big(\hat{\mathbb{1}}\otimes\hat{\mathbb{1}} - \eta_1\hat{\sigma}_1\otimes\hat{\sigma}_1 - \eta_2\hat{\sigma}_2\otimes\hat{\sigma}_2 - \eta_3\hat{\sigma}_3\otimes\hat{\sigma}_3\big). \tag{8.17}$$

The operator on the right hand side is positive, if $\boldsymbol{\eta}$ is within a tetrahedron [exercise].

In general:

$$\Lambda(\hat{\mathbb{1}}) = \hat{\mathbb{1}} + \mathbf{a} \cdot \hat{\boldsymbol{\sigma}} \tag{8.18}$$

and

$$\Lambda(\sigma_i) = \sum_j L_{ij} \sigma_j \tag{8.19}$$

($\hat{\mathbb{1}}$ does not appear in the above sum as otherwise trace would not be preserved. Maps for which $\mathbf{a} = 0$ are called *unital* as the maximally mixed state is preserved. In this case we can apply singular value decomposition to $\mathbf{L}$ which is a proper rotation, diagonal matrix and another proper rotation. Elements of the diagonal matrix have to satisfy the "tetrahedron" conditions.

8.3.1 Time evolution of a qubit density matrix is given by:

$$\frac{\mathrm{d}\hat{\varrho}}{\mathrm{d}t} = -\frac{\gamma}{2}(\hat{\sigma}_+\hat{\sigma}_-\hat{\varrho} + \hat{\varrho}\hat{\sigma}_+\hat{\sigma}_- - 2\hat{\sigma}_-\hat{\varrho}\hat{\sigma}_-) \tag{8.20}$$

where $\hat{\sigma}_- = |1\rangle\langle 0|$. Assuming a general initial state find $\hat{\varrho}(t)$, corresponding transformation of the Bloch sphere, and an exemplary set of Kraus operators.

Chapter 9

Classical information theory

9.1 Data compression

Suppose that we need to encode a message composed of eight different symbols. The most straightforward is to use three bits. We do not need three bits if some of the letters do not appear at all, but that's trivial. What if some letters appear with lower probability than others? However, let these messages appear with different probabilities. Can we do better than that? Example in the third column of Table 9.1. *Codewords* of variable length. Average number of bits:

$$\frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \frac{1}{16} \cdot 4 + 4 \cdot \frac{1}{64} \cdot 6 = 2$$

Decoding is straightforward. The wisdom that more frequent letters should have shorter codewords has been exploited by Morse. Notice that thanks to the encoding each of the bits in a message is equally likely to be 0 or 1, irrespectively of the preceding or succeeding bit values. This is the essence of the compression. With each incoming bit the receiver learns the maximal possible amount of information since he has not bias to expect 1 rather then 0 or vice versa. If we had used the standard encoding where each letter is encoded in three bits this would be not the case.

Can we do even better than that? Suppose that symbols are statistically independent. It turns out that it is not possible. The average length is bounded from below by a quantity called *Shannon entropy.* A symbol can be treated as a random variable X which can assume one of values $x_i \in \mathcal{X}$

A	$\frac{1}{2}$	0
B	$\frac{1}{4}$	10
C	$\frac{1}{8}$	110
D	$\frac{1}{8}$	1110
E	$\frac{1}{64}$	111100
F	$\frac{1}{64}$	111101
G	$\frac{1}{64}$	111110
H	$\frac{1}{64}$	111111

Table 9.1: Exemplary coding

with probabilities $p_i = p(x_i)$. Shannon entropy is defined as:

$$\mathsf{H}(X) = \sum_i p_i \log_2 \frac{1}{p_i} = -\sum_i p_i \log_2 p_i \tag{9.1}$$

An easy calculation shows that for our example $\mathsf{H}(X) = 2$. Thus our code is as good as it can get. Below we provide an intuitive argument why $H(X)$ is a proper measure of "compressibility" of a message.

Consider N independent realizations of random variable X distributed according to $p(x)$ probability distribution. For large N, i.e. long sequences, the key notion is that of a *typical sequence*. It is clear from the law of large numbers that for larger N we will most likely encounter a symbol x approximately $Np(x)$ times in the sequence. Without going into technicalities, a sequence were each of the symbols x appear approximately $Np(x)$ we refer to as *a typical sequence*. Since a symbol x appears at $Np(x)$ places, the probability that we obtain a given typical sequence (with fixed order of symbols) is:

$$p^{(N)}_{\text{typical}} \approx \prod_x p(x)^{Np(x)} = 2^{N\sum_x p(x)\log_2 p(x)} = 2^{-NH(X)}. \tag{9.2}$$

Notice the already familiar formula for the Shannon entropy appearing in the exponent. What does it have to do with the compressibility of a message? As mentioned before, for long sequence we will *almost always* generate a typical sequences. This implies that the sum of probabilities of all typical sequences is approximately 1. However, as should be clear from Eq. (9.2) each typical sequence has roughly the same probability. Hence if by $\mathcal{N}_{\text{typical}}$ we denote the number of typical sequences we have:

$$\mathcal{N}^{(N)}_{\text{typical}} p^{(N)}_{\text{typical}} \approx 1 \quad \Rightarrow \quad \mathcal{N}^{(N)}_{\text{typical}} \approx 2^{NH(X)}. \tag{9.3}$$

In other words since there are $2^{NH(X)}$ typical N symbol sequences it is in principle possible to label them using $NH(X)$ bits. Moreover, as for long sequences sequences that are *not* typical will almost never happen, it is enough to transmit the labels of typical sequences and practically all information will be safely transmitted. This is the intuitive argument why the Shannon entropy is a useful measure of compressibility of the message. Of course, apart from a theoretical argument, one needs to find a practical way in which to encode the symbols, one of simple but quite efficient solutions is the Huffman coding (see exercise 9.1.2).

Consider a random variable X with $|\mathcal{X}|$ possible values. What probability distribution will yield the maximal value of $H(X)$?. The maximum entropy will correspond to the "most random" distribution, i.e. $p(x) = 1/|\mathcal{X}|$, in which case $H(X) = \log_2 |\mathcal{H}|$. This can be proven as follows. Function $h(t) = -t \log_2 t$ is concave, which implies that for any $w_i \geq 0$ that sum up to one and any arguments t_i we have:

$$\sum_i w_i h(t_i) \leq h\left(\sum_i w_i t_i\right). \tag{9.4}$$

Making use of this property we can write:

$$\frac{1}{|\mathcal{X}|} H(X) = \sum_x \frac{1}{|\mathcal{X}|} h[p(x)] \leq h\left(\frac{1}{|\mathcal{X}|}\sum_x p(x)\right) = \frac{1}{|\mathcal{X}|}\log_2 |\mathcal{X}|. \tag{9.5}$$

Suppose that we have two random variables $X = x_i$ and $Y = y_j$ characterized by a joint probability distribution $p(x_i, y_j)$. The joint entropy of the two variables is defined as:

$$\mathsf{H}(X,Y) = -\sum_{ij} p(x_i, y_j) \log_2 p(x_i, y_j)$$

If the two variables are statistically independent then $p(x_i, y_j) = p(x_i)p(y_j)$ and:

$$\mathsf{H}(X,Y) = -\sum_{ij} p(x_i)p(y_j)[\log_2 p(x_i) + \log_2 p(y_j)] = \mathsf{H}(X) + \mathsf{H}(Y) \tag{9.6}$$

Shannon entropy is *additive*. A very important property. In particular for N indpendent realizations of X we have $H(X^N) = NH(X)$.

9.1.1 Show that the Shannon entropy H(X) is always positive

9.1.2 Apply the Huffman coding to a message consisting of 6 different symbols appearing with probabilities $p(x) = (0.25, 0.25, 0.2, 0.1, 0.1, 0.1)$. Compare the average number of bits used per symbol with the Shannon entropy. How the result would change if you applied Huffman coding to pairs of symbols?

9.1.3 Count the number of typical sequences by considering all possible ordering of the symbols assuming a given symbol x appears $Np(x)$ times, and taking the limit of large N. Try to rederive in this way the $2^{NH(X)}$ formula.

9.2 Channel capacity

Let us consider a channel which transforms the input symbol x into the output symbol y. Let $p(y|x)$ be the conditional probability describing the action of the channel. Consider N independent uses if the channel and ask what is the amount of information we can transmit without errors. This quantity will depend on the way the channel transmits the symbols i.e. $p(y|x)$ as well as on our encoding scheme i.e. the way we choose the input symbols x.

Let us define the conditional entropy of random variable Y provided the input symbol was x:

$$H(Y|x) = -\sum_y p(y|x) \log_2 p(y|x). \tag{9.7}$$

Intuitively, this quantity tells us how uncertain is the value of symbol y for a fixed input symbol x. If input symbols are send with probability $p(x)$ then on average the conditional entropy reads:

$$H(Y|X) = \sum_x p(x) H(Y|x) = -\sum_{x,y} p(x) p(y|x) \log_2 p(y|x). \tag{9.8}$$

We can interpret this quantity as: "how random is Y provided we know X". In other words: the more noisy is the channel the bigger is $H(Y|X)$.

Let us now try to give an operational interpretation of these quantities in terms of information transmission through the channel. Consider a sequence of N independent realizations of the input random variable X. We again restrict our analysis to typical sequences as the non-typical sequence are negligible in the limit of large N. In a typical sequence, each value x will appear approximately $Np(x)$ times. Consider a given value x_0 for a moment. When send through the channel we will get at the output $Np(x_0)$ symbols y distributed according to the probability distribution $p(y|x_0)$. Provided $Np(x_0)$ is large we can again apply the typical sequence argument and say that at the output we may expect $2^{Np(x_0)H(Y|x_0)}$ typical sequences. We can interpret this as the number of "typical errors" that could happen to a string of symbols x_0. Finally, taking into account all possible values x_0 we can say that the number of typical sequences at the output that arise from a given typical sequence $\{x\}^N$ at the input is equal to:

$$\mathcal{N}_{\text{typical}}^{\{x\}^N \to Y^N} \approx 2^{\sum_x Np(x)H(Y|x)} = 2^{NH(Y|X)}. \tag{9.9}$$

On the other hand the number of typical sequences that appear at the output irrespectively of the input sequence is $2^{NH(Y)}$. If we want to encode information in sequences that can be decoded unambiguously from a typical output string we cannot encode more codewords than:

$$\mathcal{N}_{\text{error free codewords}}^{X^N \to Y^N} \approx \frac{2^{NH(X)}}{2^{NH(Y|X)}} = 2^{NI(X:Y)}, \tag{9.10}$$

where the quantity

$$I(X:Y) = H(Y) - H(Y|X) \tag{9.11}$$

is called the mutual information. The mutual information may interpreted as the reduction of entropy of the variable Y thanks to learning the value of variable X. Clearly it reflects how strongly X and Y are correlated. and hence it is no surprise that it appeared in the formula for the number of codewords we can faithfully transmit through a channel. Since the number of codewords is $2^{NI(X:Y)}$, the mutual information can be interpreted as a number of logical bits that can be transmitted faithfully per single use of the channel.

$I(X:Y)$ depends both on the channel, i.e. $p(y|x)$, as well as on the input encoding described by $p(x)$. If we ask for maximal transmission rate we are free to optimize over the input probability distribution. This is captured by the notion of channel capacity:

$$C = \max_{p(x)} I(X:Y) \tag{9.12}$$

which gives the maximal amount of logical bits that can be transmitted per single use of the channel.

9.2.1 Calculate joint and marginal entropies for

Y \ X	A	B	C	D
A	0	0	0	$\frac{1}{4}$
B	$\frac{1}{4}$	$\frac{1}{4}$	0	0
C	0	0	$\frac{1}{4}$	0

Show that $H(X) = H(X, Y)$, which means that X tells us everything about Y, but $H(Y) < H(X, Y)$.

9.2.2 When $H(X) = H(X, Y)$? In words: when the knowledge of X tells us everything about Y, and we do not get any new information by learning Y.

9.2.3 Find the capacity of the binary noisy channel where $p(0|0) = p(1|1) = p$, $p(0|1) = p(1|0) = 1 - p$.

9.2.4 Find the capacity of noisy channel where errors are signaled by a third output symbol "e", i.e. $p(0|0) = p(1|1) = p$, $p(e|0) = p(e|1) = 1 - p$.

9.3 Application to quantum key distribution

Chapter 10

Communication

10.1 Von Neumann entropy

In the previous chapter we learnt how to quantify uncertainty in the context of information transmission. What about communication using quantum systems? The basic quantity in this context is the *von Neumann entropy* $S(\hat{\varrho})$ defined as:

$$S(\hat{\varrho}) = -\mathrm{Tr}(\hat{\varrho} \log_2 \hat{\varrho}). \tag{10.1}$$

It is easy to see by calculating the right hand side in the basis of eigenvectors $|u_i\rangle$ that von Neumann entropy is equal to the Shannon entropy of its eigenvalues. Obviously, $S(\hat{\varrho}) = 0$ if and only if $\hat{\varrho}$ is a pure state. Von Neumann entropy is invariant with respect to unitary transformations of the density matrix, $S(\hat{U}\hat{\varrho}\hat{U}^\dagger) = S(\hat{\varrho})$

Operational meaning: the statistical ensemble representation is not unique, in general:

$$\hat{\varrho} = \sum_i w_i |\psi_i\rangle \langle\psi_i|. \tag{10.2}$$

For any such representation, $H(\{w_i\}) \geq S(\hat{\varrho})$. Thus, von Neumann entropy is the minimum Shannon entropy associated with the probability distribution of any statistical ensemble. For a proof, see Bengtsson and Życzkowski.

Alternative: suppose that we perform a measurement composed of rank-1 projectors $\hat{M}_r$. It can be also shown that $H(\{p_r\}) \geq S(\hat{\varrho})$.

Classically, entropy of two variables is always larger than the entropy of a single one. This does not hold for Shannon entropy. A maximally entangled state of two qubits is an obvious counterexample.

10.2 Holevo bound

Suppose that Alice want to communicate a classical message. She can prepare a physical system in a state chosen from a discrete set $\{\hat{\varrho}_x\}$, and she chooses state $\hat{\varrho}_x$ with a probability $p(x)$. Bob applies a measurement $\hat{M}_y$. Channel capacity is defined by the joint distribution $p(x, y) = p(y|x)p(x)$, where $p(y|x) = \mathrm{Tr}(\hat{M}_y \hat{\varrho}_x)$. If the states overlap, our intuition is that they cannot be distinguished too well at the output. For a given ensemble, there should be an upper bound in the form of the Holevo quantity:

$$I(X:Y) \leq S\left(\sum_x p(x)\hat{\varrho}_x\right) - \sum_x p(x) S(\hat{\varrho}_x) \tag{10.3}$$

If we use pure states, the second term is zero. In general, for a d-dimensional system we obtain: $I(A:B) \leq \log_2 d$. Non-orthogonal states don't help in sending classical information.

10.3 Eavesdropping

Before we pass to the proof of Holevo bound, we will use it to analyze the security of BB84.

In Sec. 8.1 we have seen how to gain partial information about whether the qubit has been prepared in the state $|0\rangle$ or $|1\rangle$. The side effect was that a superposition $|+\rangle$ was perturbed. This is what is behind the security of the BB84 protocol: an attempt to gain information in the 0/1 basis results in errors in the +/- basis — and otherwise. This can be seen by changing the basis of the qubit A from

$$|0\rangle_A|\vartheta\rangle_E \to |0\rangle_A|\vartheta\rangle_E, \qquad |1\rangle_A|\vartheta\rangle_E \to |1\rangle_A|-\vartheta\rangle_E \tag{10.4}$$

to:

$$|+\rangle_A|\vartheta\rangle_E \to \cos\vartheta|+\rangle_A|0\rangle_E + \sin\vartheta|-\rangle_A|1\rangle_E, \tag{10.5}$$

$$|-\rangle_A|\vartheta\rangle_E \to \cos\vartheta|-\rangle_A|0\rangle_E + \sin\vartheta|+\rangle_A|1\rangle_E \tag{10.6}$$

Here is an idea for eavesdropping: after gaining information in the basis 0/1 let us introduce a second ancilla in a state $|\vartheta'\rangle_{E'}$ and try to gain information in the basis ±. This means that we extend the above equations to:

$$|+\rangle_A|\vartheta\rangle_E|\vartheta'\rangle_{E'} \to \cos\vartheta|+\rangle_A|0\rangle_E|\vartheta'\rangle_{E'} + \sin\vartheta|-\rangle_A|1\rangle_E|-\vartheta'\rangle_{E'} \tag{10.7}$$
$$|-\rangle_A|\vartheta\rangle_E|\vartheta'\rangle_{E'} \to \cos\vartheta|-\rangle_A|0\rangle_E|-\vartheta'\rangle_{E'} + \sin\vartheta|+\rangle_A|1\rangle_E|\vartheta'\rangle_{E'} \tag{10.8}$$

We can return to the 0/1 basis, which gives:

$$|0\rangle_A|\vartheta\rangle_E|\vartheta'\rangle_{E'} \to \cos\vartheta'|0\rangle_A|\vartheta\rangle_E|0\rangle_{E'} + \sin\vartheta'|1\rangle_A|\vartheta\rangle_E|1\rangle_{E'} \tag{10.9}$$
$$|1\rangle_A|\vartheta\rangle_E|\vartheta'\rangle_{E'} \to \cos\vartheta'|1\rangle_A|-\vartheta\rangle_E|0\rangle_{E'} + \sin\vartheta'|0\rangle_A|-\vartheta\rangle_E|1\rangle_{E'} \tag{10.10}$$

The above two sets of equations describe the same physical transformation, but depending on what information we want to obtain it is easier to use one or another representation. If Eve wants to gain information in the 0/1 basis about the state sent by Alice she simply tries to distinguish states $|\vartheta\rangle_E$ and $|-\vartheta\rangle_E$ of the qubit E. If she want to gain information in the ± basis, she tries to discriminate state $|\vartheta'\rangle_{E'}$ and $|-\vartheta'\rangle_{E'}$ of the qubit E'. But she also needs to measure qubit E in the basis 0/1 and make assignment accordingly.

Error rate in the $0,1$ basis: $Q_z = \sin^2\vartheta$, in the ± basis $Q_x = \sin^2\vartheta'$. Average error assuming that the two bases are equiprobable: $Q = \frac{1}{2}(Q_x+Q_y)$. Consequently mutual information between Alice and Bob: $I_{AB} = 1 - H(Q)$. Eve's information from Holevo's bound: $I_{AE} = \frac{1}{2}[H(Q_x) + H(Q_z)] \leq H(Q)$ — symmetric eavesdropping is optimal from Eve's point of view if she wants to keep the quantum bit error rate fixed at the value Q.

A sufficient information for security:

$$K = I_{AB} - I_{AE} \geq 0. \tag{10.11}$$

In our case $1 - 2H(Q) \geq 0$, which gives $Q \approx 11\%$.

10.3.1 Derive transformation of the Bloch sphere of the qubit A induced by the eavesdropping procedure described above.

10.4 Proof

The proof will be based on *strong subadditivity*: for a tripartite system described by a joint density matrix $\hat{\varrho}_{ABC}$ the following inequality holds:

$$S(\hat{\varrho}_{ABC}) + S(\hat{\varrho}_B) \leq S(\hat{\varrho}_{AB}) + S(\hat{\varrho}_{BC}). \tag{10.12}$$

Here $\hat{\varrho}_{AB}$ is the reduced density matrix for the subsystems AB and analogously for indices BC and B.

Let Q denote the quantum system used by Alice to communicate classical information to Bob. We will assume that Alice chooses an ensemble of states $\hat{\varrho}_i$ with corresponding probabilities p_i. We will write this strategy in a fully quantum mechanical form by introducing another system A whose orthogonal pure states, denoted as $|a_i\rangle$, serve Alice as "flags" that signal which state has been actually sent. The combined average initial state of AQ can be written as a formula:

$$\hat{\varrho}_{AQ}^{\text{ini}} = \sum_x p(x) |a_x\rangle \langle a_x| \otimes \hat{\varrho}_x$$

which states that the flag $|a_i\rangle$ is classically correlated in a one-to-one way with a state $\hat{\varrho}_i$ transmitted to Bob, and that this pair is used with a probability p_i. It is easy to see that:

$$\hat{\varrho}_Q^{\text{ini}} = \mathrm{Tr}_A(\hat{\varrho}_{AQ}^{\text{ini}}) = \sum_x p(x)\hat{\varrho}_i, \qquad \hat{\varrho}_A^{\text{ini}} = \mathrm{Tr}_Q(\hat{\varrho}_{AQ}^{\text{ini}}) = \sum_x p(x) |a_x\rangle \langle a_x|.$$

The second identity implies that the Shannon entropy of the distribution $\{p_i\}$, denoted here as $H(A)$, is equal to $S(\hat{\varrho}_A^{\text{ini}})$. An easy calculation, left as Exercise **??**, shows that the entropy of the state $\hat{\varrho}_{AQ}^{\text{ini}}$ is given by:

$$S(\hat{\varrho}_{AQ}^{\text{ini}}) = H(A) + \sum_x p(x) S(\hat{\varrho}_x). \tag{10.13}$$

Bob, having received the system Q from Alice, performs a generalized measurement. As discussed in Sec. **??**, such a measurement can be viewed as a unitary interaction $\hat{U}_{QB}$ with a probe system B and prepared initially in a state $|b_{\text{ini}}\rangle$. If we introduce a certain orthonormal basis $\{|b_r\rangle\}$ for the system B, the state of QB after the interaction can be written as:

$$\hat{U}_{QB}(\hat{\varrho} \otimes |b_{\text{ini}}\rangle \langle b_{\text{ini}}|) = \sum_{yy'} \hat{B}_y \hat{\varrho} \hat{B}_{y'} \otimes |b_y\rangle \langle b_{y'}|$$

where $\hat{B}_y = \langle b_y | \hat{U}_{QB} | b_{\text{ini}} \rangle$ and we assumed that the system Q is prepared initially in a pure state $|\psi\rangle$. The procedure described so far may leave the system B in a superposition of different states $|b_y\rangle$. However, for a proper projective measurement in the basis $\{|b_y\rangle\}$ we would like B to end up in a statistical mixture of states $|b_y\rangle$. We can accomplish this by introducing another system B' that is a replica of B and requiring that Bob's action results in the following unitary transformation:

$$|\psi\rangle \otimes |b_{\text{ini}}\rangle \otimes |b_{\text{ini}}\rangle \to (B_y|\psi\rangle) \otimes |b_y\rangle \otimes |b_y\rangle_{B'}.$$

It is then easy to verify that the reduced final density matrix of the system B can be written as $\sum_y \text{Tr}(B_y^\dagger B_y \hat{\varrho})|b_y\rangle\langle b_y|$, where $\text{Tr}(B_y^\dagger B_y \hat{\varrho})$ is the standard probability of obtaining the measurement outcome r given input state $|\psi\rangle$. The state of the combined systems after Bob's measurement is given by:

$$\hat{\varrho}_{AQBB'} = \sum_{xyy'} p(x)|a_x\rangle\langle a_x| \otimes (\hat{B}_y \hat{\varrho}_x \hat{B}_{y'}^\dagger) \otimes |b_y\rangle\langle b_{y'}| \otimes |b_y\rangle\langle b_{y'}|$$

and once we trace over the auxiliary system B' it reduces to:

$$\hat{\varrho}_{AQB} = \text{Tr}_{B'}(\hat{\varrho}_{AQBB'}) = \sum_{xy} p_(x)|a_x\rangle\langle a_x| \otimes (B_y \hat{\varrho}_x B_y^\dagger) \otimes |b_y\rangle\langle b_y|$$

The reduced density matrix $\hat{\varrho}_{AB}$ has the form:

$$\hat{\varrho}_{AB} = \text{Tr}_{QB'}(\hat{\varrho}_{AQBB'}) = \sum_{xy} p(y|x)p(x)|a_x\rangle\langle a_x| \otimes |b_y\rangle\langle b_y|$$

that is a statistical mixture of mutually orthogonal and thus distinguishable states $|a_i\rangle \otimes |b_r\rangle$ with probabilities $p(y|x)p(x)$. These probabilities specify that Alice chooses an xth symbol and Bob measures y. The von Neumann entropy of $\hat{\varrho}_{AB}$ is equal to the joint Shannon information $H(A, B)$ of the probability distribution $p(y|x)p(x)$. Furthermore, the Shannon entropy $H(Y)$ of Bob's outcomes is given by the von Neumann entropy $S(\hat{\varrho}_B)$.

We now have all the ingredients to prove the Holevo theorem. We will apply the strong subadditivity property specified in Eq. (10.12) to the final state after Bob's interaction, taking as A and B the systems described above, and $C = QB'$. Because a unitary evolution does not change the entropy of a density matrix:

$$S(\hat{\varrho}_{AQBB'}) = S(\hat{\varrho}_{AQ}^{\text{ini}} \otimes |b_{\text{ini}}\rangle\langle b_{\text{ini}}| \otimes |b_{\text{ini}}\rangle\langle b_{\text{ini}}|) = S(\hat{\varrho}_{AQ}^{\text{ini}}) = H(A) + \sum_x p_x S(\hat{\varrho}_x).$$

as the systems B and B' are initially uncorrelated with AQ and pure, and in the last step we used Eq. (10.13). For the same reason, $S(\hat{\varrho}_{QBB'}) = S(\hat{\varrho}_Q^{\text{ini}}) = S\left(\sum_x p(x)\hat{\varrho}_x\right)$. The remaining two entropies appearing in the inequality

(10.12) are $S(\hat{\varrho}_B) = H(Y)$ and $S(\hat{\varrho}_{AB}) = H(X, Y)$. Inserting all these values yields:

$$H(X) + \sum_x p(x) S(\hat{\varrho}_x) + H(Y) \leq H(X, Y) + S\left(\sum_x p_x \hat{\varrho}_x\right)$$

which after a trivial rearrangement is the Holevo bound.