2 Quantum tools and a first protocol

This chapter covers our first cryptographic task: we will learn how to encrypt quantum states! To prepare our entry into quantum communication and cryptography we first need to learn a little more about quantum information. Before proceeding, make sure you are comfortable with the notions introduced in Chapter 1. In this chapter we extend these notions in several ways that will be essential to model interesting cryptographic scenarii.

2.1 Probability notation

We start by recalling standard notions of probability theory, and defining associated notation which we use frequently. Consider a discrete random variable X taking values in a finite set $\mathcal{X}$. We often write $|X|$ for the size of the set $\mathcal{X}$ over which X ranges. The probability distribution of X is specified by a function $P_X(\cdot) : \mathcal{X} \to [0,1]$ such that for any $x \in \mathcal{X}$, $P_X(x)$ denotes the probability that X takes on a specific value $x \in \mathcal{X}$. Recall that for a probability distribution, the normalization condition $\sum_{x\in\mathcal{X}} P_X(x) = 1$ always holds. When the distribution is clear from context we use the shorthands

$$p_x = \Pr(X = x) = P_X(x) \ .$$

for the probability that x occurs. If P is a distribution and X a random variable, we will write $X \sim P$ to indicate that the distribution of X is P. We sometimes extend this notation and write $X \sim Y$ where X, Y are random variables to indicate that they have the same distribution.

Example 2.1.1. *Let $\mathcal{X} = \{1,2,3,4,5,6\}$ correspond to the faces of a six-sided die. If the die is fair, i.e. all sides have equal probability of occurring, then $P_X(x) = 1/6$ for all $x \in \mathcal{X}$. Using our shorthand notation this can also be written as $p_x = 1/6$. The size of the range of X is $|X| = 6$.* ∎

A random variable X ranging over a set $\mathcal{X}$ can be correlated with another random variable Y ranging over $\mathcal{Y}$. This means that they have a joint distribution $P_{XY}(\cdot,\cdot) : \mathcal{X} \times \mathcal{Y} \to [0,1]$ that is not necessarily a product. That is, $P_{XY}(x,y) \neq P_X(x)P_Y(y)$ in general, where P_X (resp. P_Y) is the marginal distribution of X (resp. Y), defined by $P_X(x) = \sum_{y\in\mathcal{Y}} P_{XY}(x,y)$ (and similarly for Y). This leads to the notion of *conditional probabilities* $P_{X|Y}(x|y)$, where $P_{X|Y}(x|y)$ is the probability that X takes on the value x, conditioned on the event that Y takes on the value y. Bayes' rule relates this conditional

probability to the joint probabilities.

$$P_{X|Y}(x|y) = \frac{P_{XY}(x,y)}{P_Y(y)} ,$$

whenever $P_Y(y) > 0$.[1] We use the following shorthands when it is clear which random variable we refer to:

$$p_{x|y} = \Pr(X = x|Y = y) = P_{X|Y}(x|y) .$$

Example 2.1.2. *Let $Y \in \mathcal{Y} = \{$"fair", "unfair"$\}$ refer to the choice of either a fair or an unfair die, each chosen with equal probability: $P_Y(\text{fair}) = 1/2$ and $P_Y(\text{unfair}) = 1/2$. If X denotes the fair or unfair die, where the unfair die always rolls a "6" (that is, $\mathcal{X} = \{1,2,3,4,5,6\}$, with $P_X(6) = 1$ and $P_X(x) = 0$ for $x \neq 6$), then $P_{X|Y}(x|\text{fair}) = 1/6$ for all x, but $P_{X|Y}(6|\text{unfair}) = 1$ and $P_{X|Y}(x|\text{unfair}) = 0$ for $x \neq 6$.* ∎

Exercise 2.1.1 Compute explicitly the joint probability $P_{XY}(x,y)$ for the random variables in Example 2.1.2.

Exercise 2.1.2 Suppose that Alice chooses between the fair or unfair die from Example 2.1.2 with probability $P_Y(\text{fair}) = P_Y(\text{unfair}) = 1/2$, but does not reveal to us which choice was wade. Imagine that we roll the (fair or unfair) die and obtain the outcome X. Suppose that we see $X = 3$. Can we guess what die Alice used? That is, what is the most likely value of Y, "fair" or "unfair"? Answer the same question in case we observe that $X = 6$.

2.2 Density matrices

Week 1, Lecture 1.2,Lecture 1: The density matrix

The quantum generalization of probability distributions, i.e. probability distributions over quantum states, are called *density matrices*. There are two main motivations for working with density matrices. The first motivation is to model the kind of scenario described above. Suppose for example that we build a device that prepares either a state $|\psi_1\rangle$, with some probability p_1, or a state $|\psi_2\rangle$, with probability p_2. Wouldn't it be nice to have a concise mathematical way to describe the "average" quantum state returned by this device, without having to resort to words as in the previous sentence? We will call such a state a *mixed state*, in contrast to the pure states which we have studied so far.

There is a second motivation for introducing density matrices, which is that they are necessary to describe the quantum state of a subsystem of a general system. To understand why this is the case, imagine given two quantum systems A and B. For example, A and B are two qubits such that the state of A and B is a normalized vector $|\psi\rangle_{AB} \in \mathbb{C}^4$. Given this situation, how can we mathematically describe the state of qubit A? Note that physically speaking, if we imagine qubits A and B as being in very far-away locations,

[1] The marginal distribution of X given $Y = y$ is undefined if y cannot occur, i.e. whenever $P_Y(y) = 0$.

then intuitively there must be a way to describe the state of A without referring to B at all. So how do we do it?

Consider first an easy case. Suppose that the joint state of A and B takes the form

$$|\psi\rangle_{AB} = |\psi_1\rangle_A \otimes |\psi_2\rangle_B \ .$$

Then the answer is clear: the state of A is the normalized vector $|\psi_1\rangle_A$. However, remember from Chapter 1 that there exists quantum states $|\psi\rangle_{AB}$ that cannot be written as a simple tensor product like this! A good example of such a state is the EPR pair

$$|\mathrm{EPR}\rangle_{AB} = \frac{1}{\sqrt{2}} |0\rangle_A |0\rangle_B + \frac{1}{\sqrt{2}} |1\rangle_A |1\rangle_B \, .$$

As shown in Exercise **??**, it is impossible to express $|\mathrm{EPR}\rangle_{AB} = |\psi_1\rangle_A \otimes |\psi_2\rangle_B$ for some states $|\psi_1\rangle_A$ and $|\psi_2\rangle_B$. In this case, how can we describe the state of A? It seems like we dug ourselves into a mathematical rabbit-hole. Either we find a way to describe the state of A, or there is a problem with our formalism. As we will see, the answer to this question is the same as the previous one: the notion of a *density matrix* will help us save the day.

2.2.1 Introduction to the formalism

We start by giving a different way to represent pure quantum states: as matrices. Recall that a ket $|\psi\rangle$ is a column vector, while the bra $\langle\psi|$ is a row vector. Therefore, $\rho = |\psi\rangle\langle\psi|$ is a rank-1 matrix: it has precisely 1 non-zero eigenvalue (equal to 1), with associated eigenstate $|\psi\rangle$. The matrix ρ is called the *density matrix* of the quantum state.

Example 2.2.1. *For the states* $|0\rangle$ *and* $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ *we obtain the density matrices*

$$|0\rangle\langle 0| = \begin{pmatrix} 1 \\ 0 \end{pmatrix} (1\ 0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} ,$$
$$|+\rangle\langle +| = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} (1\ 1) = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} .$$

■

How does writing down states as matrices help us resolve the questions above? To see how, let us first consider the first motivation that we gave: the need for a formalism that can represent probabilistic combinations of pure quantum states. But before that, let us remember that physically, the only information that we can obtain about a quantum state is obtained by performing a measurement. Moreover, if a state $|\psi\rangle$ is measured in a basis that contains the vector $|b\rangle$, then the probability of obtaining the outcome '$|b\rangle$' is given by

$$|\langle b|\psi\rangle|^2 = \langle b|\psi\rangle\langle\psi|b\rangle = \langle b|\rho|b\rangle \ , \tag{2.1}$$

where as earlier $\rho = |\psi\rangle\langle\psi|$ is the density matrix representation of the pure state $|\psi\rangle$. In words: the probability of obtaining the outcome '$|b\rangle$' is obtained by computing the *overlap* of $|b\rangle$ with ρ, which is defined as the quantity $\langle b|\rho|b\rangle$.

Remark 2.2.1. *While $|\psi\rangle \neq -|\psi\rangle$ as vectors, as matrices we have $|\psi\rangle\langle\psi| = (-|\psi\rangle)(-\langle\psi|)$. Thanks to the modulus squared in the computation of probabilities we see that the (-1) phase has no observable consequence, and so representing the state vector as a matrix does not lose information.*

Moving on, let us consider the case where our preparation device prepares one of two possible states, $|\psi_1\rangle$ or $|\psi_2\rangle$, with equal probability $p_1 = p_2 = 1/2$ as in 2.1. We claim that an accurate matrix representation of the state produced by the device can be obtained by taking the linear combination

$$\rho = \frac{1}{2}|\psi_1\rangle\langle\psi_1| + \frac{1}{2}|\psi_2\rangle\langle\psi_2| \ .$$

More generally, if the device prepares $|\psi_x\rangle$ with probability p_x, the density matrix representation of the resulting state is

$$\rho = \sum_x p_x |\psi_x\rangle\langle\psi_x| \ . \tag{2.2}$$

To verify that this choice of representation is accurate, consider what happens if we measure the state output by the device in a basis that contains the vector $|b\rangle$. If the state is $|\psi_x\rangle$ for some x, then the probability of obtaining the outcome '$|b\rangle$' is

$$q_{b|x} = |\langle b|\,\psi_x\rangle|^2 = \langle b|\psi_x\rangle\langle\psi_x|b\rangle \ .$$

Since the state $|\psi_x\rangle$ is prepared with probability p_x we expect the overall probability of obtaining the outcome '$|b\rangle$' to be

$$q_b = \sum_x p_x q_{b|x} \ .$$

Observe that

$$q_b = \sum_x p_x q_{b|x} = \sum_x p_x \langle b|\psi_x\rangle\langle\psi_x|b\rangle = \langle b| \left(\sum_x p_x |\psi_x\rangle\langle\psi_x| \right) |b\rangle = \langle b|\,\rho\,|b\rangle \ ,$$

which is precisely the same rule as (2.1). This means that the density matrix representation (2.2) captures the right amount of information about the state of the system so that the distribution of outcomes of any measurement on the state can be recovered using the generalized measurement rule (2.1).

Example 2.2.2. *Suppose more generally that a device prepares a state with density matrix ρ_x with probability p_x. Then the density matrix that describes the overall state prepared by the device is given by*

$$\rho = \sum_x p_x \rho_x \ .$$

The set of probabilities and density matrices $\mathcal{E} = \{(p_x, \rho_x)\}_x$ is called an ensemble *of states.* ∎

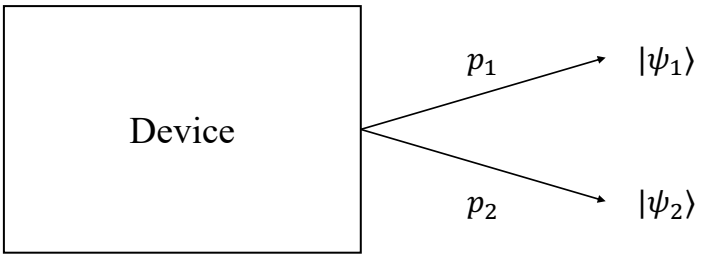

Fig. 2.1 Device that prepares two possible states with equal probability.

Example 2.2.3. *Suppose that a device prepares* $|0\rangle\langle 0|$ *with probability* $1/2$*, and* $|+\rangle\langle +|$ *with probability* $1/2$*. Then the resulting density matrix is given by*

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|+\rangle\langle +| = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix} .$$

■

Be careful that *superposition* is not the same as a *probabilistic combination*! Intuitively, the difference is that a probabilistic combination, often called a *mixture*, is an inherently classical object: there is a process that prepares one *or* the other state with some probability. In contrast, a state in a superposition is in some sense one *and* the other; it is a truly quantum phenomenon. The following example illustrates the difference between the two.

Example 2.2.4. *Consider the difference between preparing a* mixture *of* $|0\rangle\langle 0|$ *and* $|1\rangle\langle 1|$*, or creating a* superposition *over* $|0\rangle$ *and* $|1\rangle$*. First consider a source that prepares the states* $|0\rangle\langle 0|$ *and* $|1\rangle\langle 1|$ *with probabilities* $p_0 = p_1 = 1/2$*. Suppose we measure the resulting density matrix*

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}\mathbb{I}$$

in the Hadamard basis $\{|+\rangle, |-\rangle\}$*. Then the probability of each possible outcome is given by*

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

In contrast, consider now a state that is an equal superposition of $|0\rangle$ *and* $|1\rangle$*, i.e. the state* $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$*. Measuring* $|+\rangle$ *in the Hadamard basis results in* $q_+ = 1$ *and* $q_- = 0$*.*

The probabilities are different, so the two states are different! Indeed,

$$|+\rangle\langle+| = \frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} \neq \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} = \frac{1}{2}\mathbb{I}\,.$$

■

Remark 2.2.2. *Note that the same density matrix ρ can be obtained from different ensembles $\{(p_x, \rho_x)\}$. A simple example is provided by the density matrix*

$$\rho = \frac{\mathbb{I}}{2}\,,$$

which is also called the maximally mixed *state. You can verify that*

$$\frac{\mathbb{I}}{2} = \frac{1}{2}\left(|0\rangle\langle 0| + |1\rangle\langle 1|\right) = \frac{1}{2}\left(|+\rangle\langle+| + |-\rangle\langle-|\right)\,,$$

and many other equivalent decompositions are possible. (The maximally mixed state arises very frequently in cryptography, because it represents a state of complete uncertainty.*) What this means is that the two processes, generating the states $|0\rangle$ or $|1\rangle$ with probability $\frac{1}{2}$ each, or generating the states $|+\rangle$ or $|-\rangle$ with probability $\frac{1}{2}$ each, return quantum states that are physically indistinguishable: they have the same density matrix representation $\rho = (1/2)\mathbb{I}$.*

Quiz 2.2.1. *Suppose a system is produced in state $|0\rangle$ with probability $p_0 = \frac{1}{2}$ and in state $|-\rangle$ with probability $p_1 = \frac{1}{2}$. What is the resulting density matrix?*

a) $\rho = \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$

b) $\rho = \frac{1}{4}\begin{pmatrix}3 & 1\\ 1 & 1\end{pmatrix}$

c) $\rho = \frac{1}{4}\begin{pmatrix}3 & -1\\ -1 & 1\end{pmatrix}$

d) $\rho = \frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}$

2.2.2 A little bit of math

To formally define density matrices and their properties we recall some important notions from linear algebra. We start with Hermitian, and then positive semidefinite, matrices.

Definition 2.2.1 (Hermitian matrix M)**.** *A $d \times d$ complex matrix M is* Hermitian *if it satisfies $M^\dagger = M$, where recall from Definition 1.1.2 that $M^\dagger$ denotes the conjugate-transpose.*

To define density matrices formally, we need a few more mathematical concepts. The spectral theorem states that any Hermitian matrix M can be diagonalized with real eigenvalues. This means that there exists an orthonormal basis $\{|v_j\rangle\}$ of $\mathbb{C}^d$ (the *eigenvectors*) and real numbers λ_j (the *eigenvalues*) such that $M = \sum_j \lambda_j |v_j\rangle\langle v_j|$.

Definition 2.2.2 (Positive semidefinite matrix)**.** *A Hermitian matrix M is* positive semidefinite *if all its eigenvalues $\{\lambda_i\}_i$ are non-negative. This condition is denoted $M \geq 0$.*

Exercise 2.2.1 Show that a matrix M is positive semidefinite if and only if $\langle v| M |v\rangle \geq 0$ for all unit vectors $|v\rangle$. In particular, the diagonal coefficients $\langle i| M |i\rangle$ of M in any basis are non-negative.

Exercise 2.2.2 Show that the diagonal coeffiecients being positive is not a sufficient condition for M to be positive semidefinite: find an M such that the diagonal coefficients of M are all positive but M itself is not positive semidefinite.

Exercise 2.2.3 Even worse: find an M such that *all* coefficients (i.e. entries) of M are non-negative, but M is not positive semidefinite.

An important operation on matrices is the *trace*, which is simply the sum of the diagonal elements. It is convenient to note that the trace can also be expressed as follows.

Definition 2.2.3 (Trace of a matrix)**.** *The trace of a $d \times d$ matrix M is*

$$\mathrm{tr}(M) = \sum_{i=0}^{d-1} \langle i| M |i\rangle ,$$

where $\{|i\rangle\}$ is any *orthonormal basis of $\mathbb{C}^d$.*

The definition implicitly assumes that the definition of the trace does not depend on the choice of orthonormal basis. Let's verify that this is indeed the case. First, in the following exercise we verify an important property of the trace, which is that it is *cyclic*. We will frequently make use of this property in our calculations.

Exercise 2.2.4 Show that for any matrices M, N (such that both products MN and NM are well-defined) we have $\mathrm{tr}(MN) = \mathrm{tr}(NM)$. We will often use this property to perform manipulations such as

$$\langle i| A |i\rangle = \mathrm{tr}(\langle i| A |i\rangle) = \mathrm{tr}(A|i\rangle\langle i|) ,$$

where we made use of the fact the trace is cyclic with $M = \langle i|$ and $N = A|i\rangle$. (Make sure you can follow all the kets and bras!) It is worth noting that in general a non-cyclic permutation of the matrices does not preserve the trace. More precisely, for matrices M, N, P, in general

$$\mathrm{tr}(MNP) \neq \mathrm{tr}(NMP) .$$

Now, if $\{|u_i\rangle\}$ is any orthonormal basis of $\mathbb{C}^d$, we know that there exists a unitary transformation U such that $U|i\rangle = |u_i\rangle$ for all $i = 0, \ldots, d-1$. So given a $d \times d$ matrix

M,

$$\begin{aligned}\sum_i \langle u_i| M |u_i\rangle &= \sum_i \langle i| U^* M U |i\rangle \\ &= \mathrm{Tr}(U^\dagger M U) \\ &= \mathrm{Tr}(M U U^\dagger) \\ &= \mathrm{Tr}(M) \ ,\end{aligned}$$

where for the second line we used the cyclicity property and for the last line we used $UU^\dagger = \mathbb{I}$. This shows that our definition of the trace is indeed independent of the choice of orthonormal basis! In particular, by choosing the basis of eigenvectors of M, you can verify that for any Hermitian matrix M, $\mathrm{tr}(M)$ is the sum of its eigenvalues (counted with multiplicity).

2.2.3 Density matrices and their properties

Before we take the density matrix ρ as our new definition for a general quantum state, let us investigate when an arbitrary matrix ρ can be considered a valid density matrix, that is, a valid representation of a quantum state. It turns out that two properties are necessary and sufficient: the matrix ρ should be *positive semidefinite* and have *trace equal to 1*.

To see why this is true, consider the diagonalized representation of a density matrix ρ as a function of its eigenvalues $\{\lambda_j\}_j$ and corresponding eigenvectors $\{|v_j\rangle\}_j$:

$$\rho = \sum_j \lambda_j |v_j\rangle\langle v_j| \ .$$

Imagine that we measure ρ in an orthonormal basis $\{|w_k\rangle\}_k$. Based on (2.1) we know that the probability of obtaining the measurement outcome k is given by

$$q_k = \langle w_k| \rho |w_k\rangle \ . \tag{2.3}$$

For this to specify a proper distribution, it must be that $q_k \geq 0$ and $\sum_k q_k = 1$. By performing the measurement in the eigenbasis of ρ, $|w_j\rangle = |v_j\rangle$, we obtain the necessary conditions $\lambda_j \geq 0$, that is, ρ is a *positive semidefinite* matrix, and $\mathrm{tr}(\rho) = 1$, since

$$1 = \sum_j q_j = \sum_j \lambda_j \,\mathrm{tr}(|v_j\rangle\langle v_j|) = \mathrm{tr}(\rho) \ .$$

This shows that the two conditions are necessary for ρ to lead to well-defined distributions on measurement outcomes when using the rule (2.1). The following exercise asks you to show that the conditions are also sufficient.

Exercise 2.2.5 Show that for any positive semidefinite matrix ρ with trace 1, and any orthonormal basis $\{|w_k\rangle\}_k$, the numbers $q_k = \langle w_k| \rho |w_k\rangle$ are real, non-negative, and sum to 1.

We give a formal definition of a density matrix, which is the most general way of representing a quantum state.

Definition 2.2.4 (Density matrix)**.** *A* density matrix *on $\mathbb{C}^d$ is a $d \times d$ matrix ρ such that $\rho \geq 0$ and $\mathrm{tr}(\rho) = 1$. If furthermore ρ is of rank 1, then ρ is called a* pure *density matrix. Otherwise it is called a* mixed *density matrix.*

Note that by definition a pure density matrix is of the form $\rho = \lambda_1 |u_1\rangle\langle u_1|$, where the trace condition implies that necessarily $\lambda_1 = 1$. Thus for the case of pure states, density matrices and the vector representation we got used to before are in one-to-one correspondence. (Except for the phase, which as we pointed out is not relevant since there is no observation on the state that can determine it.)

We also summarize the rule for computing outcome probabilities when measuring a quantum system described by the density matrix ρ.

Definition 2.2.5 (Measuring a density matrix in a basis)**.** *Consider a density matrix ρ. Measuring ρ in the orthonormal basis $\{|b_j\rangle\}_j$ results in outcome j with probability*

$$q_j = \langle b_j | \rho | b_j \rangle \ .$$

Quiz 2.2.2. *Is $\rho = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ a valid density matrix?*

a) Yes
b) No

Quiz 2.2.3. *Is there always a unique way of preparing the state described by a given density matrix?*

a) Yes
b) No

2.2.4 Bloch representation for one-qubit mixed states

Week 1, Lecture 1.2,Lecture 3: The Bloch sphere representation of the density matrix

In Chapter 1 we saw that single-qubit states have a convenient graphical representation in terms of a vector on the Bloch sphere. In particular any pure quantum state can be described by a *Bloch vector* $\vec{r} = (\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta)$. Rather conveniently, the representation extends to mixed states. Concretely, it is always possible to write a single-qubit density matrix as

$$\rho = \frac{1}{2} \left(\mathbb{I} + v_x X + v_z Z + v_y Y \right) \ , \tag{2.4}$$

where X, Y, Z are the Pauli matrices defined in Chapter 1 and v_x, v_y, v_z are real coefficients. The fact that such an expansion always exists follows from the fact that the matrices $\mathcal{P} = \{\mathbb{I}, X, Y, Z\}$ form a basis for the space of 2×2 density matrices that correspond to a qubit.

Exercise 2.2.6 Use the fact that all matrices $M, N \in \mathcal{P}$ with $M \neq N$ anti-commute, i.e., $\{M, N\} = MN + NM = 0$ to show that $\mathrm{tr}(MN) = 0$ whenever $M \neq N \in \mathcal{P}$.

Exercise 2.2.7 Using the orthogonality condition (2.5), show that

$$|0\rangle\langle 0| = \frac{1}{2}\left(\mathbb{I} + Z\right) \ ,$$
$$|1\rangle\langle 1| = \frac{1}{2}\left(\mathbb{I} - Z\right) \ .$$

The exercise shows that the matrices $\mathbb{I}, X, Y, Z$ are orthogonal under the Hilbert-Schmidt inner product $\langle A, B\rangle = \mathrm{tr}(A^\dagger B)$. That is,

$$\mathrm{tr}(X^\dagger Y) = \mathrm{tr}(X^\dagger Z) = \mathrm{tr}(X^\dagger \mathbb{I}) = 0 \ , \tag{2.5}$$

and similarly for all other pairs of matrices. This is why we can refer to them as an orthonormal basis.

If ρ is pure you can verify that the vector $\vec{v} = (v_x, v_y, v_z)$ is precisely the Bloch vector $\vec{r}$ defined in Chapter 1. For pure states $\|\vec{v}\|_2^2 = v_x^2 + v_y^2 + v_y^2 = 1$. In other words, pure states live on the surface of the Bloch sphere. For mixed states, however, we can have $\|\vec{v}\|_2^2 \leq 1$. Mixed states thus lie in the interior of the Bloch sphere. For the case of 2×2 matrices the vector $\vec{v}$ tells us immediately whether the matrix ρ is a valid one qubit quantum state: this is the case if and only if $\|\vec{v}\|_2 \leq 1$.

Quiz 2.2.4. *A qubit density matrix with Bloch vector $v = (0.8, 0, 0.8)$ is*

a) A pure state

b) A mixed state

c) Not a valid quantum state

Quiz 2.2.5. *The matrix $\rho = \frac{1}{2}\mathbb{I}$ is*

a) A pure state

b) A mixed state

c) Not a valid quantum state

2.2.5 Combining density matrices

Week 1, Lecture 1.4,Lecture 1: Combining multiple qubits - tensor product of density matrices

Suppose we are given two quantum systems A and B, described by density matrices ρ_A and ρ_B respectively. How should their joint state ρ_{AB} be defined? In the previous chapter we saw that two pure quantum states $|v_1\rangle \in \mathbb{C}^{d_1}, |v_2\rangle \in \mathbb{C}^{d_2}$ respectively can be combined by taking their tensor product $|v_1\rangle \otimes |v_2\rangle \in \mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$. It turns out that the rule for mixed states is very similar.

Let us start with the simple case where ρ_A, ρ_B are 2×2-dimensional matrices. Then

$$\rho_A \otimes \rho_B = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \otimes \begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} = \begin{pmatrix} m_{11}\begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} & m_{12}\begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} \\ m_{21}\begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} & m_{22}\begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} \end{pmatrix}$$
$$= \begin{pmatrix} m_{11}n_{11} & m_{11}n_{12} & m_{12}n_{11} & m_{12}n_{12} \\ m_{11}n_{21} & m_{11}n_{22} & m_{12}n_{21} & m_{12}n_{22} \\ m_{21}n_{11} & m_{21}n_{12} & m_{22}n_{11} & m_{22}n_{12} \\ m_{21}n_{21} & m_{21}n_{22} & m_{22}n_{21} & m_{22}n_{22} \end{pmatrix}.$$

For example, if we have two density matrices $\rho_A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\rho_B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, then

$$\rho_{AB} = \rho_A \otimes \rho_B = \begin{pmatrix} 1 \cdot \rho_B & 0 \cdot \rho_B \\ 0 \cdot \rho_B & 0 \cdot \rho_B \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

This definition easily extends to larger matrices as follows.

Definition 2.2.6 (Tensor product)**.** *Consider any $d' \times d$ matrix ρ_A and $k' \times k$ matrix ρ_B,*

$$\rho_A = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1d} \\ m_{21} & \ddots & \ddots & m_{2d} \\ \vdots & \ddots & \ddots & \vdots \\ m_{d'1} & m_{d'2} & \cdots & m_{d'd} \end{pmatrix}, \qquad \rho_B = \begin{pmatrix} n_{11} & n_{12} & \cdots & n_{1k} \\ n_{21} & \ddots & \ddots & n_{2k} \\ \vdots & \ddots & \ddots & \vdots \\ n_{k'1} & n_{k'2} & \cdots & n_{k'k} \end{pmatrix}.$$

Their tensor product is given by the $d'k' \times dk$ matrix

$$\rho_{AB} = \rho_A \otimes \rho_B = \begin{pmatrix} m_{11}\rho_B & m_{12}\rho_B & \cdots & m_{1d}\rho_B \\ m_{21}\rho_B & \ddots & \ddots & m_{2d}\rho_B \\ \vdots & \ddots & \ddots & \vdots \\ m_{d'1}\rho_B & m_{d'2}\rho_B & \cdots & m_{d'd}\rho_B \end{pmatrix}.$$

As a word of caution, beware that the tensor product, as the usual matrix product, is non-commutative.

Example 2.2.5. *Consider the density matrices $\rho_A = \frac{1}{4}\begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}$ and $\rho_B = \frac{1}{2}\begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix}$.*

Then

$$\rho_A \otimes \rho_B = \frac{1}{8}\begin{pmatrix} 1 & -i & 1 & -i & 0 & 0 \\ i & 1 & i & 1 & 0 & 0 \\ 1 & -i & 2 & -2i & 1 & -i \\ i & 1 & 2i & 2 & i & 1 \\ 0 & 0 & 1 & -i & 1 & -i \\ 0 & 0 & i & 1 & i & 1 \end{pmatrix},$$

and

$$\rho_B \otimes \rho_A = \frac{1}{8}\begin{pmatrix} 1 & 1 & 0 & -i & -i & 0 \\ 1 & 2 & 1 & -i & -2i & -i \\ 0 & 1 & 1 & 0 & -i & -i \\ i & i & 0 & 1 & 1 & 0 \\ i & 2i & i & 1 & 2 & 1 \\ 0 & i & i & 0 & 1 & 1 \end{pmatrix} \neq \rho_A \otimes \rho_B \,.$$

■

Quiz 2.2.6. $\frac{1}{2}\left(\rho_A^1 + \rho_A^2\right) \otimes \rho_B = \frac{1}{2}\left(\rho_A^1 \otimes \rho_B + \rho_A^2 \otimes \rho_B\right)$ *for all* ρ_A^1, ρ_A^2 *and* ρ_B*. True or false?*

a) True
b) False

Quiz 2.2.7. $\rho_A \otimes \rho_B = \rho_B \otimes \rho_A$ *for all* ρ_A *and* ρ_B*. True or false?*

a) True
b) False

Quiz 2.2.8. *What is the tensor product* $\rho_{AB} = \rho_A \otimes \rho_B$ *of* $\rho_A = |1\rangle\langle 1|$ *and* $\rho_B = \frac{\mathbb{I}}{2}$*?*

a) $\rho_{AB} = \frac{1}{4}\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$

b) $\rho_{AB} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

c) $\rho_{AB} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$

2.2.6 Classical-quantum states

Week 1, Lecture 1.5,Lecture 1: Classical-quantum states

In quantum cryptography we frequently find ourselves in a situation in which the "honest parties" have some classical information X about which an "adversary" — such as an eavesdropper Eve — may hold quantum information Q. In other words, the quantum state Q is correlated with the classical information X. Since classical information is a special case of quantum information, the joint state of both X and Q can be represented by a density matrix ρ_{XQ}. How does such a density matrix look like?

Classical states

As a first step, let us pause to think about what it means for X to contain "classical information". In full generality, classical information can be modeled by a probability distribution over strings of bits x. Here x denotes the information and p_x the probability that this is the information contained in X. Suppose then that we are given a probability distribution over symbols x taken from the alphabet $\mathcal{X} = \{0, \ldots, d-1\}$, and let p_x denote the probability of symbol x. Identifying each possible value in $\mathcal{X}$ with an element of the standard basis $\{|0\rangle, \ldots, |d-1\rangle\}$ we can describe a system that is initialized in state $|x\rangle$ with probability p_x using the density matrix

$$\rho = \sum_{x=0}^{d-1} p_x |x\rangle\langle x| \ .$$

Note that ρ is a matrix which has the probabilities p_x on the diagonal and has all other entries equal to zero. As such, ρ is just another way to represent the distribution p_x: instead of a sequence of numbers, or a vector, we wrote the numbers on the diagonal of a matrix. Moreover, you can verify that measuring ρ in the standard basis results in outcome "x" with probability precisely p_x. In this sense, ρ is an accurate representation of the system X described above.

Definition 2.2.7 (Classical state)**.** *Let $\{|x\rangle\}_{x=0}^{d-1}$ denote the standard basis for $\mathbb{C}^d$. A system X is said to be in a classical state, or* c-state*, if its density matrix ρ_X is diagonal in the standard basis, i.e. ρ_X has the form*

$$\rho = \sum_{x=0}^{d-1} p_x |x\rangle\langle x| \ ,$$

where $\{p_x\}_{x=0}^{d-1}$ is a probability distribution.

Thus from now on we equate "classical state" or "classical density matrix" with "diagonal in the standard basis". The choice of the standard basis is arbitrary, as from a mathematical point of view all orthonormal bases are equivalent. Nevertheless, it is an important convention and serves as a point of connection between the classical and quantum worlds.

Classical-quantum states

Now, let's move to states which are partially classical and partially quantum. Let's start with an example. Suppose that with probability $1/2$ system X is in the classical state $|0\rangle$ and system Q is in the mixed state $\frac{\mathbb{I}}{2}$, and with probability $1/2$ system X is in the classical state $|1\rangle$ and system Q is in the pure state $|+\rangle$. How do we write down the density matrix of the joint system XQ? In the first case, the density matrix is $|0\rangle\langle 0|_X \otimes (\mathbb{I}/2)_Q$, and in the second it is $|1\rangle\langle 1|_X \otimes |+\rangle\langle +|_Q$. Since both probabilities are equal to $\frac{1}{2}$, overall we obtain

$$\rho_{XQ} = \frac{1}{2}|0\rangle\langle 0|_X \otimes \frac{\mathbb{I}_Q}{2} + \frac{1}{2}|1\rangle\langle 1|_X \otimes |+\rangle\langle +|_Q \, .$$

Check for yourself that ρ_{XQ} is a valid density matrix (remember the two conditions that need to be verified). This kind of density matrix is called a *classical-quantum state*, or cq-state for short. The reason is that the X part of the state is classical. More generally we give the following definition.

Definition 2.2.8. *A classical-quantum state, or simply* cq-state*, is a state of two subsystems, X and Q, such that its density matrix has the form*

$$\rho_{XQ} = \sum_x p_x |x\rangle\langle x|_X \otimes \rho_x^Q \, ,$$

where $\{p_x\}$ is a probability distribution and for every x, $|x\rangle$ designates the standard basis state on X and ρ_x^Q is an arbitrary density matrix on Q.

In applications to cryptography x will often represent some (partially secret) classical string that Alice creates during a quantum protocol, and ρ_x^Q some quantum information that an eavesdropper may have gathered during the protocol and which may be correlated with the string x. By convention we will usually reserve the letters X, Y, Z to denote classical registers, and use the other letters for quantum information. (More letters for quantum!)

Quiz 2.2.9. *Which of the following states is in general a classical-quantum state?*

a) $\rho_{AB} = \frac{1}{2}\left(\rho_A^0 \otimes \rho_B^0 + \rho_A^1 \otimes \rho_B^1\right)$
b) $\rho_{AB} = \frac{1}{2}\left(\rho_A^0 \otimes \rho_B^0 + |0\rangle\langle 0|_A \otimes |1\rangle\langle 1|_B\right)$
c) $\rho_{AB} = \frac{1}{2}\left(|0\rangle\langle 0|_A \otimes \rho_B^0 + |1\rangle\langle 1|_A \otimes \rho_B^1\right)$

Quiz 2.2.10. *Alice prepares uniformly at random (each with probability $p_i = 1/3$) one out of three quantum states ρ_B^i, where $i \in \{0, 1, 2\}$, and sends this state to Bob. After preparation, the information about the state she prepared becomes encoded in a classical memory $|i\rangle\langle i|_A$ that Alice keeps. What is the correct description of the joint state that Alice and Bob share?*

a) $\rho_{AB} = \frac{1}{3}\left(|0\rangle\langle 0|_A \otimes \rho_B^0 + |1\rangle\langle 1|_A \otimes \rho_B^1 + |2\rangle\langle 2|_A \otimes \rho_B^2\right)$
b) $\rho_{AB} = \frac{1}{9}\left(|0\rangle\langle 0|_A + |1\rangle\langle 1|_A + |2\rangle\langle 2|_A\right) \otimes \left(\rho_B^0 + \rho_B^1 + \rho_B^2\right)$
c) $\rho_{AB} = \frac{1}{2}\left(|0\rangle\langle 0|_A \otimes |1\rangle\langle 1|_A \otimes |2\rangle\langle 2|_A + \rho_B^0 \otimes \rho_B^1 \otimes \rho_B^2\right)$

2.3 General measurements

Week 1, Lecture 1.6,Lecture 1: General measurements

So far we have described how to measure a quantum state, pure or mixed, in a given orthonormal basis. Quantum mechanics allows a much more refined notion of measurement, which plays an important role both in quantum information theory and in cryptography. Indeed, in quantum information theory certain tasks, such as the task of discriminating between multiple states, can be solved more efficiently using these generalized measurements. Moreover, taking an adversarial viewpoint, in quantum cryptography it is essential to prove security for the most general kind of attack, including all measurements that an attacker could possibly make! This includes using extra qubits to make measurements, which is effectively how such generalised measurements can be realized.

2.3.1 POVMs

If we are only interested in computing the probabilities of measurement outcomes – but do not require a complete specification of what happens to the quantum state once the measurement has been performed – then the most general kind of measurement that is allowed in quantum mechanics can be described mathematically by a positive operator-valued measure, or POVM for short.

Definition 2.3.1 (POVM)**.** *A POVM on $\mathbb{C}^d$ is a set of positive semidefinite matrices $\{M_x\}_{x\in\mathcal{X}}$ such that*

$$\sum_x M_x = \mathbb{I}_d \,.$$

The subscript x is used as a label for the measurement outcome.

Quiz 2.3.1. *Which of the following is a valid POVM?*

I. $\left\{\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}, \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}\right\}$

II. $\left\{\begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{3} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}\right\}$

III. $\left\{\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}\right\}$

a) I. and II.
b) I. and III.
c) only I.
d) only III.

Having generalized our notion of measurement, we need to extend the measurement

rule, i.e. the rule that specifies the probability of obtaining each possible outcome when performing the measurement on a given state with density matrix ρ.

Definition 2.3.2 (Generalized measurement rule)**.** *Let $\{M_x\}$ be a POVM. Then the probability p_x of observing outcome x when performing the measurement $\{M_x\}$ on a density matrix ρ is*

$$p_x = \mathrm{tr}(M_x \rho) \ .$$

This expression is sometimes called the Born rule.

The next two examples show that the generalized Born rule is compatible with the measurement rule we had introduced before.

Example 2.3.1. *Consider a probability distribution (p_x) and the associated classical mixture $\rho = \sum_x p_x |x\rangle\langle x|$. If we measure ρ in the standard basis, with associated POVM $M_x = |x\rangle\langle x|$ as in Example 2.3.2, we obtain outcome x with probability*

$$\mathrm{tr}(|x\rangle\langle x|\rho) = \langle x| \rho |x\rangle = p_x \ ,$$

as expected: ρ indeed captures the classical distribution given by the probabilities p_x. ∎

Example 2.3.2. *Recall that when measuring a state $|\psi\rangle = \sum_x \alpha_x |x\rangle$ in a basis such as $\{|x\rangle\}_x$, the probability of observing outcome x is given by $|\alpha_x|^2$. Let us verify that this rule is recovered as a special case of the POVM formalism. For each x let $M_x = |x\rangle \langle x|$, so that M_x is positive semidefinite (in fact, it is a projector, i.e. $M_x^2 = M_x$) and $\sum_x M_x = \mathbb{I}$ (this can be verified by using that $\{|x\rangle\}$ is a basis), as required. Let $\rho = |\psi\rangle\langle\psi|$. We can use the Born rule to compute*

$$\begin{aligned} p_x &= \mathrm{tr}(M_x \rho) \\ &= \mathrm{tr}(|x\rangle\langle x|\rho) \\ &= \langle x| \rho |x\rangle \\ &= \sum_{x',x''} \alpha_{x'} \alpha_{x''}^* \langle x| x'\rangle\langle x''| x\rangle \\ &= |\alpha_x|^2 \ . \end{aligned}$$

∎

Beyond the calculation of outcome probabilities it can be important to know what happens to a quantum state after a generalized measurement has been performed. For the case of measuring in a basis, we already know the answer: the state collapses to the basis element associated with the outcome of the measurement that is obtained.

In the case of a POVM it turns out that the information given by the operators $\{M_x\}$ is not sufficient to fully determine the post-measurement state. The reason for this is because such a measurement may not fully collapse the state, meaning that the post-measurement state may not be pure (this corresponds to the case where M_x has rank more than 1). Intuitively, if the measurement operator M_x does not have rank 1 there is some freedom

in choosing exactly where the post-measurement state lies without affecting the outcome probabilities.

2.3.2 Generalized measurements

The additional information needed to specify post-measurement states is a *Kraus operator representation* of the POVM.

Definition 2.3.3 (Kraus operators)**.** *Let $M = \{M_x\}$ be a POVM on $\mathbb{C}^d$. A* Kraus operator representation *of M is a set of $d' \times d$ matrices A_x such that $M_x = A_x^\dagger A_x$ for all x.*

For any positive semidefinite matrix N, if $N = \sum_i \lambda_i |v_i\rangle\langle v_i|$ is the spectral decomposition of N, then N has a unique positive semidefinite square root which is given by $\sqrt{N} = \sum_i \sqrt{\lambda_i} |v_i\rangle\langle v_i|$. Thus a Kraus decomposition of any POVM always exists by setting $A_x = \sqrt{M_x}$. In particular, if $M_x = |u_x\rangle\langle u_x|$ is a projector then $\sqrt{M_x} = M_x$ and we can take $A_x = M_x$. But for any unitary U_x on $\mathbb{C}^d$, $A'_x = U_x \sqrt{M_x}$ is also a valid decomposition. Hence, there is no unique Kraus representation for a given POVM. In fact, the definition even allows matrices A_x that are not square.

This means we cannot go from POVM to Kraus operators. However, given Kraus operators we can find the POVM. Thus the most general form to write down a quantum measurement is through the full set of Kraus operators $\{A_x\}_x$. Let's see how knowledge of the Kraus operators allows us to compute post-measurement states.

Definition 2.3.4 (Post-measurement state)**.** *Let ρ be a density matrix and $M = \{M_x\}_x$ a POVM with Kraus decomposition given by operators $\{A_x\}_x$. Suppose the measurement is performed on a density matrix ρ, and the outcome x is obtained. Then the state of the system after the measurement, conditioned on having obtained the outcome x, is*

$$\rho_{|x} = \frac{A_x \rho A_x^\dagger}{\mathrm{tr}(A_x^\dagger A_x \rho)} \, .$$

If $\mathrm{tr}(A_x^\dagger A_x \rho) = 0$ then the formula for $\rho_{|x}$ is meaningless. However, in that case the outcome x has probability 0 of occurring and so there is no need to define an associated post-measurement state.

You may want to convince yourself that when measuring a pure state $|\psi\rangle$ in an arbitrary orthonormal basis, with Kraus decomposition $A_x = \sqrt{M_x} = |x\rangle\langle x|$, the post-measurement state as defined above is precisely the basis state associated to the measurement outcome.

An important class of generalized measurements is given by the case where the M_x are projectors onto orthogonal subspaces (not necessarily of rank 1).

Definition 2.3.5. *A* projective measurement*, also called a* von Neumann measurement*, is given by a set of orthogonal projectors $M_x = \Pi_x$ such that $\sum_x \Pi_x = \mathbb{I}$. For such a measurement, unless otherwise specified we will always use the default Kraus decomposition $A_x = \sqrt{M_x} = \Pi_x$. For such a measurement the probability q_x of observing measurement*

outcome x can be expressed as

$$q_x = \mathrm{tr}(\Pi_x \rho),$$

and the post-measurement states are

$$\rho_{|x} = \frac{\Pi_x \rho \Pi_x}{\mathrm{tr}(\Pi_x \rho)} \, .$$

The following example shows how to use the formalism of generalized measurements to perform a certain task in different ways.

Example 2.3.3. *Suppose given a two-qubit state ρ, such that we would like to measure the parity (in the standard basis) of the two qubits. A first way to do this would be to measure ρ in the standard basis, obtain two bits, and take their parity. In this case the probability of obtaining the outcome "even" would be*

$$q_{\mathrm{even}} = \langle 00| \, \rho \, |00\rangle + \langle 11| \, \rho \, |11\rangle \, ,$$

and the post-measurement state would be the mixture of the two post-measurement states associated with outcomes $(0,0)$ and $(1,1)$, so

$$\rho_{|\mathrm{even}} = \frac{1}{q_{\mathrm{even}}} \big(\langle 00| \, \rho \, |00\rangle \big) \, |00\rangle \langle 00| + \big(\langle 11| \, \rho \, |11\rangle \big) \, |11\rangle \langle 11| \, .$$

Now suppose that we attempt to measure the parity using a generalized measurement which directly projects onto the relevant subspaces, without measuring the qubits individually. That is, consider the projective measurement $\Pi_{\mathrm{even}} = |00\rangle\langle 00| + |11\rangle\langle 11|$ and $\Pi_{\mathrm{odd}} = \mathbb{I} - \Pi_{\mathrm{even}} = |01\rangle\langle 01| + |10\rangle\langle 10|$. With this measurement the probability of obtaining the outcome "even" is

$$q'_{\mathrm{even}} = \mathrm{tr}(\Pi_{\mathrm{even}} \rho) = \langle 00| \, \rho \, |00\rangle + \langle 11| \, \rho \, |11\rangle \ \, ,$$

as before. However, the post-measurement state is now

$$\rho'_{|\mathrm{even}} = \frac{1}{q'_{\mathrm{even}}} \Pi_{\mathrm{even}} \rho \Pi_{\mathrm{even}} \, .$$

To see the difference, consider the state $\rho = |\mathrm{EPR}\rangle \langle \mathrm{EPR}|$ where $|\mathrm{EPR}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Then clearly the parity measurement should report the outcome "even" with probability 1, and you can check that this is the case for both measurements. However, the post-measurement states are different. In the first case,

$$\rho_{|\mathrm{even}} = \frac{1}{2} \, |00\rangle \langle 00| + \frac{1}{2} \, |11\rangle \langle 11| \, ,$$

while in the second case,

$$\rho'_{|\mathrm{even}} = |\mathrm{EPR}\rangle \langle \mathrm{EPR}|$$

is unchanged! This is one of the key advantages of using generalized measurements, as opposed to basis measurements: they allow to compute certain simple quantities on multi-qubit states (such as the parity) without fully "destroying" the state. ∎

Exercise 2.3.1 Use a projective measurement to measure the parity, *in the Hadamard basis*, of the state $|00\rangle\langle 00|$.

Exercise 2.3.2 For the same scenario as the previous exercise, compute the probabilities of obtaining measurement outcomes "even" and "odd", and the resulting post-measurement states. What would the post-measurement states have been if you had first measured the qubits individually in the Hadamard basis, and then taken the parity?

2.4 The partial trace

Week 1, Lecture 1.7,Lecture 1: The partial trace
Week 1, Lecture 1.7,Lecture 2: Another way to compute the partial trace

Going back to our second motivation for introducing density matrices, let us now give an answer to the following question: given a multi-qubit state, how do we write down the "partial state" associated to a subset of the qubits? More generally, suppose ρ_{AB} is a density matrix on a tensor product space $\mathbb{C}_A^{d_A} \otimes \mathbb{C}_B^{d_B}$, and suppose that Alice holds the part of ρ corresponding to system A and Bob holds the part corresponding to system B. How do we represent the state ρ_A of Alice's system?

2.4.1 An operational viewpoint

The operation that takes us from ρ_{AB} to ρ_A is called the *partial trace*. It can be specified in purely mathematical terms, and we do so in the next section. Before we do that, let us try to think about the problem from an operational point of view. First, consider an easy case: if $\rho_{AB} = \rho_A \otimes \rho_B$, where ρ_A and ρ_B are both density matrices, then clearly Alice's system is defined by ρ_A. In this case, we would say that the partial trace of ρ_{AB}, when "tracing out" system B, is the density matrix ρ_A.

A slightly more complicated case is when

$$\rho_{AB} = \sum_i p_i \, \rho_i^A \otimes \rho_i^B \tag{2.6}$$

is a mixture of tensor products (we will later see that this is called a "separable state"). Using the interpretation that this represents a state that is in state $\rho_i^A \otimes \rho_i^B$ with probability p_i, it would be natural to claim that Alice's share of the state is ρ_i^A with probability p_i, i.e. the partial trace of ρ_{AB}, when tracing out — i.e. ignoring — system B, is now $\rho_A = \sum_i p_i \rho_i^A$.

How about a general ρ_{AB}? Remember from Exercise **??** that there exists some ρ that do not have a decomposition of the form (2.6), such as for example the EPR pair. Our idea is to "force" such a decomposition by performing the following little thought experiment. Let us *imagine* that Bob performs a complete basis measurement on his system, using an arbitrary basis $\{|u_x\rangle\}_x$. Let us introduce a POVM on the joint system of Alice and Bob that models

this measurement: since Alice does nothing, we can set $M_x = \mathbb{I}_A \otimes |u_x\rangle\langle u_x|_B$, which you can check indeed defines a valid POVM. Moreover, this is a projective measurement, so we can take the Kraus operators $A_x = \sqrt{M_x} = M_x$. By definition the post-measurement states are given by

$$\rho_{|x}^{AB} = \frac{M_x \rho_{AB} M_x}{\mathrm{tr}\left(M_x \rho_{AB}\right)} = \frac{\left((\mathbb{I}_A \otimes \langle u_x|)\rho_{AB}(\mathbb{I}_A \otimes |u_x\rangle)\right)_A \otimes |u_x\rangle\langle u_x|_B}{\mathrm{tr}\left((\mathbb{I}_A \otimes |u_x\rangle\langle u_x|_B)\rho_{AB}\right)}.$$

Notice how we wrote the state as a tensor product of a state on A and one on B. Make sure you understand the notation in this formula, and that it specifies a well-defined state.

The key step is to realize that, whatever the state of Alice's system A is, it shouldn't depend on any operation that Bob performs on B. After all, it may be that A is here on Earth, and B is on Mars. Since quantum mechanics does not allow faster than light communication, as long as the two of them remain perfectly isolated, meaning that Alice doesn't get to learn the measurement that Bob performs or its outcome, then her state should remain unchanged. We can thus describe it as follows: "With probability $q_x = \mathrm{Tr}(M_x \rho_{AB})$, Alice's state is the A part of $\rho_{|x}^{AB}$". Using the rule for computing post-measurement states, we get

$$\rho_A = \sum_x q_x \frac{\left((\mathbb{I} \otimes \langle u_x|)\rho_{AB}(\mathbb{I} \otimes |u_x\rangle)\right)_A}{\mathrm{Tr}\left((\mathbb{I} \otimes |x\rangle\langle x|)\rho_{AB}\right)} = \sum_x (\mathbb{I} \otimes \langle u_x|)\rho_{AB}(\mathbb{I} \otimes |u_x\rangle). \tag{2.7}$$

Although we derived the above expression for Alice's state using sensible arguments, there is something you should be worried about: doesn't it depend on the choice of basis $\{|u_x\rangle\}_x$ we made for Bob's measurement? Of course, it should not, as our entire argument is based on the idea that Alice's reduced state should not depend on any operation performed by Bob. The next exercise asks you to verify that this is indeed the case. (We emphasize that this is only the case as long as Alice doesn't learn the measurement outcome! If we fix a particular outcome x then it's a completely different story. Beware of this subtlety, it will repeatedly come up throughout the book.)

Exercise 2.4.1 Verify that the state ρ_A defined in Eq.(2.7) does not depend on the choice of basis $\{|u_x\rangle\}$. *[Hint: first argue that if two density matrices ρ, σ satisfy $\langle\phi|\,\rho\,|\phi\rangle = \langle\phi|\,\sigma\,|\phi\rangle$ for all unit vectors $|\phi\rangle$ then $\rho = \sigma$. Then compute $\langle\phi|\,\rho_A\,|\phi\rangle$, and use the POVM condition $\sum_x M_x = \mathbb{I}$ to check that you can get an expression independent of the $\{|u_x\rangle\}$. Conclude that ρ_A itself does not depend on $\{|u_x\rangle\}$.]*

Example 2.4.1. *Consider the example of the EPR pair*

$$|\mathrm{EPR}\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)\ . \tag{2.8}$$

Writing this as a density operator we have

$$\rho_{AB} = |\mathrm{EPR}\rangle\langle\mathrm{EPR}|_{AB} = \frac{1}{2}\left(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|\right)\ . \tag{2.9}$$

Let's measure system B in the standard basis: taking A into account we consider the

POVM $M_0 = \mathbb{I}_A \otimes |0\rangle\langle 0|_B$ *and* $M_1 = \mathbb{I}_A \otimes |1\rangle\langle 1|_B$. *We can then compute*

$$\begin{aligned} q_0 &= \mathrm{Tr}(M_0\rho) \\ &= \frac{1}{2}\,\mathrm{Tr}\left((\mathbb{I}\otimes|0\rangle\langle 0|)(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|)\right) \\ &= \frac{1}{2}\big(1+0+0+0\big) = \frac{1}{2}\,, \end{aligned}$$

and similarly $q_1 = 1/2$. *The post-measurement stated on* A *is then*

$$\rho^A_{|0} = \frac{1}{2}(\mathbb{I}\otimes\langle 0|)\rho_{AB}(\mathbb{I}\otimes|0\rangle) + \frac{1}{2}(\mathbb{I}\otimes\langle 1|)\rho_{AB}(\mathbb{I}\otimes|1\rangle) = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|\,.$$

Now do the same calculation using a measurement in the Hadamard basis on B, *and check that you get the same result!* ∎

Quiz 2.4.1. *Suppose that Alice and Bob share the state* $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. *Bob measures his qubit in the basis* $\{|+\rangle, |-\rangle\}$ *and obtains* $|+\rangle$. *What is the post-measurement state of Alice's qubit?*

a) $|-\rangle$
b) $|0\rangle$
c) $|+\rangle$

Quiz 2.4.2. *Suppose instead that Alice and Bob share the state* $\frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)$. *Bob again measures his qubit in the basis* $\{|+\rangle, |-\rangle\}$ *and obtains* $|+\rangle$. *What is the post-measurement state of Alice's qubit?*

a) $|-\rangle$
b) $|0\rangle$
c) $-|+\rangle$

2.4.2 A mathematical definition

Armed with our "operational" definition of what the partial trace *should* achieve, we now give the precise, mathematical definition of this operation.

Definition 2.4.1 (Partial Trace)**.** *Consider a general matrix*

$$M_{AB} = \sum_{ijk\ell} \gamma_{ij}^{k\ell} |i\rangle\langle j|_A \otimes |k\rangle\langle \ell|_B\,, \tag{2.10}$$

where $|i\rangle_A, |j\rangle_A$ *and* $|k\rangle_B, |\ell\rangle_B$ *run over orthonormal bases of* A *and* B *respectively.*

Then the partial trace over B is defined as

$$\begin{aligned} M_A &= \mathrm{tr}_B(M_{AB}) \\ &= \sum_{ijk\ell} \gamma_{ij}^{k\ell} |i\rangle\langle j|_A \otimes \mathrm{tr}(|k\rangle\langle \ell|_B) \\ &= \sum_{ijk\ell} \gamma_{ij}^{k\ell} |i\rangle\langle j|_A \otimes \langle \ell|\, k\rangle_B) \\ &= \sum_{ijk\ell} \gamma_{ij}^{k\ell} |i\rangle\langle j|_A \otimes \delta_{k\ell} \\ &= \sum_{ij} \left(\sum_k \gamma_{ij}^{kk} \right) |i\rangle\langle j|_A \, . \end{aligned}$$

Similarly, the partial trace over A is

$$M_B = \mathrm{tr}_A(M_{AB}) = \sum_{ijk\ell} \gamma_{ij}^{k\ell} \, \mathrm{tr}(|i\rangle\langle j|) \otimes |k\rangle\langle \ell| = \sum_{k\ell} \left(\sum_j \gamma_{jj}^{k\ell} \right) |k\rangle\langle \ell| \, .$$

If $M_{AB} = \rho_{AB}$ is a density matrix, meaning that it is positive semidefinte and $\mathrm{tr}(M_{AB}) = 1$, then it is a good exercise to verify that the partial traces $\rho_A = \mathrm{tr}_B(\rho_{AB})$ and $\rho_B = \mathrm{tr}_A(\rho_{AB})$ are again density matrices. We refer to them as the *reduced states* of the system AB on system A and system B respectively.

The formal definition directly gives us a recipe for computing the partial trace of a state ρ_{AB}, as follows.

1 Write a decomposition of ρ_{AB} in the form (2.10). Note that you may do this for any choice of orthonormal bases that you like for systems A and B. The coefficients $\gamma_{ij}^{k\ell}$ are then simply the entries of the matrix ρ_{AB} at position $|i\rangle\langle j| \otimes |k\rangle\langle \ell|$.
2 Put a trace on the "B part" and use cyclicity of the trace to finish the computation.

In many cases, a decomposition of ρ as in (2.10) is easily found.

Example 2.4.2. *Let us consider again the example of the EPR pair*

$$|\mathrm{EPR}\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \, ,$$

with associated density matrix

$$\begin{aligned} \rho_{AB} &= |\mathrm{EPR}\rangle \langle \mathrm{EPR}|_{AB} \\ &= \frac{1}{2}\big(|0\rangle\langle 0|_A \otimes |0\rangle\langle 0|_B + |0\rangle\langle 1|_A \otimes |0\rangle\langle 1|_B + |1\rangle\langle 0|_A \otimes |1\rangle\langle 0|_B + |1\rangle\langle 1|_A \otimes |1\rangle\langle 1|_B\big). \end{aligned}$$

Using the definition we can compute

$$\begin{aligned} \mathrm{tr}_B(\rho_{AB}) = \frac{1}{2}\big(&|0\rangle\langle 0|_A \otimes \mathrm{tr}(|0\rangle\langle 0|_B) + |0\rangle\langle 1|_A \otimes \mathrm{tr}(|0\rangle\langle 1|_B) \\ &+ |1\rangle\langle 0|_A \otimes \mathrm{tr}(|1\rangle\langle 0|_B) + |1\rangle\langle 1|_A \otimes \mathrm{tr}(|1\rangle\langle 1|_B)\big) \, . \end{aligned}$$

Since the trace is cyclic, $\mathrm{tr}(|0\rangle\langle 1|) = \langle 1|0\rangle = 0$*, similarly* $\mathrm{tr}(|1\rangle\langle 0|) = 0$*, but* $\mathrm{tr}(|0\rangle\langle 0|) = \mathrm{tr}(|1\rangle\langle 1|) = 1$ *and hence*

$$\mathrm{tr}_B(\rho_{AB}) = \frac{1}{2}\left(|0\rangle\langle 0| + |1\rangle\langle 1|\right) = \frac{\mathbb{I}}{2}\,. \tag{2.11}$$

Convince yourself that when we take the partial trace operation over A*, and hence look at the state of just Bob's qubit we also get*

$$\mathrm{tr}_A(\rho_{AB}) = \frac{\mathbb{I}}{2}\,. \tag{2.12}$$

This is consistent with our calculations in Example 2.4.2. ■

Exercise 2.4.2 If $\rho_{AB} = |\Phi\rangle\langle\Phi|$ is the singlet $|\Phi\rangle = \left(|01\rangle - |10\rangle\right)/\sqrt{2}$, compute ρ_A and ρ_B.

Example 2.4.3. *The notion of partial trace allows us to verify that performing a unitary operation on* A *has no effect on the state of* B*, i.e., it does not change* ρ_B*.*

$$(U_A \otimes \mathbb{I}_B)\rho_{AB}(U_A \otimes \mathbb{I}_B)^\dagger = \sum_{ijk\ell} \gamma_{ij}^{k\ell} U_A|i\rangle\langle j|_A U_A^\dagger \otimes |k\rangle\langle\ell|_B. \tag{2.13}$$

Computing again the partial trace we have

$$\begin{aligned}
\mathrm{tr}_A(U_A \otimes \mathbb{I}_B \rho_{AB} U_A^\dagger \otimes \mathbb{I}_B) &= \sum_{ijk\ell} \gamma_{ij}^{k\ell}\, \mathrm{tr}(U_A|i\rangle\langle j|U_A^\dagger) \otimes |k\rangle\langle\ell| && (2.14)\\
&= \sum_{ijk\ell} \gamma_{ij}^{k\ell}\, \mathrm{tr}(|i\rangle\langle j|U_A^\dagger U_A) \otimes |k\rangle\langle\ell| && (2.15)\\
&= \sum_{ijk\ell} \gamma_{ij}^{k\ell}\, \mathrm{tr}(|i\rangle\langle j|) \otimes |k\rangle\langle\ell| && (2.16)\\
&= \sum_{k\ell} \left(\sum_j \gamma_{jj}^{k\ell}\right) |k\rangle\langle\ell| = \rho_B\,. && (2.17)
\end{aligned}$$

Can you convince yourself that performing a measurement on A *also has no effect on* B*?* ■

Quiz 2.4.3. *What are Alice and Bob's reduced states in the joint state*

$$\rho_{AB} = \begin{pmatrix} \frac{1}{4} & 0 & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & -\frac{1}{4} & 0 \\ 0 & -\frac{1}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & 0 & 0 & \frac{1}{4} \end{pmatrix}?$$

a) $\rho_A = \rho_B = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$

b) $\rho_A = \rho_B = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$

c) $\rho_A = \begin{pmatrix} \frac{3}{4} & 0 \\ 0 & \frac{1}{4} \end{pmatrix}, \rho_B = \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & \frac{3}{4} \end{pmatrix}$

Quiz 2.4.4. *Alice and Bob share a state ρ_{AB}. If Alice's reduced state is $\rho_A = |0\rangle\langle 0|$, we know that ρ_{AB} is ...*

a) pure
b) mixed
c) not enough information

Quiz 2.4.5. *Alice and Bob share a state ρ_{AB}. If Alice's reduced state is $\rho_A = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$, we know that ρ_{AB} is ...*

a) pure
b) mixed
c) not enough information

2.5 Secure message transmission

With all the math behind us we are ready to turn to our first serious cryptographic task — in fact, the most serious task of all: the secure transmission of messages. To set things up, imagine two protagonists, Alice and Bob. Alice and Bob would like to exchange classical messages between each other (e.g. they want to chat!). However, Alice and Bob are worried that the messages they exchange could be intercepted by a malicious eavesdropper, Eve. This is because, although Alice and Bob each trust that they have full control over their own secure laboratory (i.e. their bedroom), they really don't know what happens on the communication line, such as the airwaves for a cell phone conversation or the post office truck for a snail mail conversation. Alice and Bob's goal is to limit, and if possible reduce to zero, the useful information that Eve may be able to get: they want to make sure that even if Eve intercepts all the classical messages they exchange, these messages look like complete rubbish to Eve!

If you think about this setup, you will see that we are faced with a symmetry-breaking problem. This is because, if Alice wants to send a message to Bob but Eve can listen to all messages exchanged, then Eve receives everything that Bob does. So then, how can Bob understand what Alice wants to say to him, but somehow Eve has no information, even though she read the same message?

To break the symmetry we will make a crucial assumption. We will assume that Alice and Bob are in possession of a *secret key*, that is known to them but completely unknown to Eve. This key will be used to hide the messages that they exchange. Cryptosystems which make use of this assumption are called *private-key* cryptosystems.

For the time being, we will not justify our assumption about the key: this is just some secret key Alice and Bob have in common, a secret they may have agreed on a long time in the past, when they were in the same place and could whisper to each other's ears. In later chapters we will see how quantum information can be used to establish such a secret even when Alice and Bob are physically separated.

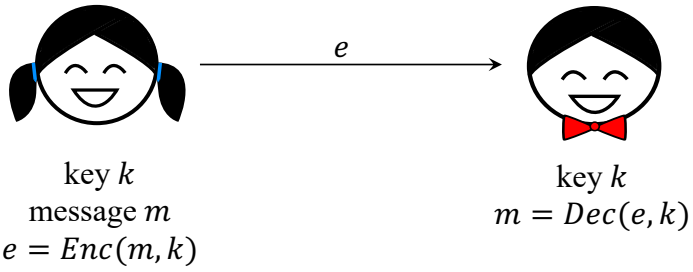

Fig. 2.2 Alice sends an encrypted message to Bob.

2.5.1 Shannon's secrecy condition and the need for large keys

The mathematical framework for the description of secret communication schemes was first developed by Claude Shannon in the 1940s, well before quantum information made its apparition. According to Shannon's formalism, an encryption scheme consists of two functions. The first is the *encryption function* $\mathrm{Enc}(k, m) = e$, that takes the key k and the message m and maps it to some encrypted message e. The original message m is often called the *plaintext*, and e the *ciphertext*. The second function is the *decryption function* $\mathrm{Dec}(k, e) = m$, that takes the key k and the ciphertext e back to the plaintext (2.2).

Definition 2.5.1. *An encryption scheme* $(\mathrm{Enc}, \mathrm{Dec})$ *is called* correct *if for every key* k *and every plaintext* m, $\mathrm{Dec}(k, \mathrm{Enc}(k, m)) = m$. *It is called* perfectly secure *if for any distribution* $p(\cdot)$ *over the space* $\mathcal{M}$ *of plaintexts the following two distributions on plaintexts are identical:*

1 *Generate a random plaintext* $m \in \mathcal{M}$ *with probability* $p(m)$.
2 *Select an arbitrary ciphertext* e. *Generate a uniformly random key* $k \in K$. *Generate a random plaintext* $m \in \mathcal{M}$ *with probability* $p(m|\mathrm{Enc}(k, m) = e)$.

In the definition of perfectly secure the key k is chosen uniformly at random. This is an important condition. It expresses our assumption that Eve has no information whatsoever about the key. So from her point of view every possible key has the same a priori probability: for every k in the key space K, it holds that $p_k = 1/|K|$.

The definition may be a little hard to understand the first time that you read it. So let's paraphrase using words. We call an encryption scheme perfectly secure whenever an eavesdropper Eve ignorant of the key does not gain any additional information about a plaintext message m from its encryption e. In other words, the probability $p(m)$ of the message m is the same a priori (as anyone could guess) as it is from the point of view of Eve, who has

obtained e. Observe that this is a very strong notion of security: absolutely no information is gained from having access to e!

This definition is so strong that it may even seem impossible to realize: if e has "no information" about m, then how can e be decrypted to recover m? As we will soon see, there is no contradiction: it is possible that e has no information at all about m *from the point of view of an Eavesdropper who does not know the secret key* k, yet e still has full information about m from the point of view of a honest party Bob *who does know the secret key*. This is a very subtle point: make sure you fully understand the distinction.

Note that it would be easy to come up with an encryption scheme which is "just" secret: Alice simply sends a randomly chosen e to Bob. Then, because e is random and independent of any message, of course learning e does not reveal information about the message. But clearly this scheme would not be correct: Bob cannot recover Alice's message. Similarly, it is easy to devise a scheme which is "just" correct: Alice sends $e = m$ to Bob. Clearly this is not secure since Even also learns m. In summary, the art of encryption is to design schemes that are both correct *and* secure.

In our presentation we assumed that Alice and Bob share a secret key k, and we informally argued that such a key was needed to "break the symmetry" between Bob and the eavesdropper Eve. Is this argument watertight — is a key really needed? As it turns out, not only it is needed but in fact the number of possible keys needs to be as large as the number of possible messages that Alice may wish to send. The following theorem, due to Shannon, proves this.

Theorem 2.5.1. *An encryption scheme* $(\mathrm{Enc}, \mathrm{Dec})$ *can only be* perfectly secure *and* correct *if the number of possible keys* $|K|$ *is at least as large as the number of possible messages* $|M|$*, that is,* $|K| \geq |M|$.

Proof Suppose for contradiction that there exists a correct scheme using fewer keys, i.e., $|K| < |M|$. We will show that such a scheme cannot be perfectly secure. Let p be the uniform distribution over M. Consider an eavesdropper who has intercepted the ciphertext e. She can compute

$$\mathcal{S} = \{\hat{m} \mid \exists k, \hat{m} = \mathrm{Dec}(k, e)\} \ , \tag{2.18}$$

that is, the set of all messages $\hat{m}$ for which there exists a key k that could have resulted in the observed ciphertext e. Note that the size $|\mathcal{S}|$ of this set is $|\mathcal{S}| \leq |K|$, since for each possible key k we get at most one message $\hat{m}$. Since $|K| < |M|$, we thus have $|\mathcal{S}| < |M|$. This means that there exists at least one message m such that $m \notin \mathcal{S}$, and hence $p(m|e) = 0$. However, by definition $p(m) = 1/|M|$. This contradicts the definition of perfect security given in Definition 2.5.1. □

Can the bound given in the lemma be achieved: does there exist an encryption scheme that is both correct *and* secure, and which uses precisely the minimal number of keys $|K| = |M|$? The answer is yes! We construct such a scheme in the next section.

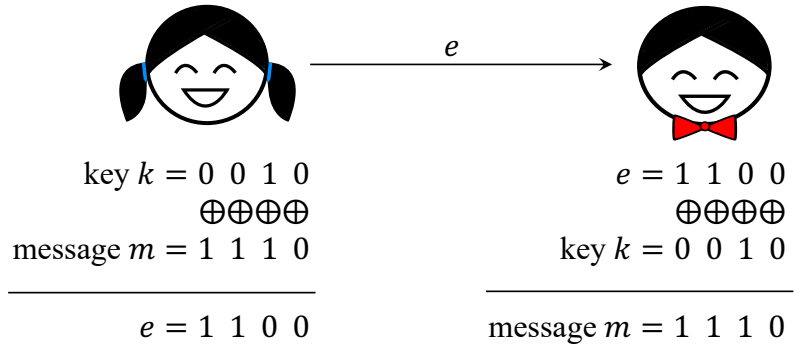

Fig. 2.3 An example of a one-time pad between Alice and Bob.

2.5.2 The (quantum) one-time pad

Week 1, Lecture 1.1,Lecture 1: The one-time pad
Week 1, Lecture 1.3,Lecture 1: Encrypting qubits with the quantum one-time pad

The *one-time pad* is arguably the simplest, yet also the most secure, encryption scheme known. We start with the "classical" version, that allows encryption of classical messages.

The classical one-time pad

Imagine that Alice (the sender) wants to send a secret message m to Bob (the receiver). For simplicity, we take the message space M to be the set of all n-bit strings: $M = \{0,1\}^n$. Let us furthermore assume that Alice and Bob already share a key $k \in \{0,1\}^n$ which is just as long as the message, and is uniformly random from the point of view of the adversary Eve. In the following definition, we use the notation $a \oplus b$ for the bitwise XOR, or equivalently addition modulo 2: for $a, b \in \{0,1\}$, $a \oplus b = a + b \bmod 2$.

Protocol 1. *The classical* one-time *pad is an encryption scheme in which the encryption of a message $m \in \{0,1\}^n$ using the key $k \in \{0,1\}^n$ is given by*

$$\mathrm{Enc}(k,m) = m \oplus k = (m_1 \oplus k_1, m_2 \oplus k_2, \ldots, m_n \oplus k_n) = (e_1, \ldots, e_n) = e\,.$$

The decryption is given by

$$\mathrm{Dec}(k,e) = e \oplus k = (e_1 \oplus k_1, e_2 \oplus k_2, \ldots, e_n \oplus k_n).$$

2.3 shows an example of the one-time pad. Note that since for any $j \in \{1,\ldots,n\}$, $m_j \oplus k_j \oplus k_j = m_j$, the scheme is correct. Is it secure?

To see that it satisfies Shannon's definition, consider any distribution p on M. For a

uniformly random choice of key k and a fixed message m, the associated ciphertext $e = \mathrm{Enc}(k, m)$ is uniformly distributed over all n-bit strings: for any e,

$$p(\mathrm{Enc}(k,m) = e|m) = p(m \oplus k = e|m) = p(k = e \oplus m|m) = \frac{1}{2^n} ,$$

since k is chosen uniformly at random. Since this holds for any message m,

$$p(e) = \sum_m p(m)p(e|m) = \frac{1}{2^n} .$$

Applying Bayes' rule we get that

$$p(m|e) = \frac{p(m,e)}{p(e)} = \frac{p(e|m)p(m)}{p(e)} = p(m) ,$$

independent of m. Thus $p(m|e) = p(m)$ and the scheme is perfectly secure.

Note that our argument crucially relies on the key being uniformly distributed and independent from the eavesdropper, a condition that has to be treated with care! In Chapter 6 we will introduce a method called *privacy amplification* that can be used to "improve" the quality of a key about which the eavesdropper may have partial information.

Remark 2.5.2. *While the one-time pad is "perfectly secure" according to Shannon's definition, it does not protect against an adversary changing bits in the messages exchanged between Alice and Bob. Indeed, you can verify that for any key k, and any string x, $\mathrm{Enc}(m \oplus x, k) = \mathrm{Enc}(m, k) \oplus x$. What this means is that flipping bits of the ciphertext is equivalent to flipping bits of the plaintext, and there is no way for Bob to detect if such an operation has taken place. This would be an issue for bank transactions, since an adversary could flip the transaction amount in an arbitrary way (without learning any information about the amount itself!). For this reason, one-time pads are generally supplemented by checksums or message authentication codes which allow changes to be detected (and corrected). These are well-known classical techniques, and we will not get into them in more detail here.*

Quiz 2.5.1. *Bob received from Alice a message encoded using the one-time pad: $e = 0010111$. Bob has the key needed to decrypt the message: $k = 1001011$. What is the message that Alice sent him?*

a) 1001011
b) 0010111
c) 1011100
d) 1011111

There is another way to look at the classical one time pad that brings it much closer to the quantum version we will consider next. Consider the encryption of a single-bit message $m \in \{0, 1\}$. Recall that we can represent this message as a pure quantum state $|m\rangle$, or equivalently as the density matrix $|m\rangle\langle m|$. When we apply the XOR operation the result is that the bit m is flipped whenever the key bit $k = 1$, and unchanged if $k = 0$. That is, when $k = 1$ the state is transformed as $|m\rangle \mapsto X|m\rangle$, where recall that X is the Pauli bit-flip

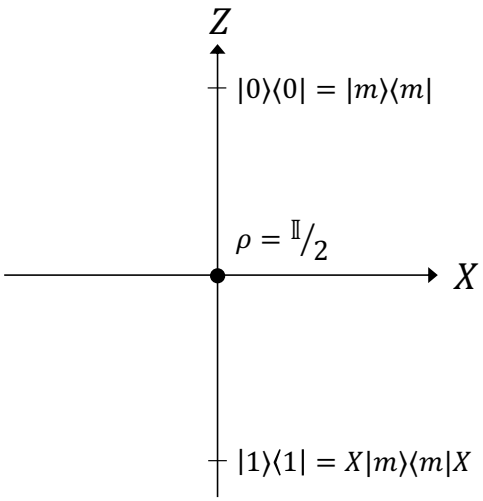

Fig. 2.4 Classical one-time pad in the xz-plane of the Bloch sphere for $m = 0$.

matrix. Thus in this case encryption implements the transformation $|m\rangle\langle m| \mapsto X|m\rangle\langle m|X$ visualized in 2.4.

If Alice and Bob choose a uniformly random key bit k then we can write the density matrix for the entire system KM, where K contains the key and M the message, as

$$\rho_{KM} = \frac{1}{2}|0\rangle\langle 0|_K \otimes |m\rangle\langle m|_M + \frac{1}{2}|1\rangle\langle 1|_K \otimes X|m\rangle\langle m|_M X \ .$$

From the point of view of Eve, who does not have access to the system K containing the key, the state of the message is represented by the density matrix

$$\rho_M = \frac{1}{2}|m\rangle\langle m|_M + \frac{1}{2}X|m\rangle\langle m|_M X = \frac{\mathbb{I}}{2} \ .$$

Note that ρ_M does *not* depend on m! Whatever m is, we get that $\rho_M = \frac{\mathbb{I}}{2}$. Since all information that can be gained from receiving the encrypted message is captured in the density matrix ρ_M, it follows that absolutely no information about m can be gained from intercepting the encryption.

The quantum one-time pad

We are finally ready for our first element of quantum cryptography: the quantum one-time pad! Let us consider the task of encrypting a qubit, instead of a classical bit (see 2.5). As a first attempt we might try to use the classical one-time pad, and see if it works for qubits as well. Does it? Well, the qubits $|0\rangle$ and $|1\rangle$ are encrypted just like classical messages are. But what about qubit in state $|+\rangle$? Because $X|+\rangle = |+\rangle$, in this case whatever key Alice and Bob share, the qubit is "encrypted" to itself. This is certainly not secure!

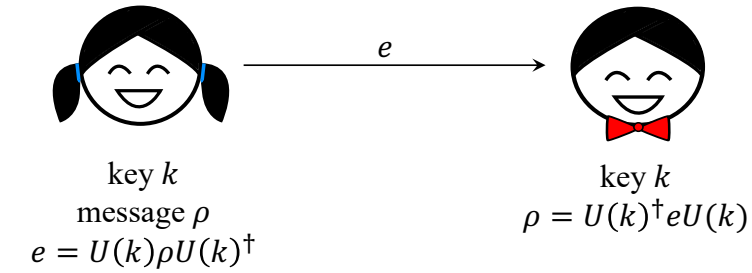

Fig. 2.5 General form of a quantum one-time pad. Alice encrypts the message qubit ρ with key k by applying unitary $U(k)$. Bob decrypts by undoing the unitary according to the key k.

The difficulty is that a quantum encryption scheme should hide information in all possible bases the qubit could be encoded in. In the classical case applying the bit flip operator X allowed us to encrypt any bit expressed in the standard basis. If we are allowed other bases, we should also encrypt a bit encoded in the Hadamard basis. This could be done by applying a Z instead of an X, because $Z|+\rangle = |-\rangle$ and vice-versa.

But what about other bases, what operation do we need to apply to encrypt information encoded in them? And how do we combine all these operations so that the same encryption scheme works for *all* qubits?

At this point it may seem miraculous that quantum encryption is at all possible using only a finite amount of key! But it is possible, and in face all we need are *two* bits of key, for every qubit.

Amazingly, it is in fact enough to handle both the standard and the Hadamard bases, and all other bases will follow. Let's see how this works. To flip in both bases, we apply the unitary operator $X^{k_1}Z^{k_2}$, where $k_1, k_2 \in \{0,1\}$ are two key bits chosen uniformly at random. With this choice of encryption operation, an arbitrary single-qubit ρ is transformed as

$$\rho \mapsto \frac{1}{4}\sum_{k_1,k_2\in\{0,1\}} X^{k_1}Z^{k_2}\rho Z^{k_2}X^{k_1}\ . \tag{2.19}$$

Now let's verify that this securely encrypts *any* single-qubit density matrix ρ. For this, remember the Bloch sphere representation of ρ. Remember also the fact that the Pauli matrices pairwise anti-commute. Using this we can make a small calculation,

$$\begin{aligned}\frac{1}{4}\big(X+XXX+ZXZ+XZXZX\big) &= \frac{1}{4}\big(X+X-ZZX-XZZXX\big)\\ &= \frac{1}{4}\big(X+X-X-X\big)\\ &= 0\ ,\end{aligned}$$

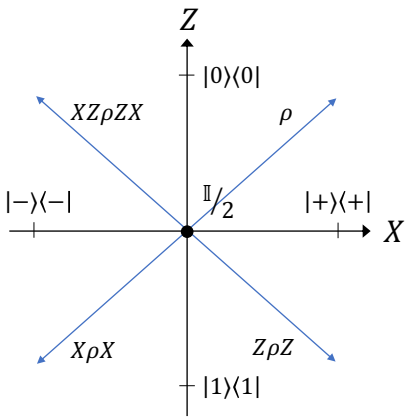

Fig. 2.6 Quantum one-time pad in the xz-plane of the Bloch sphere. A qubit ρ is encoded by two key bits: the operations $\mathbb{I}, X, Z, XZ$ are performed on the qubit with equal probability. The resulting mixture of states is the maximally mixed state (represented by the origin of the diagram).

where we used the fact that the Pauli matrices are observables (i.e. they are Hermitian and square to identity), and $\{X, Z\} = XZ + ZX = 0$. The interpretation of this calculation is that if we apply either $\mathbb{I}$, X, Z or XZ with equal probability to the Pauli matrix X then we obtain 0. Moreover, the same calculation can be done on the matrices Y and Z, and we obtain the same result, 0.

Exercise 2.5.1 Show that similarly, for any $M \in \{X, Y, Z\}$ we have

$$\frac{1}{4} \sum_{k_1,k_2 \in \{0,1\}} X^{k_1} Z^{k_2} M Z^{k_2} X^{k_1} = 0 \,. \tag{2.20}$$

Now let's use that any single-qubit state can be written as

$$\rho = \frac{1}{2} \left(\mathbb{I} + v_x X + v_y Y + v_z Z \right) \,.$$

By linearity and the calculation in Exercise 2.5.2 we then get that for any ρ,

$$\frac{1}{4} \sum_{k_1,k_2 \in \{0,1\}} X^{k_1} Z^{k_2} \rho Z^{k_2} X^{k_1} = \frac{\mathbb{I}}{2} \,. \tag{2.21}$$

What this equation means is precisely that from the point of view of anyone who does not know k_1, k_2 the bit- and phase-flipped state is completely independent of the input ρ, which means that all information contained in ρ is hidden from the eavesdropper who only "sees" $\frac{\mathbb{I}}{2}$ independently of ρ. This leads to the following quantum encryption scheme.

Protocol 2. *The quantum one-time pad is an encryption scheme for qubits. The key* $k =$

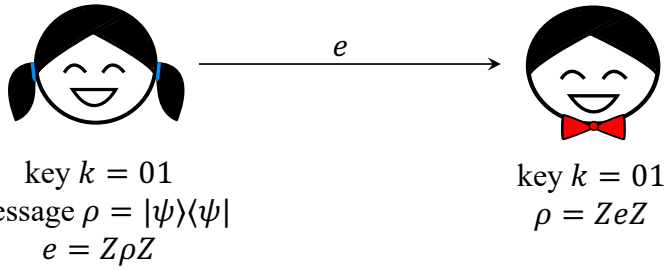

Fig. 2.7 An example of a one-time pad using unitary operations.

(k_1, k_2) is chosen uniformly at random in $K = \{0,1\}^2$. To encrypt a qubit in state ρ, Alice applies the unitary operation $X^{k_1}Z^{k_2}$ to ρ. To decrypt, Bob applies the inverse operation $(X^{k_1}Z^{k_2})^\dagger = Z^{k_2}X^{k_1}$.

The fact that the scheme is correct follows by definition, since the decryption operation is the inverse of the encryption operation. See 2.7 for an example. For security, we haven't given a complete formal definition for the quantum case. Doing so would take us too far so early in the book; if you are impatient you may jump ahead to Chapter 12. Intuitively, the scheme is perfectly secure because just as for the classical case, when we compute the reduced density matrix of an encrypted qubit, having traced out the key, we obtain something that is completely independent from the message itself.

To conclude we observe that the quantum one-time pad can easily be extended to n qubits by applying either $\mathbb{I}$, X, Z or XZ on each qubit, depending on two key bits associated with that qubit. This means that to encrypt n qubits, we use $2n$ bits of classical key. In Chapter 12 we will show that this is optimal: quantum information requires twice as many bits of key as classical information for perfectly secure encryption.

Exercise 2.5.2 Show that the collection of all (normalized) tensor products of Pauli matrices

$$P^s = \frac{1}{2^n} X^{s_1} Z^{s_2} \otimes X^{s_3} Z^{s_4} \otimes \ldots \otimes X^{s_{2n-1}} Z^{s_{2n}}$$

with $s \in \{0,1\}^{2n}$ form an orthogonal basis for the vector space of all $2^n \times 2^n$ matrices, i.e. for all $s, t \in \{0,1\}^{2n}$, $\mathrm{tr}((P^s)^\dagger P^t) = \delta_{st}$. In particular, any density matrix ρ on n qubits has a unique decomposition of the form

$$\rho = \frac{1}{2^n}\left(\mathbb{I}^{\otimes n} + \sum_{s \neq 0} v_s P^s\right), \tag{2.22}$$

for some complex coefficients v_s.

Quiz 2.5.2. *Alice encodes the qubit $|\psi\rangle$ using the quantum one-time pad. Eve is ignorant about the key bits k_1 and k_2. What is the state of the encoded qubit as seen by Eve?*

a) $\rho = \frac{1}{2}\left(|\psi\rangle\langle\psi| + XZ\,|\psi\rangle\langle\psi|\,ZX\right)$
b) $\rho = \frac{\mathbb{I}}{2}$
c) $\rho = X^{k_1}Z^{k_2}\,|\psi\rangle\langle\psi|\,Z^{k_2}X^{k_1}$

Quiz 2.5.3. *What is the state of the encoded qubit as seen by Bob who does know the key bits k_1 and k_2?*

a) $\rho = \frac{1}{2}\left(|\psi\rangle\langle\psi| + XZ\,|\psi\rangle\langle\psi|\,ZX\right),$
b) $\rho = \frac{\mathbb{I}}{2},$
c) $\rho = X^{k_1}Z^{k_2}\,|\psi\rangle\langle\psi|\,Z^{k_2}X^{k_1}$

2.6 Chapter notes

For additional background on probability theory you may consult any one of the many textbooks available, such as [Kel94, Ros10]. For the new elements of the quantum formalism introduced in this chapter we recommend the textbook [NC00]. For a more advanced discussion focused on quantum information theory, the book [Wil13] provides a wealth of information.

Shannon in his 1949 paper [Sha49] formally introduced the notion of perfect secrecy for classical communication and showed that the one-time pad achieves perfect secrecy. The task of encrypting quantum information is first considered in [AMTdW00a, BR00], who introduce the quantum one-time pad and show that it achieves perfect secrecy for quantum encryption. We will return to the topic of quantum encryption in Chapter 12.

2.7 Cheat Sheet

Trace

Given a matrix M, its trace is given by $\mathrm{tr}(M) = \sum_i M_{ii}$, i.e. the sum of its diagonal elements. The trace operation is cyclic, i.e. for any two matrices M, N, $\mathrm{tr}(MN) = \mathrm{tr}(NM)$.

Density Matrices

If we prepare a quantum system in the state ρ_x with probability p_x, then the state of the system is given by the density matrix

$$\rho = \sum_x p_x \rho_x.$$

Bloch representation of density matrices: any qubit density matrix can be written as

$$\rho = \frac{1}{2}\left(\mathbb{I} + v_x X + v_z Z + v_y Y\right),$$

and the Bloch vector $\vec{v} = (x_x, v_y, v_z)$ satisfies $\|\vec{v}\| \leq 1$, with equality if and only if ρ is pure.

Probability of measurement outcomes on a density matrix

If a quantum state with density matrix ρ is measured in the basis $\{|w_j\rangle\}_j$, then the probabilities of obtaining each outcome $|w_j\rangle$ is given by

$$p_{w_j} = \langle w_j| \rho |w_j\rangle = \mathrm{tr}(\rho |w_j\rangle\langle w_j|).$$

Combining density matrices

For density matrices $\rho_A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and $\rho_B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$ representing qubits A and B, the joint density matrix is given by

$$\rho_{AB} = \rho_A \otimes \rho_B := \begin{pmatrix} a_{11}\rho_B & a_{12}\rho_B \\ a_{21}\rho_B & a_{22}\rho_B \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{pmatrix}.$$

Partial trace

Given a bipartite matrix ρ_{AB} which has a decomposotion of the form

$$\rho_{AB} = \sum_{ijkl} \gamma_{ij}^{kl} |i\rangle\langle j|_A \otimes |k\rangle\langle l|_B,$$

where $\{|i\rangle_A\}$ and $\{|k\rangle_B\}$ are orthonormal bases of A and B respectively, the partial trace over system A yields the reduced state ρ_B

$$\rho_B = \mathrm{tr}_A(\rho_{AB}) = \sum_{ijk\ell} \gamma_{ij}^{k\ell} \mathrm{tr}(|i\rangle\langle j|) \otimes |k\rangle\langle \ell|_B = \sum_{k\ell} \left(\sum_j \gamma_{jj}^{k\ell} \right) |k\rangle\langle \ell|_B .$$

Properties of the Pauli Matrices X, Y, Z

For any $S_1, S_2 \in \{X, Y, Z\}$, $\{S_1, S_2\} = 2\delta_{S_1 S_2}\mathbb{I}$ where the anti-commutator is $\{A, B\} = AB + BA$. This implies the following properties.

- Zero trace: $\mathrm{tr}(S_1) = 0$.
- Orthogonality: $\mathrm{tr}(S_1^\dagger S_2) = 0$.
- Unitary: $S_1^\dagger S_1 = S_1 S_1^\dagger = \mathbb{I}$.
- Square to identity: $S_1^2 = \mathbb{I}$.

2.8 Quiz solutions

Quiz 2.2.1 c)
Quiz 2.2.2 b)
Quiz 2.2.3 b)
Quiz 2.2.4 c)
Quiz 2.2.5 b)
Quiz 2.2.6 a)
Quiz 2.2.7 b)
Quiz 2.2.8 b)
Quiz 2.2.9 c)
Quiz 2.2.10 a)
Quiz 2.3.1 d)
Quiz 2.4.1 c)
Quiz 2.4.2 a)
Quiz 2.4.3 a)
Quiz 2.4.4 c)
Quiz 2.4.5 c)
Quiz 2.5.1 c)
Quiz 2.5.2 b)
Quiz 2.5.3 c)