Contents

Chapter 3

The power of entanglement

We already encountered quantum entanglement in the form of the EPR pair $|\mathrm{EPR}\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$. In this chapter we will define entanglement more formally and explore some of the reasons that make it such an interesting topic in quantum information. To wet your appetite, let it already be said that in later chapters we will see that entanglement allows us to guarantee the security of communications based only on the laws of nature. We also know that entanglement is a necessary ingredient in the most impressive quantum algorithms, such as Shor's algorithm for factoring, and for quantum error correction.

3.1 Entanglement

If we combine two qubits A and B, each of which is in a pure state, the joint state of the two qubits is given by

$$|\psi\rangle_{AB} = |\psi_1\rangle_A \otimes |\psi_2\rangle_B \ . \tag{3.1}$$

Any two-qubit state that is either directly of this form, or is a mixture of states of this form, is called *separable*. Entangled state are states which are *not* separable. In other words, a pure state $|\psi\rangle$ is entangled if and only if

$$|\psi\rangle \neq |\psi_1\rangle \otimes |\psi_2\rangle \ , \tag{3.2}$$

for any possible choice of $|\psi_1\rangle$ and $|\psi_2\rangle$. A mixed state ρ is entangled if and only if it cannot be written as a convex combination of pure product states of the form in Eq. (3.1).

Definition 3.1.1 (Entanglement)**.** *Consider two quantum systems A and B. The joint state ρ_{AB} is* separable *if there exists a probability distribution $\{p_i\}_i$, and sets*

of density matrices $\{\rho_i^A\}_i, \{\rho_i^B\}_i$ *such that*

$$\rho_{AB} = \sum_i p_i \, \rho_i^A \otimes \rho_i^B. \tag{3.3}$$

If there exists no such decomposition ρ_{AB} *is called* entangled.
If $\rho_{AB} = |\Psi\rangle\langle\Psi|_{AB}$ *is a pure state, then* $|\Psi\rangle_{AB}$ *is separable if and only if there exists* $|\psi\rangle_A, |\psi\rangle_B$ *such that*

$$|\Psi\rangle_{AB} = |\psi\rangle_A \otimes |\psi\rangle_B . \tag{3.4}$$

Example 3.1.1. *An example of an entangled state of two qubits is the EPR pair*

$$|\mathrm{EPR}\rangle_{AB} = \frac{1}{\sqrt{2}} \left(|00\rangle_{AB} + |11\rangle_{AB}\right) . \tag{3.5}$$

When we learn more about entanglement later on, we will see that this state is, in a precise sense, the "most entangled" state of two qubits. The EPR pair is thus often referred to as a maximally entangled *state. (There are other two-qubit states which are different from the EPR pair but have just about the same "amount" of entanglement; we will learn about these other maximally entangled states later.)* ■

As we already saw in Example **??**, the EPR pair has the special property that it can be written in many symmetric ways. For instance, in the Hadamard basis

$$|\mathrm{EPR}\rangle_{AB} = \frac{1}{\sqrt{2}} \left(|{+}{+}\rangle_{AB} + |{-}{-}\rangle_{AB}\right) . \tag{3.6}$$

Thus measurements of both qubits of $|\mathrm{EPR}\rangle_{AB}$ in the standard basis, or the Hadamard basis, always produce the same outcome. In a few weeks we will see that this property can even be used to *characterize* the EPR pair: it is the only two-qubit state having this property!

Exercise 3.1.1. *Suppose that* ρ_{AB} *is a two-qubit separable state. Show that if a measurement of both qubits of* ρ_{AB} *in the standard basis always yields the same outcome, then a measurement of both qubits in the Hadamard basis necessarily has non-zero probability of giving different outcomes. Deduce a proof that the EPR pair* (3.5) *is not a separable state.* ■

Entanglement has another interesting property which we will see later, called "monogamy". Monogamy states that if two systems are maximally entangled with each other then they cannot have any entanglement with any other system: equivalently, they must be in tensor product with the remainder of the universe.

Example 3.1.2. *Consider the state*

$$|\psi\rangle_{AB} = \frac{1}{\sqrt{2}} \left(|01\rangle_{AB} + |11\rangle_{AB} \right) \ . \tag{3.7}$$

In contrast to the EPR pair in Example 3.1.1 this state is not entangled, since $|\psi\rangle_{AB} = \frac{1}{\sqrt{2}} \left(|0\rangle_A + |1\rangle_A \right) \otimes |1\rangle_B = |+\rangle_A \otimes |1\rangle_B$. ∎

Example 3.1.3. *Consider the density matrix*

$$\rho_{AB} = \frac{1}{2}|0\rangle\langle 0|_A \otimes |1\rangle\langle 1|_B + \frac{1}{2}|+\rangle\langle +|_A \otimes |-\rangle\langle -|_B. \tag{3.8}$$

Such a state is in the form of Eq. (3.3)*, so it is not entangled: it is separable. Note that this does not imply that the systems A and B are necessarily independent: here they are correlated, but not entangled. (We would typically say that they are "classically correlated".)* ∎

Example 3.1.4. *Any cq-state, i.e. a state of the form* $\rho_{XQ} = \sum_x p_x |x\rangle\langle x|_X \otimes \rho_x^Q$*, is separable.* ∎

It is important to make the distinction between the two states

$$\rho_{AB} = \frac{1}{2}|0\rangle\langle 0|_A \otimes |0\rangle\langle 0|_B + \frac{1}{2}|1\rangle\langle 1|_A \otimes |1\rangle\langle 1|_B$$

and

$$\sigma_{AB} = |\mathrm{EPR}\rangle\langle \mathrm{EPR}|_{AB} \ .$$

For the state ρ_{AB}, if A is measured in the standard basis then whenever $|0\rangle_A$ is observed the state on B is $|0\rangle_B$; likewise when $|1\rangle_A$ is observed, the state on B is $|1\rangle_B$. This is also true for σ_{AB}. However, consider measuring system A of ρ_{AB} in the Hadamard basis. The corresponding measurement operators are $|+\rangle\langle +|_A \otimes \mathbb{I}_B, |-\rangle\langle -|_A \otimes \mathbb{I}_B$. The post-measurement state conditioned on obtaining the outcome $|+\rangle_A$ is then

$$\rho^{AB}_{|+_A} = \frac{(|+\rangle\langle +|_A \otimes \mathbb{I}_B)\rho_{AB}(|+\rangle\langle +|_A \otimes \mathbb{I}_B)}{\mathrm{tr}((|+\rangle\langle +|_A \otimes \mathbb{I}_B)\rho_{AB})} \tag{3.9}$$

$$= 2 \cdot \left(\frac{1}{2}\frac{1}{2}|+\rangle\langle +|_A \otimes |0\rangle\langle 0|_B + \frac{1}{2}\frac{1}{2}|+\rangle\langle +|_A \otimes |1\rangle\langle 1|_B \right) \tag{3.10}$$

$$= |+\rangle\langle +|_A \otimes \frac{\mathbb{I}_B}{2}, \tag{3.11}$$

and we see that the reduced state on B, $\rho^B_{|+_A} = \frac{\mathbb{I}_B}{2}$, is maximally mixed. In contrast, using that the state $|\mathrm{EPR}\rangle_{AB}$ can be rewritten as

$$|\mathrm{EPR}\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle_{AB} + |11\rangle_{AB}) = \frac{1}{\sqrt{2}}(|{+}{+}\rangle_{AB} + |{-}{-}\rangle_{AB}), \qquad (3.12)$$

when σ_{AB} is measured with respect to the Hadamard basis on system A, conditioned on the outcome $|+\rangle_A$, the reduced state on B is $\sigma^B_{|+_A} = |+\rangle\langle+|_B$. In particular this state is pure: it is very different from the totally mixed state we obtained by performing the same experiment on ρ_{AB}. This is a sense in which the correlations in σ_{AB} are stronger than those in ρ_{AB}.

3.2 Purifications

In Chapter **??** we learned about the partial trace operation, which provides a way to describe the state of a subsystem when given a description of the state on a larger composite system. Even if the state of the larger system is pure, the reduced state can sometimes be mixed, and this is a signature of entanglement in the larger state.

Is it possible to reverse this process? Suppose given a density matrix ρ_A describing a quantum state on system A. Is it always possible to find a pure state $\rho_{AB} = |\Psi\rangle\langle\Psi|_{AB}$ such that $\mathrm{tr}_B(\rho_{AB}) = \rho_A$? Such a state is called a *purification* of ρ_A.

Definition 3.2.1 (Purification)**.** *Given any density matrix ρ_A, a pure state $|\Psi_{AB}\rangle$ is a* purification *of A if* $\mathrm{tr}_B(|\Psi\rangle\langle\Psi|_{AB}) = \rho_A$.

Let's see how an arbitrary density matrix ρ_A can be purified. As a first step, diagonalize ρ_A, expressing it as a mixture

$$\rho_A = \sum_{j=1}^{d_A} \lambda_j |\phi_j\rangle\langle\phi_j| \,, \qquad (3.13)$$

where λ_j are the (necessarily non-negative) eigenvalues of ρ_A and $|\phi_j\rangle$ the eigenstates. Since ρ_A is a density matrix the λ_j are non-negative and sum to 1. We've seen an interpretation of density matrices before: here we would say that ρ_A describes a quantum system that is in a probabilistic mixture of being in state $|\phi_j\rangle$ with probability λ_j. But who "controls" which part of the mixture A is in?

Let's introduce an imaginary system B which achieves just this. Let $\{|j\rangle_B\}_{j\in\{1,\ldots,d_B\}}$ be the standard basis for a system B of dimension $d_B = d_A$, and consider the pure

state

$$|\Psi\rangle_{AB} = \sum_{j=1}^{d_A} \sqrt{\lambda_j}\, |\phi_j\rangle_A \otimes |j\rangle_B \ , \tag{3.14}$$

where $\{|j\rangle_B\}_j$ is the standard basis on system B. Suppose we were to measure the B system of $|\Psi\rangle_{AB}$ in the standard basis. We know what will happen: we will obtain outcome j with probability $\langle\Psi|_{AB} M_j |\Psi\rangle_{AB}$, where $M_j = \mathbb{I}_A \otimes |j\rangle\langle j|_B$, and a short calculation will convince you this equals λ_j. Since we're using a projective measurement, we can describe the post-measurement state easily as being proportional to $M_j|\Psi\rangle\langle\Psi|_{AB}M_j$, and looking at the A system only we find that it is $|\phi_j\rangle\langle\phi_j|_A$.

To summarize, a measurement of system B gives outcome j with probability λ_j, and the post-measurement state on A is precisely $|\phi_j\rangle\langle\phi_j|$. This implies that $\mathrm{Tr}_B(|\Psi\rangle\langle\Psi|_{AB}) = \rho_A$, a fact which can be verified directly using the mathematical definition of the partial trace operation.

Are purifications unique? You'll notice that in the above construction we made the choice of the standard basis for system B, but any other basis would have worked just as well. So it seems like we at least have a choice of basis on system B: there is a "unitary degree of freedom". To see that this is the only freedom that we have in choosing a purification, we first need to learn about a very convenient representation of bipartite pure states, the *Schmidt decomposition*.

3.2.1 The Schmidt decomposition

The purification that we constructed in (3.14) has a special form: it is expressed as a sum, with non-negative coefficients whose squares sum to 1, of tensor products of basis states for the A and B systems respectively. As we saw, this particular form is convenient because it lets us compute the reduced states in A and B very easily. Unfortunately, not every state is always given in this way: for example, if we write $|\Psi\rangle_{AB} = \frac{1}{\sqrt{2}}(|0\rangle_A |0\rangle_B + |+\rangle_A |1\rangle_B)$ then the two states $|0\rangle_A, |+\rangle_A$ on A are not orthogonal. But maybe the same state can be written in a more convenient form? The answer is yes, and it is given by the Schmidt decomposition.

Theorem 3.2.1 (Schmidt decomposition)**.** *Consider quantum systems A and B with dimensions d_A, d_B respectively, and let $d = \min(d_A, d_B)$. Any pure bipartite state $|\Psi\rangle_{AB}$ has a Schmidt decomposition*

$$|\Psi\rangle_{AB} = \sum_{i=1}^{d} \sqrt{\lambda_i}\, |u_i\rangle_A |v_i\rangle_B \,, \tag{3.15}$$

where $\lambda_i \geq 0$ *and* $\{|u_i\rangle_A\}_i, \{|v_i\rangle_B\}_i$ *are collections of orthonormal vectors. The coefficients* $\sqrt{\lambda_i}$ *are called the* Schmidt coefficients *and* $|u_i\rangle_A, |v_i\rangle_B$ *the* Schmidt vectors.

You can also find a detailed proof of the theorem in Section 2.5 of [**nielsen&chuang:qc**]. The main idea is to start by expressing $|\Psi\rangle_{AB} = \sum_{j,k} \alpha_{j,k} |j\rangle_A |k\rangle_B$ using the standard bases of A and B, and then write the singular value decomposition of the $d_A \times d_B$ matrix with coefficients $\alpha_{j,k}$ to recover the $\sqrt{\lambda_i}$ (the singular values), the $|u_i\rangle_A$ (the left eigenvectors), and the $|v_i\rangle_B$ (the right eigenvectors).

The Schmidt decomposition has many interesting consequences. A first consequence is that it provides a simple recipe for computing the reduced density matrices: given a state of the form (3.15), we immediately get $\rho_A = \sum_i \lambda_i |u_i\rangle \langle u_i|_A$ and $\rho_B = \sum_i \lambda_i |v_i\rangle \langle v_i|_B$. An important observation is that ρ_A and ρ_B have the same eigenvalues, which are precisely the squares of the Schmidt coefficients. As a consequence, given any two density matrices ρ_A and ρ_B, there exists a pure bipartite state $|\Psi\rangle_{AB}$ such that $\rho_A = \mathrm{Tr}_B(|\Psi\rangle \langle\Psi|_{AB})$ and $\rho_B = \mathrm{Tr}_A(|\Psi\rangle \langle\Psi|_{AB})$ if and only if ρ_A and ρ_B have the same spectrum! Without the Schmidt decomposition this is not at all an obvious fact to prove.

The same observation also implies that the Schmidt coefficients are uniquely defined: they are the square roots of the eigenvalues of the reduced density matrix. The Schmidt vectors are also unique, up to degeneracy and choice of phase: if an eigenvalue has an associated eigenspace of dimension 1 only then the associated Schmidt vector must be the corresponding eigenvector. If the eigenspace has dimension more than 1 we can choose as Schmidt vectors any basis for the subspace. And note that in (3.15) we can always multiply $|u_i\rangle$ by $e^{i\theta_i}$, and $|v_i\rangle$ by $e^{-i\theta_i}$, so there is a phase degree of freedom.

Another important consequence of the Schmidt decomposition is that it provides us with a way to measure entanglement between the A and B systems in a pure state $|\Psi_{AB}\rangle$. A first, rather rough but convenient such measure is given by the number of non-zero coefficients $\sqrt{\lambda_j}$. This measure is called the *Schmidt rank*. An important is that, since the Schmidt rank is uniquely defined for any given state, it allows us to tell if a state is entangled or not: if the Schmidt rank is 1 then the state is a product state, and if it is strictly larger than 1 then the state is entangled.

Definition 3.2.2 (Schmidt rank)**.** *For any bipartite pure state with Schmidt decomposition* $|\Psi\rangle_{AB} = \sum_{i=1}^{d} \sqrt{\lambda_i} |a_i\rangle_A |b_i\rangle_B$*, the* Schmidt rank *is defined as the number of non-zero coefficients* $\sqrt{\lambda_i}$*. It is also equal to* $\mathrm{rank}(\rho_A)$ *and* $\mathrm{rank}(\rho_B)$*.*

The Schmidt coefficients provide a finer way to measure entanglement than the Schmidt rank. A natural measure, called "entropy of entanglement", consists in taking the entropy of the distribution specified by the squares of the coefficients.

If the entropy is 0 then there is only a single coefficient equal to 1, and the state is not entangled. But as soon as the entropy is positive the state is entangled. This measure is finer than the Schmidt rank. For example, it distinguishes the entanglement in the two states

$$|\Psi\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle \qquad \text{and} \qquad |\phi\rangle = \sqrt{1-\varepsilon}\,|00\rangle + \sqrt{\varepsilon}\,|11\rangle\,.$$

For small $0 < \varepsilon < 1/2$ both states have the same Schmidt rank, but the first one has entanglement entropy 1 whereas the second has entanglement entropy $H(\varepsilon)$ (where H is the binary entropy function, defined as $H(x) = -x\log(x) - (1-x)\log(1-x)$ for $x \in [0,1]$) going to 0 as $\varepsilon \to 0$. This is the reason why we call the EPR pair "maximally entangled": its entanglement entropy is maximal among all two-qubit states.

3.2.2 Uhlmann's theorem

Let's return to the topic of the freedom in choosing purifications of a density matrix. We saw that there is at least a unitary degree of freedom by choosing a basis on the purifying system B. Uhlmann's theorem states that this is precisely the only freedom we have.

Theorem 3.2.2 (Uhlmann's theorem)**.** *Suppose given a density matrix ρ_A and a purification of A given by $|\Psi\rangle_{AB}$. Then another state $|\Phi\rangle_{AB}$ is also a purification of A if and only if there exists a unitary U_B such that*

$$|\Phi\rangle_{AB} = \mathbb{I}_A \otimes U_B\,|\psi\rangle_{AB}\,. \tag{3.16}$$

We already saw a proof of the "if" part of the theorem. To show the converse, i.e. that two purifications must always be related by a unitary, consider the Schmidt decomposition:

$$|\Phi\rangle_{AB} = \sum_i \sqrt{\lambda_i}\,|u_i\rangle_A\,|v_i\rangle_B\,,$$
$$|\Psi\rangle_{AB} = \sum_i \sqrt{\mu_i}\,|w_i\rangle_A\,|z_i\rangle_B\,.$$

As we know the λ_i are uniquely defined: they are the eigenvalues of ρ_A. So if $|\Phi\rangle_{AB}$ and $|\Psi\rangle_{AB}$ are both purifications of the same ρ_A, we must have $\lambda_i = \mu_i$. Now suppose for simplicity that all eigenvalues are non-degenerate. (The degenerate case can be treated by a simple extension of this argument.) Then the $|u_i\rangle_A$ are also uniquely determined: they are the eigenvectors of ρ_A associated to the λ_i.

Therefore $|u_i\rangle_A = |w_i\rangle_A$ as well! Thus we see that the only choice we have left are the $|v_i\rangle_B$, or $|z_i\rangle_B$: since the density matrix ρ_B of the purification is not specified a priori, we may choose any orthonormal basis of the B system. Since any two orthonormal bases of the same space are related by a unitary matrix, this choice of basis is precisely the degree of freedom that is guaranteed by Uhlmann's theorem.

3.3 Two applications

Let's discuss a couple cryptographic application of the notions we just introduced. Both applications are based on the same following four states:

$$\begin{aligned} |\psi_{00}\rangle_{AB} &= \frac{1}{\sqrt{2}}(|00\rangle_{AB} + |11\rangle_{AB}), \qquad |\psi_{01}\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle_{AB} - |11\rangle_{AB}), \\ |\psi_{10}\rangle_{AB} &= \frac{1}{\sqrt{2}}(|01\rangle_{AB} + |10\rangle_{AB}), \qquad |\psi_{11}\rangle_{AB} = \frac{1}{\sqrt{2}}(|01\rangle_{AB} - |10\rangle_{AB}). \end{aligned} \tag{3.17}$$

These states are called the *Bell states*. Observe that they are orthonormal and thus form a basis of $\mathbb{C}^2 \otimes \mathbb{C}^2$. In Example **??** we calculated the reduced density on Alice's system A of one of those states, the EPR pair $|\psi_{00}\rangle_{AB}$:

$$\begin{aligned} \rho_{00}^A &= \mathrm{tr}_B(|\psi_{00}\rangle\langle\psi_{00}|_{AB}) \\ &= \frac{1}{2}\big(|0\rangle\langle 0|_A \,\mathrm{tr}_B(|0\rangle\langle 0|_B) + |0\rangle\langle 1|_A \,\mathrm{tr}_B(|0\rangle\langle 1|_B) \\ &\qquad + |1\rangle\langle 0|_A \,\mathrm{tr}_B(|1\rangle\langle 0|_B) + |1\rangle\langle 1|_A \,\mathrm{tr}_B(|1\rangle\langle 1|_B)\big) \\ &= \frac{1}{2}(|0\rangle\langle 0|_A + |1\rangle\langle 1|_A) = \frac{\mathbb{I}_A}{2}. \end{aligned}$$

Calculating the reduced states on either A or B for each each of the states in (3.17) always gives the same result,

$$\rho_{00}^A = \rho_{01}^A = \rho_{10}^A = \rho_{11}^A = \frac{\mathbb{I}}{2}, \tag{3.18}$$

$$\rho_{00}^B = \rho_{01}^B = \rho_{10}^B = \rho_{11}^B = \frac{\mathbb{I}}{2}. \tag{3.19}$$

Thus the four Bell states in (3.17) are perfectly distinguishable (because they are orthogonal), yet they all have the same reduced density matrices, and thus are locally indistinguishable. These two facts are key to the following applications.

Secret sharing

Our first application is called *secret sharing*. Imagine a country owns nuclear weapons, but wants to make sure that both the queen (Alice) and king (Bob) have to come together to activate them. One solution would be to give half of the launch codes $s = (s_1, \ldots, s_\ell) \in \{0,1\}^\ell$ to Alice, and the other half to Bob, thereby making sure that they both need to reveal their share of the information in order for the weapons to be activated. A drawback of this scheme is that each of them does have significant information about the launch codes, namely half of the bits. And what if there is only one bit? (Although that wouldn't be very secure, would it...)

The goal in a secret sharing scheme is to divide the information s into shares in such a way that any unauthorized set of parties (in the example, Alice or Bob alone) cannot learn *anything* at all about the secret. Remembering the idea behind the one-time pad, a much better scheme would be to choose a random string $r \in \{0,1\}^\ell$ and give r to Alice and $r \oplus s$ to Bob. In this case neither Alice nor Bob individually has any information about s; their respective secrets appear uniformly random. Yet when they come together they can easily recover s!

From this example we see that given a random classical bit one can construct a secret sharing scheme between Alice and Bob that shares a single secret bit s. However they can do better if they are each given a qubit instead. Consider the case that Alice and Bob are given one of the four Bell states in (3.17). Since the reduced state of each of those states on each subsystem is maximally mixed, neither Alice nor Bob can gain any information on which of the states $|\psi_{00}\rangle_{AB}, |\psi_{01}\rangle_{AB}, |\psi_{10}\rangle_{AB}, |\psi_{11}\rangle_{AB}$ they have one qubit of! However, due to the fact that these states together form a basis, when Alice and Bob come together they can perform a measurement in that basis that perfectly distinguishes which state they have, yielding two bits of information.

Exercise 3.3.1. *Suppose there are now three parties, Alice, Bob and Charlie (the prime minister is also given a share of the nuclear codes!). Give a secret sharing scheme, based on a tripartite entangled state, such that no individual party has any information about the secret but the three of them together are able to recover the secret. Better: can you give a scheme such that no two of them has any information about the secret. Different: give a scheme such that no individual has any information about the secret, but any group of two can recover it.* ∎

Superdense coding

A different application of the usefulness of entanglement is to *superdense coding*. The task in dense coding consists in sending classical bits of information from Alice to Bob by encoding them in a quantum state that is as small as possible.

Let's see how using entanglement we can send two classical bits using a single qubit.

Suppose Alice and Bob share the state $|\psi_{00}\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle_{AB} + |11\rangle_{AB})$, and that Alice performs a unitary on her qubit as indicated in Table 3.1, depending on which bits $ab \in \{00, 01, 10, 11\}$ she wants to send to Bob.

Classical information a, b	Unitary $X_A^a Z_A^b$	Final joint state
00	$\mathbb{I}_A$	$\frac{1}{\sqrt{2}}(\|00\rangle_{AB} + \|11\rangle_{AB})$
01	X_A	$\frac{1}{\sqrt{2}}(\|10\rangle_{AB} + \|01\rangle_{AB})$
10	Z_A	$\frac{1}{\sqrt{2}}(\|00\rangle_{AB} - \|11\rangle_{AB})$
11	$-X_A Z_A$	$\frac{1}{\sqrt{2}}(\|01\rangle_{AB} - \|10\rangle_{AB})$

Table 3.1: Unitary operation performed by Alice in order to encode her two classical bits $ab \in \{0, 1\}^2$.

As we already saw, the four states on the right-hand side in Table 3.1 form the Bell basis, and in particular they are perfectly distinguishable. Hence if Alice sends her qubit over to Bob, he can perform a measurement in the Bell basis and recover both of Alice's classical bits.

3.4 Bell-Nonlocality

Entanglement has many counter-intuitive properties. A very important one is that it allows correlations between two particles — two qubits — that cannot be replicated classically. The very first example of such correlations was demonstrated in [**bell1964einstein**], where Bell proved that the predictions of quantum theory are incompatible with those of any classical theory satisfying a natural notion of *locality*.

The modern way to understand Bell non-locality is by means of so-called non-local games. Let's imagine that we play a game with two players, which we'll again call Alice and Bob. Alice has a system A, and Bob has some system B. In this game, we will ask Alice and Bob questions, and collect answers. Let us denote the possible questions to Alice and Bob x and y, and label the answers a and b. We will play this game many times, and in each round choose the questions to ask with some probability $p(x, y)$. As you might expect our game has some rules. We denote these rules using a predicate $V(a, b|x, y)$, which takes the value

"1" if a and b are winning answers for questions x and y. To be fair, Alice and Bob know the rules of the game given by $V(a,b|x,y)$, and also the distribution $p(x,y)$. They can agree on any strategy before the game starts. However, once we start asking questions they are no longer allowed to communicate. Of interest to us will be the probability that Alice and Bob win the game, maximized over all possible strategies. That is,

$$p_{\text{win}} = \max_{strategy} \sum_{x,y} p(x,y) \sum_{a,b} V(a,b|x,y)\, p(a,b|x,y) \;, \tag{3.20}$$

where $p(a,b|x,y)$ is the probability that Alice and Bob produce answers a and b given x and y according to their chosen strategy.

What strategies are allowed? In a classical world, Alice and Bob can only have a classical strategy. A deterministic classical strategy is simply given by functions $f_A(x) = a$ and $f_B(y) = b$ that take the questions x and y to answers a and b. We then have $p(a,b|x,y) = 1$ whenever $a = f_A(x)$ and $b = f_B(y)$, and $p(a,b|x,y) = 0$ otherwise. Possibly, Alice and Bob also use shared randomness. That is, they have another string r, which they share with probability $p(r)$. In physics, r is also referred to as a hidden variable, but we will take the more operational viewpoint of shared randomness. In a strategy using shared randomness r, classical Alice and Bob can still only apply functions, except that now the function can also depend on the shared randomness r: $a = f_A(x,r)$ and $b = f_B(y,r)$. In terms of the probabilities we then have $p(a,b|x,y,r) = 1$ if $a = f_A(x,r)$ and $b = f_B(y,r)$ and $p(a,b|x,y,r) = 0$ otherwise. This gives

$$p(a,b|x,y) = \sum_r p(r) p(a,b|x,y,r) \;. \tag{3.21}$$

Does shared randomness help Alice and Bob? Note that for a classical strategy based on shared randomness we have

$$p_{\text{win}} = \max_{class.strat.} \sum_{x,y} p(x,y) \sum_{a,b} V(a,b|x,y) \sum_r p(r) p(a,b|x,y,r) \tag{3.22}$$

$$= \max_{class.strat.} \sum_r p(r) \Big(\sum_{x,y} p(x,y) \sum_{a,b} V(a,b|x,y)\, p(a,b|x,y,r) \Big) \;. \tag{3.23}$$

Note that the quantity in brackets is largest for some particular value(s) of r. Since Alice and Bob want to maximize their winning probability, they can fix the best possible r. Doing so gives them a deterministic strategy $a = f_A(x,r)$ and $b = f_A(y,r)$ where r is now fixed.

Why would we care about this at all? It turns out that for many games, a *quantum* strategy can achieve a higher winning probability. This is of fundamental

importance for our understanding of nature. Specifically, a *quantum strategy* means that Alice and Bob can pick a state ρ_{AB} to share, and agree on measurements to perform depending on their respective questions. That is, x and y will label a choice of measurement, and a and b are the outcomes of that measurement.

What's more, observing a higher winning probability is a signature of entanglement: quantumly, Alice and Bob can achieve a higher winning probability *only* if they are entangled, making such games into *tests* for entanglement. Testing whether the stated shared by Alice and Bob is entangled forms a crucial element in quantum key distribution, as we will see in later chapters.

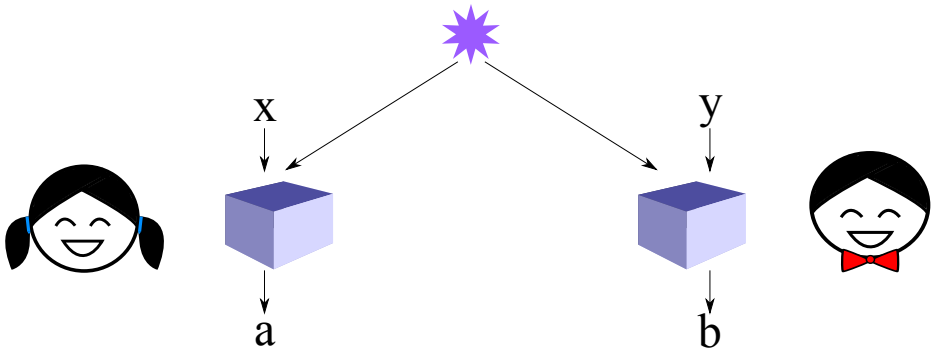

Figure 3.1: A non-local game. Alice and Bob are given questions x and y, and must return answers a and b. If Alice and Bob are quantum, then x and y label measurement settings and a and b are measurement outcomes.

3.4.1 Example of a non-local game: CHSH

Let us have a look at a very simple game based on the famous CHSH inequality. It will turn out to be extremely useful for quantum cryptography. At the start of the game, we send two bits x and y to Alice and Bob respectively, where we choose x with uniform probabilities $p(x = 0) = p(x = 1) = 1/2$ and y with probabilities $p(y = 0) = p(y = 1) = 1/2$. In turn, Alice and Bob will return answer bits a and b. Alice and Bob win the game if and only if

$$x \cdot y = a + b \mod 2 \,. \tag{3.24}$$

In terms of the predicate $V(a, b|x, y)$ this means that $V(a, b|x, y) = 1$ if $x \cdot y = a + b \mod 2$ and $V(a, b|x, y) = 0$ otherwise. We are interested in the probability that

Alice and Bob win the game. This probability can be written as

$$p_{\text{win}}^{\text{CHSH}} = \frac{1}{4} \sum_{x,y \in \{0,1\}} \sum_{\substack{a,b \\ a+b \mod 2 = x \cdot y}} p(a,b|x,y) \,, \tag{3.25}$$

where $p(a,b|x,y)$ is the probability that Alice and Bob answer a and b given questions x and y. What can Alice and Bob do to win this game?

Classical winning probability

If Alice and Bob are entirely classical beings, then each of their answers must be a direct function of their question. For example, if $x = 0$, then Alice and Bob could agree as part of their strategy that Alice will then always answer $a = 0$. We see that as long as $x = 0$ or $y = 0$, then $x \cdot y = 0$. In this case, Alice and Bob want to achieve $a + b \mod 2 = 0$. However, if $x = y = 1$ then they would like to give answers such that $a + b \mod 2 = 1$. What makes this difficult for Alice and Bob is that they cannot communicate during the game. This means in particular that Alice's answer a can only depend on x (but not on y) and similarly Bob's answer b can only depend on y (but not on x).

It it not difficult to see (you may wish to check!) by trying out all possible deterministic strategies for Alice and Bob (remember that we showed that even though shared randomness is allowed in principle, it never helps improve upon the best deterministic strategy), that classically the maximum winning probability that can be achieved is

$$p_{\text{win}}^{\text{CHSH}} = \frac{3}{4} \,. \tag{3.26}$$

Alice and Bob can achieve this winning probability with the strategy of answering $a = b = 0$ always, which means $a + b \mod 2 = 0$, which is correct in 3 out of the 4 possible cases. Only when $x = y = 1$ will Alice and Bob make a mistake.

Quantum winning probability

It turns out that Alice and Bob can do significantly better with a quantum strategy, using shared entanglement. Indeed, suppose that Alice and Bob share an EPR pair, where we label the qubit held by Alice (A) and the one held by Bob (B).

$$|\Psi\rangle_{AB} = \frac{1}{\sqrt{2}} \left(|0\rangle_A |0\rangle_B + |1\rangle_A |1\rangle_B \right) . \tag{3.27}$$

Suppose now that when $x = 0$, Alice measures her qubit in the basis $\{|0\rangle, |1\rangle\}$. Otherwise when $x = 1$, she measures in the basis $\{|+\rangle, |-\rangle\}$. Bob does something

slightly different: when his question $y = 0$, Bob measures his qubit in the basis $|v_1\rangle, |v_2\rangle$ where

$$|v_1\rangle = \cos(\pi/8)\,|0\rangle + \sin(\pi/8)\,|1\rangle\,, \qquad |v_2\rangle = -\sin(\pi/8)\,|0\rangle + \cos(\pi/8)\,|1\rangle\,, \tag{3.28}$$

and when $y = 1$, he measures in the basis $|w_1\rangle, |w_2\rangle$, where

$$|w_1\rangle = \cos(\pi/8)\,|0\rangle - \sin(\pi/8)\,|1\rangle\,, \qquad |w_2\rangle = \sin(\pi/8)\,|0\rangle + \cos(\pi/8)\,|1\rangle\,. \tag{3.29}$$

Consider the case where $x = 0, y = 0$. This means Alice measures in the basis $\{|0\rangle, |1\rangle\}$ and Bob in the basis $\{|v_1\rangle, |v_2\rangle\}$. The probability of winning, conditioned on $x = 0, y = 0$ is given by

$$p_{\text{win}|x=0,y=0} = p(a=0,b=0|x=0,y=0) + p(a=1,b=1|x=0,y=0) \tag{3.30}$$

$$= |\,\langle 0|_A \langle v_1|_B\,\Psi\rangle_{AB}|^2 + |\,\langle 1|_A \langle v_2|_B\,\Psi\rangle_{AB}|^2 \tag{3.31}$$

$$= 2\left|\frac{1}{\sqrt{2}}\cos\frac{\pi}{8}\right|^2 = \cos^2\frac{\pi}{8}. \tag{3.32}$$

The probability of winning, conditioned on $x = 0, y = 1$ is given by a similar expression

$$p_{\text{win}|x=1,y=0} = p(a=0,b=0|x=1,y=0) + p(a=1,b=1|x=1,y=0) \tag{3.33}$$

$$= |\,\langle 0|_A \langle w_1|_B\,\Psi\rangle_{AB}|^2 + |\,\langle 1|_A \langle w_2|_B\,\Psi\rangle_{AB}|^2 \tag{3.34}$$

$$= 2\left|\frac{1}{\sqrt{2}}\cos\frac{\pi}{8}\right|^2 = \cos^2\frac{\pi}{8}. \tag{3.35}$$

Exercise 3.4.1. *Show similarly that*

$$p_{\text{win}|x=0,y=1} = p_{\text{win}|x=1,y=1} = \cos^2\frac{\pi}{8}\,.$$

■

As a result, the quantum strategy for Alice and Bob succeeds with overall probability $\cos^2\frac{\pi}{8} \approx 0.85$, which is strictly larger than the best classical strategy! This is the power of entanglement.

3.4.2 Implications

This counterintuitive effect of entanglement has far reaching consequences. The first is of a conceptual nature. You may have already been wondering what *actually* happens when we measure a quantum particle. Sure, there is a probabilistic rule that we learned about. But "in reality", shouldn't it be the case that the outcome is a predetermined property that the particle — it's just that our "formalism" doesn't really allow us to say it, but only gives us access to probabilities? Maybe every particle has a classical "cheat sheet" attached to it, so that the cheat sheet can be used to specify the outcome for any possible measurement that we can make on it?

Now observe that such a cheat sheet, when computed for the two particles in an EPR pair, could be used to construct a classical strategy in the CHSH game: For every x, we'd look up Alice's answer a in the "cheat sheet" associated to her qubit. In physics, such cheat sheets are also called local hidden variables.

The fact that quantum strategies can beat classical strategies in the CHSH game implies that nature does not work that way! There are no classical cheat sheets, and nature is inherently quantum. Many experiments of ever increasing accuracy have been performed that verify that Alice and Bob can indeed achieve a higher winning probability in the CHSH game than the classical world would allow. This tells us that the world is not classical, but we need more sophisticated tools to describe it - such as quantum mechanics. It also means that when trying to build the ultimate computing and communication devices, we should make full use of what nature allows and go quantum.

We will see in the coming chapters how to use this simple game to verify the presence of entanglement, test unknown quantum devices, and even create secure encryption keys.

3.5 The monogamy of entanglement

Let's get back to the property mentioned in the very beginning of this chapter: that entanglement is monogamous. We know that two systems A and B can be in a joint pure state that is entangled, such as the "maximally entangled" EPR pair $|\mathrm{EPR}\rangle_{AB}$. All our examples, however, had to do with entanglement between two systems A and B. But what about a third system, call it C for Charlie? Of course we could always consider three EPR pairs, $|\mathrm{EPR}\rangle_{AB}$, $|\mathrm{EPR}\rangle_{BC}$ and $|\mathrm{EPR}\rangle_{AC}$. If this is the state of the three systems however we can't really talk about tripartite entanglement, because the correlations are always between any two of the three parties. Is it possible to create a joint state $|\Psi\rangle_{ABC}$ in which the strong correlations of the EPR pair are shared simultaneously between all three systems?

Let's first argue that, if we require that A and B are strictly in an EPR pair, then it is impossible for C to share any correlation with the qubits that form the EPR pair.

Example 3.5.1. *Let $\rho_{AB} = |\Psi\rangle\langle\Psi|_{AB}$ be an arbitrary pure state. Then since ρ_{AB} is pure its only nonzero eigenvalue is $\lambda_1 = 1$. Thus by Uhlmann's theorem any purification of ρ_{AB} must have the form $\rho_{ABC} = |\Psi\rangle\langle\Psi|_{AB} \otimes |\Phi\rangle\langle\Phi|_C$ for an arbitrary state $|\Phi\rangle_C$ of system C. But this is a pure state, whose Schmidt rank across the partition $AB : C$ is equal to 1: it is not entangled! If furthermore we take $|\Psi\rangle_{AB}$ to be an EPR state, then you can further compute that $\rho_{AC} = \frac{\mathbb{I}}{2} \otimes \rho_C$, meaning that not only C is uncorrelated with A, but from the point of view of C A looks maximally mixed, i.e. it completely random. The same holds for ρ_{BC}.* ∎

3.5.1 Quantifying monogamy

The previous example demonstrates monogamy of the maximally entangled EPR pair. What about more general states, could they demonstrate entanglement across three different parties? This is possible to some extent. For example, consider the three-qubit *GHZ state*

$$|GHZ\rangle_{ABC} = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle) \ .$$

Then you can verify that this state is entangled across any of the three possible partititions of the three qubits, $A : BC$, $AB : C$ or $AC : B$. However, the following exercise shows that this entanglement is *not* maximal as soon as one of the qubits is "dropped".

Exercise 3.5.1. *Computed the reduced density ρ_{AB} of the state $|GHZ\rangle_{ABC}$ on the first two qubits. Show that this reduced density is separable, by computing an explicit decomposition $\rho_{AB} = \sum_i p_i \rho_A^{(i)} \otimes \rho_B^{(i)}$.* ∎

As the exercise shows, the correlations in the GHZ state, when considering any given pair of qubits, are weaker than those of a maximally entangled state (indeed, they are not even entangled at all).

To quantify entanglement in mixed states more finely than the "entangled/not entangled" distinction, wne possibility is to use so-called *entanglement measures* $E(A : B)$. An entanglement measure is any function of bipartite density matrices that satisfies certain desirable properties. We already saw such a measure, the Schmidt rank; however it only applies to pure bipartite states. For states that are not pure the situation is much more complicated, and there is no standard entanglement measure that satisfies all the properties that we would like. Among these properties,

there is one which expresses monogamy as follows: for any tripartite density matrix ρ_{ABC} it requires that

$$E(A:B) + E(A:C) \leq E(A:BC). \tag{3.36}$$

One way to interpret this inequality is that, whatever the *total* entanglement that A has with B and C (right-hand side), this entanglement must split additively between entanglement with B and with C (left-hand side). You may think this is obvious — but in fact very few entanglement measures are known to satisfy the monogamy inequality (3.36)! Finding good measures of entanglement is an active area of research.

3.5.2 A three-player CHSH game

Another, more intuitive way of measuring monogamy is through the use of nonlocal games, such as the CHSH game that we discussed in Section 3.4. First consider a three-player variant of this game where Alice would be required to successfully play the CHSH game simultaneously with two different partners, Bob and Charlie. That is, Alice would be sent a random x, Bob a random y and Charlie a random z; they would have to provide answers a, b and c respectively such that $xy = a + b \mod 2$ *and* $xz = a + c \mod 2$. Can they do it? The fact, discussed in Example 3.5.1, that the EPR pair has no entangled extension to three parties should give you a hint that things are going to be difficult for Alice!

In fact it is possible to make an even stronger statement. Consider the following three-player variant of the CHSH game:

- The referee selects two of the three players at random, and sends each of them the message "You've been selected!", together with a label that indicates which player in the CHSH game they are supposed to "impersonate".
- The referee plays the CHSH game with the selected players, sending each of them a random question and checking their answers for the CHSH condition. The third player is completely ignored.

Now, what do you think is the players' maximum success probability in this game? For the case of classical players the answer should be clear: $3/4$. Indeed, there is nothing more or less they can do in this variant than in the original two-player CHSH. (Make sure you are convinced of this fact. What is an optimal strategy for the three players?)

What about quantum players? Can they win with probability $\cos^2(\pi/8)$? Why not? Let's think of a possible extension of the two-player strategy we saw in Section 3.4.1. First of all we need the three players, Alice, Bob and Charlie, to decide

on an entangled state to share. Given they know two of them are going to be asked to play CHSH, it is natural to set things up with three EPR pairs, one between Alice and Bob, another between Bob and Charlie, and the third between Alice and Charlie.

Now the game starts, and two players are told they are to play the game. However, the crucial point to observe is that each of the selected players is not told with whom they are to play the game! So, for instance Alice will know she has been selected, but will not be told who is her partner — Bob or Charlie. Which EPR pair is she going to use to implement her strategy?

It turns out there is no answer to this question: Alice is stuck! Although we won't do it here, it is possible to show that the optimal winning probability in the three-player CHSH game described above, for quantum players, is no larger than the classical optimum: $3/4$. This is a powerful demonstration of monogamy of entanglement, showing in particular that there is no nice extension of the EPR pair to a tripartite state — at least not one that allows any two of them to win the CHSH game! We will return to a similar manifestation of monogamy by analyzing a "tripartite guessing game" next week.

3.6 Important identities for calculations

Purification of states

Given any density matrix diagonalized as $\rho_A = \sum_i \lambda_i |\phi_i\rangle\langle\phi_i|_A$, a purification of A is

$$|\Psi\rangle_{AB} = \sum_i \sqrt{\lambda_i}\, |\phi_i\rangle_A\, |w_i\rangle_B\,, \tag{3.37}$$

for any set of orthonormal vectors $\{|w_i\rangle_B\}_i$.

Schmidt decomposition of bipartite pure states

Any bipartite pure state $|\Psi\rangle_{AB}$ can be written into the form

$$|\Psi\rangle_{AB} = \sum_{i=1}^{d} \sqrt{\lambda_i}\, |a_i\rangle_A\, |b_i\rangle_B\,, \tag{3.38}$$

where $\{|a_i\rangle_A\}_i, \{|b_i\rangle_B\}_i$ are orthonormal vector sets, and $\sum_{i=1}^{d} \lambda_i = 1$.

CHSH game winning probability

Consider Alice and Bob playing in a game, where questions $x, y \in \{0, 1\}$ are sent to them, and they respond with answers $a, b \in \{0, 1\}$ respectively. Alice and Bob win the game if $a + b \pmod 2 = x \cdot y$. The winning probability is given by

$$p_{\text{win}}^{\text{CHSH}} = \frac{1}{4} \sum_{x,y \in \{0,1\}} \sum_{\substack{a,b \\ a+b \\ \mod 2 = x \cdot y}} p(a, b|x, y)\,. \tag{3.39}$$

For any classical strategy, $p_{\text{win}}^{\text{CHSH}} = \frac{3}{4}$.
If Alice and Bob shares an EPR pair, then $p_{\text{win}}^{\text{CHSH}} = \cos^2 \frac{\pi}{8} \approx 0.85$.

Chapter 5

From imperfect information to (near) perfect security

This week we discuss *privacy amplification*. This task is an essential component of many cryptographic protocols; in particular it forms the final step in the quantum key distribution protocols we'll see in the coming weeks. Moreover, we'll see that privacy amplification can be achieved using a beautiful family of objects from theoretical computer science called *randomness extractors* — themselves well worth studying in their own right!

5.1 Privacy amplification

Let's start by introducing the task of privacy amplification. Imagine (as usual!) that Alice and Bob want to use cryptography to exchange messages securely. For this they have access to a classic public communication channel: they can send each other any messages they like, *but* the channel is public: the malicious eavesdropper Eve may be listening in on the whole communication. Our only cryptographic assumption on the channel is that it is *authenticated*, meaning that when Alice (or Bob) receives a message she has the guarantee that it came directly from Bob (or Alice). (We will return to the topic of authentication in Week 6; for the time being think of it as a convenient assumption that will usually be met in practice. We will also assume the channel is noiseless, which in practice is easily ensured by a proper use of error-correcting codes.)

Alice and Bob would like to use symmetric-key cryptography: they know (as you do!) that the one-time pad is unconditionally secure, so the only thing they need is to come up with a shared secret key. Moreover, Alice and Bob being old-time friends, they already have a lot of shared secrets, such as the flavor of the first

ice-cream cone they shared. By putting all these secrets together and translating them in a string of bits, they're pretty confident they can come up with some value, call it $x \in \{0,1\}^n$, that's fairly secret...but only "fairly" so. Unfortunately they're not fully confident about which parts of x can be considered a secret, and which may have leaked. Alice might have told her best friend Mary about the ice-cream. She definitely wouldn't have told Mary about her (embarrassing) all-time favorite cheeky cartoon, but then her little brother John might now about this. Is there a way for Alice and Bob to somehow "boil down" the secrecy that x contains, throwing away some of the bits but without knowing a priori which are secure and which may potentially have been leaked?

Answer: yes! This is precisely what privacy amplification will do for them. To describe the task more precisely, consider the following scenario. Two mutually trusting parties, Alice and Bob, each holds a copy of the same string of bits x, which we'll call a "weak secret". This secret is taken from a certain distribution p_x, which we can represent through a random variable X; later on we'll call X the "source". The distribution of X itself is not known, but the sample x is available to both parties. An eavesdropper has side information E that may be correlated to X; for example E could be the first bit of X, the parity of X, or an arbitrary quantum state ρ_x^E. Given this setup, the goal for Alice and Bob is to each produce the same string z, which could be shorter than x but must be such that the distribution of z (represented via a random variable Z) is (close to) uniform, even from the point of view of the eavesdropper.

To summarize using symbols, privacy amplification is the transformation:

$$\rho_{XE} \qquad \xmapsto{\mathrm{PA}_X \otimes \mathbb{I}_E} \qquad \rho_{ZE} \approx_\varepsilon \frac{\mathbb{I}_Z}{|Z|} \otimes \rho_E. \tag{5.1}$$

Of course this will only be possible under some assumption on X: for example if $X = E$ always there is not much we can do. Given what we've already learned, it's natural to measure the "potential for privacy amplification" of a source X through the min-entropy (equivalently, the guessing probability) $\mathrm{H}_{\min}(X|E)$, as this is a measure of "uncertainty" the eavesdropper has about X. But we're getting ahead of ourselves. First let's see how to perform a simpler but closely related task, *randomness extraction*. Then we'll see how to use this to achieve privacy amplification.

Before diving in, consider the following warm-up exercise:

Exercise 5.1.1. *Suppose that $X \in \{0,1\}^3$ is uniformly distributed, and $E = X_1 \oplus X_2 \in \{0,1\}$. Give a protocol for privacy amplification that outputs two secure bits (without any communication). What if $E = (X_1 \oplus X_2, X_2 \oplus X_3) \in \{0,1\}^2$, can you still do it? If not, give a protocol extracting just one bit.*

Suppose the eavesdropper is allowed to keep any two of the bits of X as side information. Give a protocol for Alice and Bob to produce a Z which contains a single bit that is always uniformly random, irrespective of which two bits of X are stored by the eavesdropper. How about an R that contains two bits — can they do it? ∎

5.2 Randomness extractors

In the task of randomness extraction there is a single party, Alice, who has access to an n-bit string x with distribution X. We call X the *source*. X is unknown, and it may be correlated to an additional system E over which Alice has no control. For example, E could contain some information about the way in which the source was generated, or some information that an adversary has gathered during the course of an earlier protocol involving the use of X. The only promise that is given to Alice is a lower bound on the min-entropy, $\mathrm{H}_{\min}(X|E) \geq k$. Alice's goal is to produce a new string Z that is close to uniform and uncorrelated with E. (As you can see, this problem is very similar to privacy amplification, but without the added complication of Alice having to coordinate with Bob!)

Now, of course Alice could dump X and create her own uniformly random Z, say by measuring a $|0\rangle$ qubit in the Hadamard basis. To make the problem interesting we won't allow any quantum resources to Alice. She also doesn't have that much freely accessible randomness — maybe she can get some, but it will be slow and costly. So Alice's goal is to leverage what she has to the best she can: she wants to *extract* randomness from X, not import it from some magical elsewhere!

5.2.1 Randomness sources

Let's see some concrete examples of sources X, and how it is possible to extract uniform bits from them.

I.I.D. sources

The simplest case of a randomness source is the i.i.d. source, where the term i.i.d. stands for *independent and identically distributed*. A (classical) i.i.d. source $X \in \{0,1\}^n$ has a distribution $\{p_x\}$ which has a product form: there is a distribution $\{p_0, p_1\}$ on a single bit such that for all $(x_1, \ldots, x_n) \in \{0,1\}^n$,

$$\Pr[X = (x_1, \ldots, x_n)] = \Pr[X_1 = x_1] \cdots \Pr[X_n = x_n] = p_{x_1} \cdots p_{x_n}.$$

Such sources are sometimes called *von Neumann* sources, since they were already considered by von Neumann. If you are curious about the history of randomness extraction, go look up the von Neumann extractor online!

Can we extract uniformly random bits from an i.i.d. source? As a warmup, let's consider how we could obtain a nearly uniform bit from a source such that each bit X_i is 0 with probability $p_0 = 1/4$ and 1 with probability $p_1 = 3/4$. Suppose we let $Z = X_1 \oplus X_2 \oplus \ldots \oplus X_n \in \{0,1\}$ be the parity of all n bits of X. Our goal is to show that $\Pr[Z = 0] \approx 1/2 \pm \epsilon$ for reasonably small ε.

- Let's first consider $n = 2$. How well does our strategy work? We can compute

$$\begin{aligned}\Pr[Z = 0] &= \Pr[X_1 = 0 \wedge X_2 = 0] + \Pr[X_1 = 1 \wedge X_2 = 1] \\ &= p_0^2 + p_1^2 = 1/16 + 9/16 = 0.625,\end{aligned}$$

  and using a similar calculation we find $\Pr[Z = 1] = 0.375$. Not quite uniform, but closer than what we started with!

- By doing the calculation for increasingly large values of n you will see that the trace distance ε of Z from a uniformly distributed random variable gets smaller and smaller. At what rate? Give a bound on ε as a function of n. Do you find our procedure efficient?

Independent bit sources

A slightly broader class of sources are *independent bit* sources. As their name suggests such sources are characterized by the condition that each bit is chosen independently; however the distribution could be different for different bits. Clearly, any i.i.d. source is also an independent bit source, but the converse does not hold.

Exercise 5.2.1. *Show that there exists an independent 2-bit source X such that $\Pr[X = (0,0)] = \Pr[X = (1,1)] = 3/16$, but there is no i.i.d. source satisfying the same condition.* ∎

It turns out that taking the parity of all the bits in the string generated by an independent bit source still results in a bit that is increasingly close to uniform as $n \to \infty$, provided each bit from the source is not fully biased to start with: $0 < \Pr[X_j = 0] < 1$ for all j.

Exercise 5.2.2. *Let X be an independent n-bit source such that $\delta < \Pr[X_j = 0] < 1 - \delta$ for some $\delta > 0$ and all $j \in \{1, \ldots, n\}$. Give an upper bound on the distance from uniform of the parity of the bits of x, as a function of the number of bits n of X and δ.* ∎

Bit-fixing sources

Bit-fixing sources are a special case of independent sources where each bit of X can be of one of two kinds only: either the bit is completely fixed, or it is uniformly random. For example, the three-bit source X such that $\Pr[X = (1, 0, 0)] = \Pr[X = (1, 1, 0)] = 1/2$, with all other probabilities being 0, is a bit-fixing source: the first bit is fixed to 1, the second is uniformly random, and the third is fixed to 0.

You can verify for yourselves that, just as for the previous two types of sources we considered, taking the parity of all bits from a bit-fixing source gives a uniformly random bit. This time, we do even better: as long as at least one of the bits from the source is not fixed, the parity is (exactly) uniformly random.

General sources

The randomness sources we just discussed all have something in common: they produce a string in which each bit is chosen *independently*. What if we relax this condition?

Consider a tricky example, called an *adversarial* bit-fixing source: this is the same as a bit-fixing source, except the value taken by the fixed bits can depend on the previous bits. For example, the three-bit source X such that $\Pr[X = (1, 0, 0)] = \Pr[X = (1, 1, 1)] = 1/2$, with all other probabilities being 0, is an adversarial bit-fixing source: the first bit is fixed to 1, the second is uniformly random, and the third is fixed to, either 0 if the second was a 0, or 1 if the second was a 1. To see that this kind of source can be much more tricky, first check that our earlier choice of Z as the parity of all the bits of X no longer works on the example. However, the parity of the first two, or the first and last, bits does work on that example. Nevertheless, show that for any fixed choice of a subset of bits, there exists an adversarial bit-fixing source such that only one bit is fixed, but nevertheless the parity of the bits in the chosen subset is a constant — arbitrarily far from uniform!

As you can imagine there is a whole jungle of possible kinds of sources. How do we classify them? For the purposes of extracting randomness, we aim to measure the inherent uncertainty of the source, or in other words its *entropy*. It turns out that the min-entropy provides just the "right" measure of extractable randomness, in a precise way that we'll soon see.

Definition 5.2.1. *A random variable X is a k-source if* $\mathrm{H}_{\min}(X) \geq k$.

Before we move on, we should realize there is something crucial missing from this definition. Remember we're going to apply the idea of randomness extraction to a cryptographic task, privacy amplification. But we forgot to account for the

eavesdropper! The process of randomness extraction is not going to happen in a void: we ought to take into account the possibility for an additional system E that may be correlated with X. Call E the *side information*. X is a classical string of bits, but E may be quantum. How do we model this? The proper way to do it is to introduce a cq state ρ_{XE}, which in general takes the form

$$\rho_{XE} = \sum_x |x\rangle\langle x|_X \otimes \rho_x^E,$$

where each ρ_x^E is positive semidefinite and $\mathrm{Tr}(\rho_{XE}) = \sum_x \mathrm{Tr}(\rho_x^E) = 1$. Using side information gives us a convenient way to model any source X as the result of an initially uniform string about which the adversary has gained partial information. For instance, you can think of a bit-fixing source as a uniform source correlated with a system E which contains some of the bits of x.

Exercise 5.2.3. *Let X be an independent source, where the i-th bit X_i has distribution $\{p_i, 1-p_i\}$. Show that there exists a pair of correlated random variables (Y, Z) on $\{0,1\}^n \times \{0,1\}^n$ such that Y is uniformly distributed in $\{0,1\}^n$ but for any $z \in \{0,1\}^n$ the random variable $V = Y_{|Z=z}$ is such that V_i has the same distribution as X_i if $z_i = 0$, and as $1 - X_i$ if $z_i = 1$.* ∎

Let's update our definition:

Definition 5.2.2. *A cq state ρ_{XE} is called a k-source if $\mathrm{H}_{\min}(X|E) \geq k$.*

Can we construct extraction procedures that produce uniformly random bits from any k-source, without knowing anything else about the source?

5.2.2 Strong seeded extractors

In all examples we've seen so far we applied a fixed function, call it Ext, to the source X; for example we considered $\mathrm{Ext}(X) = X_1 \oplus \cdots \oplus X_n$. Such a function is known as a deterministic extractor, meaning that is it just one fixed function $Z = \mathrm{Ext}(X)$ that does not introduce any randomness beside what is already present in X.

Ideally we'd like to show that it is possible to extract randomness from any k-source using such a deterministic function. Unfortunately this is not possible: there is no fixed deterministic procedure that can be used to extract even a *single* bit of randomness from every possible k-source, even when k is almost maximal, $k = n - 1$! This is a bit disappointing, but let's understand why.

Lemma 1. *For any function $\mathrm{Ext} : \{0,1\}^n \to \{0,1\}$ there exists an $(n-1)$-source X such that $\mathrm{Ext}(X)$ is constant.*

Proof. Let $b \in \{0,1\}$ be such that $|S_b| \geq 2^n/2 = 2^{n-1}$ with $S_b = \{x \mid \mathrm{Ext}(x) = b\}$. Note that there must exist such a b. Choose a subset $S' \subseteq S_b$ such that $|S'| = 2^{n-1}$. Define X by the following distribution:

$$p_x = \begin{cases} 1/2^{n-1} & \text{if } x \in S' \,, \\ 0 & \text{otherwise} \,. \end{cases} \tag{5.2}$$

Clearly, $\mathrm{H}_{\min}(X) = n-1$, but $\mathrm{Ext}(X) = b$ is a constant! □

Have we reached the end of the road — are we stuck to designing special-purpose functions which only work for this or that special kind of source, as we did with independent sources? Luckily there is a way out, but we're going to need an additional resource: a little extra randomness. This extra randomness will be called the *seed* of the extractor; think of it as a second input $Y \in \{0,1\}^d$ to which Alice has access and is promised to be uniformly random and independent from X and E. This gives us the notion of a *seeded extractor*:

Definition 5.2.3. *A* (k,ε)-weak seeded randomness extractor *is a function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *such that for any* k*-source* ρ_{XE}*,*

$$D\Big(\rho_{\mathrm{Ext}(X,Y)E}, \frac{\mathbb{I}}{2^m} \otimes \rho_E\Big) \leq \varepsilon \,, \tag{5.3}$$

where $Y \sim U_d$ *is uniformly distributed and independent from* X *and* E*, and*

$$\rho_{\mathrm{Ext}(X,Y)E} = \sum_z |z\rangle\langle z|_Z \otimes \rho_z^E \qquad \textit{with} \qquad \rho_z^E = 2^{-d} \sum_y \sum_{x:\,\mathrm{Ext}(x,y)=z} \rho_x^E .$$

If the seed is perfectly uniform, why don't we just return it as our output: define $\mathrm{Ext}(X,Y) = Y$? Well, this satisfies the definition. So maybe there is something wrong with the definition? Remember that our goal is to extract randomness from X, and that additional uniform randomness should not be considered free. So we want to keep Y as small as possible, even though X, and k, could be very large, in which case we'd like to maintain a long output (large m) with only a small help from the seed (small d).

A better answer considers our ultimate goal of privacy amplification. Remember that in that setting Alice and Bob share a weak secret X, and they want to produce a uniformly random secret R. Our solution of an extractor outputting its seed would be similar to asking Alice and Bob to throw away their initial secret X and share a fresh random string Y. But they only have access to a public communication channel, how would they agree on the same Y without the eavesdropper learning it as well?

This motivates a stronger definition of extractor, which is the one we'll use from now on:

Definition 5.2.4. *A* (k, ε)-strong seeded randomness extractor *is a function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *such that for any* k*-source* ρ_{XE}*,*

$$D\big(\rho_{\mathrm{Ext}(X,Y)YE}, \frac{\mathbb{I}}{2^m} \otimes \rho_{YE}\big) \leq \epsilon \,, \tag{5.4}$$

where $Y \sim U_d$ *is uniformly distributed and independent from* X *and* E*.*

Before we start trying to construct strong extractors, let's consider the notion of k-source a little more closely. Why do we think that the min-entropy provides the right way to quantify the amount of randomness that can be extracted from a given source?

Let's first argue informally that the min-entropy is an upper bound on the amount of extractable randomness: there is no strong extractor that has output length more than $\mathrm{H}_{\min}(X|E)$. To see why this is the case, recall that $\mathrm{H}_{\min}(X|E) = -\log P_{\mathrm{guess}}(X|E)$. Suppose now that we apply some function f to X. How hard is it to guess $f(X)$ given E, i.e., what's $P_{\mathrm{guess}}(f(X)|E)$? Clearly, since one way to guess $f(X)$ is to guess X, and then apply f to our guess, we have $P_{\mathrm{guess}}(f(X)|E) \geq P_{\mathrm{guess}}(X|E)$. However, this is equivalent to

$$\mathrm{H}_{\min}(f(X)|E) \leq \mathrm{H}_{\min}(X|E) \,. \tag{5.5}$$

This means that also the output of the extractor, which for fixed seed y is obtained as a function $f(X) = \mathrm{Ext}(X, y)$, must have min-entropy at most $\mathrm{H}_{\min}(X|E)$, which implies that the output $\mathrm{Ext}(X, Y)$, conditioned on $Y = y$, can be uniform on at most $\mathrm{H}_{\min}(X|E)$ bits!

Now, how about a converse: does there exist a strong extractor that can extract approximately $\mathrm{H}_{\min}(X|E)$ bits from *any* k-source ρ_{XE}? The answer turns out to be yes, and we're going to see how this can be done in the next section.

5.3 An extractor based on hashing

Much research has gone into constructing randomness extractors, and they have found many applications throughout computer science and mathematics. The quality of an extractor is measured by the parameters it achieves, and different applications require different trade-offs. The main targets consist in extracting as much randomness as possible (large m) using the smallest possible seed (small d) and with the best possible error (small ε), all from arbitrary sources with min-entropy (at least) k.

By using a probabilistic argument (select a function Ext at random from all possible functions, and fix it to be the extractor), for any given input length k

and min-entropy k the best trade-offs that can be achieved are seed length $d = \log(n-k) + 2\log(1/\varepsilon) + O(1)$ and output length $m = k + d - 2\log(1/\varepsilon) + O(1)$ [**radhakrishnan2000bounds**]. Moreover, there are efficient constructions known that achieve essentially both parameters simultaneously! Rather than aiming for optimal, but often intricate, constructions, here we will focus on a simple construction which nevertheless achieves very good parameters for the application we have in mind (privacy amplification!).

Going back to the intuition we developed on the examples, we saw that taking the parity of a random subset of the bits of the source often (but not always) provides a good way to extract a bit of randomness. In this case we can think of the seed of the extractor as specifying the subset of bits whose parity is taken. We could repeat this procedure to extract even more bits, each chosen as the parity of a different random subset. It is a good exercise to show that this procedure works, but it has one major drawback: it is excessively costly in terms of seed length, requiring an investment of approximately n bits of randomness (to specify the subset of bits whose parity is taken) for each new bit produced!

Let's see how we can do better. For this we'll have to make a little detour and learn about certain families of hash functions.

5.3.1 Two-universal families of hash functions

Informally, a hash function is a function that maps long strings to shorter strings, with the property that the output of the hash functions tends to be "well-distributed". What this means depends on the application we have in mind for the hash function — indeed, the term "hash function" can be interpreted in many different ways, with the only standard requirement, as its name indicates, being that a hash function should not increase the length of its input! An additional reasonable requirement, which formalizes the "well-distributed" aspect of the output of a hash function, is the following:

Definition 5.3.1 (1-universal family)**.** *A family of hash functions $\mathcal{F} = \{f : \{0,1\}^n \to \{0,1\}^m\}$, where $m \leq n$, is called* 1-universal *if for every string $x \in \{0,1\}^n$ and $z \in \{0,1\}^m$ we have*

$$\Pr_{f \in \mathcal{F}}[f(x) = z] = \frac{1}{2^m} \,. \tag{5.6}$$

It is worth reading this definition carefully: in (5.6) both x and z are fixed, and the probability is taken over a uniformly random function from the family. The condition is equivalent to saying that for any fixed x, the random variable $F(x)$, where F is uniformly distributed over all f in $\mathcal{F}$, in uniformly distributed in $\{0,1\}^m$. Let's see an example of a 1-universal family of hash functions.

Exercise 5.3.1. *For any $y \in \{0,1\}^n$ let $f_y : \{0,1\}^n \to \{0,1\}^n$ be defined by $f_y(x) = x \oplus y$, where the parity is taken bitwise. Show that the family of functions $\mathcal{F} = \{f_y,\ y \in \{0,1\}^n\}$ is 1-universal.* ∎

You may want to convince yourself that a family of 1-universal hash functions is already sufficient to construct a *weak* seeded extractor: use the seed to select a random function from the family, and output the value of the function evaluated at the source. The property of 1-universality ensures that the output will be uniformly distributed, even if the input is fixed. However, recall our earlier criticism: in this case it is apparent that we are "cheating", and that all the randomness is coming from the seed. Indeed, it turns out that the property of 1-universality is not sufficient to obtain a *strong* seeded extractor. We'll need the following stronger property, first introduced by Carter and Wegman:

Definition 5.3.2 (2-universal family)**.** *A family of hash functions $\mathcal{F} = \{f : \{0,1\}^n \to \{0,1\}^m\}$ is called* 2-universal *if for every two strings $x, x' \in \{0,1\}^n$ with $x \neq x'$, and any two $z, z' \in \{0,1\}^m$, we have*

$$\Pr_{f \in \mathcal{F}}[f(x) = z \ \wedge\ f(x') = z'] = \frac{1}{2^{2m}}\ . \tag{5.7}$$

Condition (5.7) in the definition would be satisfied if $f(x)$ and $f(x')$ were *jointly* chosen uniformly and independently at random in $\{0,1\}^m$. This is a stronger condition than (5.6): we now require that the pair of random variables $(F(x), F(x'))$, for F uniformly distributed over $f \in \mathcal{F}$, are jointly uniform (as an exercise, verify that the family of hash functions from Exercise 5.3.1 is *not* 2-universal).

You can check that for any $m \leq n$ the set of all possible functions $f : \{0,1\}^n \to \{0,1\}^m$ is 2-universal. But it is too big a set: it has size $|\mathcal{F}| = 2^{m2^n}$, so that selecting a function at random from the set would require a seed length $d = m2^n$! Let's see a much more efficient construction.

Let $q = 2^n$ and $\mathbb{F}_q$ the finite field with 2^n elements. (If you have never seen this field before, the details of its construction will not be matter to us, but you may still want to check it out online! The multiplication rule is *not* the same as multiplication over the integers, $\bmod 2^n$.) For any $(a,b) \in \mathbb{F}_q^2$ let

$$f_{a,b} : \mathbb{F}_q \to \mathbb{F}_q, \qquad f_{a,b}(x) = ax + b,$$

where addition and multiplication are done in $\mathbb{F}_q$. Then $\mathcal{F} = \{f_{a,b},\ (a,b) \in \mathbb{F}_q^2\}$ is a 2-universal family of only $q^2 = 2^{2n}$ hash functions. To show this we need to verify that equation (5.7) from the definition holds. So let's fix distinct $x \neq x' \in \mathbb{F}_q$ and two $z, z' \in \mathbb{F}_q$. What is the probability, over a uniformly random choice of (a,b), that $f_{a,b}(x) = z$ and $f_{a,b}(x') = z'$? The two conditions are equivalent to

$ax+b = z$ and (taking the difference) $a(x'-x) = z'-z$, thus $a = (z'-z)/(x'-x)$, where the condition $x \neq x'$ and the fact that $\mathbb{F}_q$ is a field allows us to perform the division. This equation determines a unique possible value for a. Moreover, once a is fixed there is a unique possible value for b: $b = z - ax$ (this shouldn't be a surprise, since we started with two linear equations and two unknowns). Out of 2^{2m} possibilities, we end up with a single one: $\Pr_{a,b}[f_{a,b}(x) = z \wedge f_{a,b}(x') = z'] = 2^{-2m}$, as desired.

One last technicality: recall that our goal was to construct a 2-universal family of functions $f : \{0,1\}^n \to \{0,1\}^m$, for arbitrary n and $m \leq n$, whereas what we managed to construct so far are functions from $\mathbb{F}_q \to \mathbb{F}_q$. Since $|\mathbb{F}_q| = q = 2^n$ the domain of f can be identified with $\{0,1\}^n$ in an arbitrary way. The range of f may be bigger than $\{0,1\}^m$, but there is a simple solution: throw away the last $(n-m)$ bits of $f(x)$! I'll let you verify that this works, i.e. it preserves the property of 2-universality.

5.3.2 The two-universal extractor

Equipped with an arbitrary family of 2-universal hash functions, we define an extractor as follows.

Definition 5.3.3 (2-universal extractor)**.** *Let $\mathcal{F} = \{f_y : \{0,1\}^n \to \{0,1\}^m, y \in \{0,1\}^d\}$ be a 2-universal family of hash functions such that $|\mathcal{F}| = 2^d$. The associated 2-universal extractor is*

$$\mathrm{Ext}_{\mathcal{F}} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m, \qquad \mathrm{Ext}_{\mathcal{F}}(x,y) = f_y(x).$$

One way to think of $\mathrm{Ext}_{\mathcal{F}}$ is as using its seed y to select a function from the family $\mathcal{F}$ uniformly at random, and then returning the output of the function when evaluated on the source X.

How good is this extractor? The key result required to analyze it is known as the *leftover hash lemma*. It was first proven by Impagliazzo, Levin, and Luby for the case when there is no side information E, and later extended to the case of quantum E by Renner. Here is a statement of the lemma when there is no side information.

Theorem 5.3.1 (Leftover hash lemma)**.** *Let n and $k \leq n$ be arbitrary integers, $\varepsilon > 0$, $m = k - 2\log(1/\varepsilon)$, and $\mathcal{F} = \{f : \{0,1\}^n \to \{0,1\}^m\}$ a 2-universal family of hash functions. Then the 2-universal extractor $\mathrm{Ext}_{\mathcal{F}}$ is a (k,ε)-strong seeded randomness extractor.*

In the previous section we saw how to construct a 2-universal family with 2^{2n} functions, meaning that the seed length of the two-universal extractor is $2n$. This is

much longer than the optimal length $d \approx O(\log(n/\varepsilon))$, and it can be a drawback in some applications for which the randomness required to produce the seed is particularly costly. However, for our application to privacy amplification, and especially later to quantum key distribution, it is not a significant limitation. Much more important for us is the dependence of the output length on the initial min-entropy, which will ultimately govern the length of key that we are able to produce. In this respect the two-universal construction is essentially optimal, a good reason to use it!

5.3.3 Analysis with no side information

We first prove the leftover hash lemma in the case when there is no side information, stated in Theorem 5.3.1. This will be a good warm-up for the general case, which will follow the same structure.

The proof proceeds in two steps. In the first step we reduce our ultimate goal, bounding the error of the extractor, i.e. the trace distance between the extractor's output and the uniform distribution, to bounding a different quantity called the *collision probability*. In the second step we show that the collision probability is sufficiently small to imply the desired bound on the error of the extractor.

(i) From trace distance to collision probability. Our goal is to bound $D(\rho_{\mathrm{Ext}(X,Y)Y}, 2^{-(m+d)}\mathbb{I})$, where X has min-entropy at least k and Y is uniformly distributed over d-bit strings. The joint distribution of $(Z = \mathrm{Ext}(X,Y), Y)$ is given by

$$p_{zy} = \Pr[(\mathrm{Ext}(X,Y),Y) = (z,y)] = 2^{-d} \sum_{x:f_y(x)=z} p_x. \tag{5.8}$$

Using the definition of the trace distance, we get

$$\begin{aligned} D(\rho_{\mathrm{Ext}(X,Y)Y}, 2^{-(d+m)}\mathbb{I}) &= \frac{1}{2}\sum_{z,y}\Big|2^{-d}\sum_{x:f_y(x)=z} p_x - 2^{-d-m}\Big| \\ &\leq 2^{\frac{m}{2}-1}\Big(2^{-d}\sum_{z,y}\Big|\sum_{x:f_y(x)=z} p_x - 2^{-m}\Big|^2\Big)^{1/2} \\ &= 2^{\frac{m}{2}-1}\Big(2^{d}\sum_{z,y} p_{zy}^2 - 2^{-m}\Big)^{1/2}, \end{aligned}$$

where for the second line we applied the Cauchy-Schwarz inequality. This completes our first step. The quantity $CP(ZY) = \sum_{z,y} p_{zy}^2$ is called the collision probability of (Z,Y), and we turn to bounding it next.

(ii) A bound on the collision probability. Using the definition (5.8) and expanding the square,

$$\begin{aligned}\sum_{z,y} p_{zy}^2 &= 2^{-2d} \sum_{y,z} \sum_{\substack{x,x': \\ f_y(x)=f_y(x')=z}} p_x p_{x'} \\ &= 2^{-2d} \sum_{y,z} \Big(\sum_{\substack{x\neq x': \\ f_y(x)=f_y(x')=z}} p_x p_{x'} + \sum_{x:\, f_y(x)=z} p_x^2 \Big) \\ &= 2^{-(d+m)} \sum_{x\neq x'} p_x p_{x'} + 2^{-d} \sum_x p_x^2 \\ &\leq 2^{-(d+m)} + 2^{-(d+k)}.\end{aligned}$$

Here the crucial step is in bounding the summation over $x \neq x'$ when going from the second to the third line: we are using the property of 2-universality to argue that for any $x \neq x'$ there is a fraction exactly 2^{-m} of all f_y that map both x and x' to the same value. To bound the second term in going from the second-last to last lines we used $\sum_x p_x^2 \leq \max_x p_x = 2^{-\mathrm{H}_{\min}(X)}$ and the assumption $\mathrm{H}_{\min}(X) \geq k$.

Plugging this back into the bound on the trace distance from (i) we obtain

$$D(\rho_{\mathrm{Ext}(X,Y)Y}, 2^{-(d+m)}\mathbb{I}) \leq 2^{\frac{m-k}{2}-1},$$

proving the lemma.

5.3.4 The pretty good measurement and quantum side information

We would like to extend the proof in the previous section to the case where the source X is correlated with some quantum side information E, that is, $\rho_{XE} = \sum_x |x\rangle\langle x| \otimes \rho_x^E$ is an arbitrary cq state such that $\mathrm{H}_{\min}(X|E) \geq k$. Before diving into this, let's make a small detour by considering the related problem of optimally distinguishing between a set of quantum states.

The pretty-good measurement

Let $\rho_{XE} = \sum_x |x\rangle\langle x| \otimes \rho_x^E$ be a cq state. What is the optimal probability with which Eve, holding the quantum system E, can successfully guess x? We've seen this problem already: the answer is captured by the guessing probability,

$$P_{\mathrm{guess}}(X|E)_\rho = \max_{\{M_x\}} \sum_x \mathrm{Tr}(M_x \rho_x^E), \tag{5.9}$$

where the maximum is taken over all POVM $\{M_x\}$ on E. But what is the best POVM? If $x \in \{0,1\}$ takes only two values you've already seen the answer: in this case we can write

$$\begin{aligned}\mathrm{Tr}(M_0\rho_0^E) + \mathrm{Tr}(M_1\rho_1^E) &= \mathrm{Tr}\left(\frac{M_0+M_1}{2}\cdot\left(\rho_0^E+\rho_1^E\right)\right) + \mathrm{Tr}\left(\frac{M_0-M_1}{2}\cdot\left(\rho_0^E-\rho_1^E\right)\right)\\ &\leq \frac{1}{2} + \frac{1}{2}D(\rho_0^E,\rho_1^E),\end{aligned}$$

and moreover the last inequality is an equality if M_0 and M_1 are the projectors on the positive and negative eigenspaces of the Hermitian matrix $\rho_0^E - \rho_1^E$ respectively.

When $|X| > 2$ unfortunately the situation is a bit more murky. The problem of finding the optimal measurement can be solved efficiently with a computer by expressing the optimization problem (5.9) as a *semidefinite program*, a generalization of linear programs for which there are efficient algorithms. But what we'd really like is a nice, clean mathematical expression for what the optimal measurement is, so that we can work with it in our proofs! No such simple closed form is known. However, what we can do is find a simple measurement that always achieves *close* to the optimum: the *pretty-good measurement*.

So what is this "pretty-good" measurement? To get some intuition first consider the case where the states ρ_x^E are perfectly distinguishable; for example $\rho_x^E = p_x |x\rangle\langle x|$ is simply a classical copy of X. Then it is clear what we should do: measure in the computational basis, and recover x! Observe that in this case the POVM elements M_x are directly proportional to ρ_x: we can think of the states as "pointing" in some direction correlated with x, and it is natural to make a measurement along that direction.

Can we generalize this idea? Let's try defining $M_x = \rho_x^E$. This is positive semidefinite, so it satisfies the first condition for a POVM. However, $\sum_x M_x = \sum_x \rho_x^E = \rho^E$ is not necessarily the identity, as required by the second condition. The solution? Normalize!

Definition 5.3.4. *Given a collection of positive semidefinite matrices $\{\rho_x\}$, the* pretty-good measurement *(PGM) associated to the collection is the POVM with elements*

$$M_x = \rho^{-1/2}\,\rho_x\,\rho^{-1/2},$$

where $\rho = \sum_x \rho_x$ and the inverse is the Moore-Penrose pseudo-inverse, i.e. we use the convention $0^{-1} = 0$.

Note how we dealt with division by zero in the definition. Defining division by zero may seem odd, but this convention makes sense in the context of linear operators. If ρ is orthogonal to some subspace, i.e. it is an eigenspace of eigenvalue

0, then the pseudo-inverse ρ^{-1} should also be orthogonal to that subspace. A useful property of this convention is that it makes it so that if P is an orthogonal projection and $P\rho P$ is invertible, then $(P\rho P)^{-1} = P\rho^{-1}P$.

How well does the pretty-good measurement compare to the optimal guessing measurement? Let $\{N_x\}$ be an optimal guessing POVM for Eve. Then by definition

$$\begin{aligned} P_{\text{guess}}(X|E) &= \sum_x \operatorname{Tr}\left(N_x \rho_x^E\right) \\ &= \sum_x \operatorname{Tr}\left((\rho^{1/4} N_x \rho^{1/4})(\rho^{-1/4} \rho_x^E \rho^{-1/4})\right) \\ &\leq \Big(\sum_x \operatorname{Tr}\left(\rho^{1/2} N_x \rho^{1/2} N_x\right)\Big)^{1/2} \Big(\sum_x \operatorname{Tr}\left(\rho^{-1/2} \rho_x^E \rho^{-1/2} \rho_x^E\right)\Big)^{1/2} \\ &\leq \big(\mathrm{PGM}(X|E)\big)^{1/2}, \end{aligned}$$

where

$$\mathrm{PGM}(X|E) = \sum_x \operatorname{Tr}(M_x \rho_x^E) = \sum_x \operatorname{Tr}\left(\rho^{-1/2} \rho_x \rho^{-1/2} \rho_x\right) \tag{5.10}$$

is the success probability of the PGM in the guessing task. The second and third lines are the most important here. To go from the first to the second line we inserted factors $\rho^{1/4}$ and $\rho^{-1/4}$ that cancel each other out (using cyclicity of the trace), but are important for normalization. To go from the second to the third line we used the Cauchy-Schwarz inequality twice: first, for each x we apply a matrix version of the inequality,

$$\big|\operatorname{Tr}(AB)\big| \leq \big(\operatorname{Tr}(AA^\dagger)\big)^{1/2} \big(\operatorname{Tr}(BB^\dagger)\big)^{1/2}, \tag{5.11}$$

with $A = \rho^{1/4} N_x \rho^{1/4}$ and $B = \rho^{-1/4} \rho_x^E \rho^{-1/4}$; and second, we apply the usual version

$$\Big|\sum_x a_x b_x\Big| \leq \Big(\sum_x a_x^2\Big)^{1/2} \Big(\sum_x b_x^2\Big)^{1/2},$$

valid for any real a_x and b_x (here $a_x = \operatorname{Tr}(\rho^{1/2} N_x \rho^{1/2} N_x)$ and $b_x = \operatorname{Tr}(\rho^{-1/2} \rho_x^E \rho^{-1/2} \rho_x^E)$). Finally to get to the last line we used $\sum_x N_x = \mathbb{I}$ to bound the first term, and the definition of the pretty-good measurement for the second.

Proof of the leftover hash lemma with quantum side information

The proof follows the same structure as the proof we saw for the case with no side information, but it is slightly more involved technically. We will use the following inequality: for any positive Hermitian σ and positive semidefinite τ such that

$\mathrm{Tr}(\tau) = 1$ and the support of τ contains the support of σ,

$$\mathrm{Tr}\left(|\sigma|\right) \leq \mathrm{Tr}\left((\tau^{-1/4}\sigma\tau^{-1/4})^2\right)^{1/2}. \tag{5.12}$$

To prove the inequality, observe that

$$\begin{aligned}\mathrm{Tr}\left(|\sigma|\right) &= \mathrm{Tr}\left(\tau^{1/4}\tau^{-1/4}|\sigma|\tau^{-1/4}\tau^{1/4}\right)\\ &= \mathrm{Tr}\left(\tau^{1/4}|\tau^{-1/4}\sigma\tau^{-1/4}|\tau^{1/4}\right)\\ &= \mathrm{Tr}\left(|\tau^{-1/4}\sigma\tau^{-1/4}|\tau^{1/2}\right).\end{aligned}$$

Here the second line is obtained by computing the trace in the eigenbasis of $\tau^{-1/4}\sigma\tau^{-1/4}$; see [**renner2008security**] for details of the calculation. To conclude the proof of (5.12), apply (5.11) with the choice $A = \tau^{-1/4}\sigma\tau^{-1/4}$ and $B = \tau^{1/2}$.

(i) From trace distance to collision probability. Our goal is to bound $D(\rho_{\mathrm{Ext}(X,Y)YE}, 2^{-(m+d)}\mathbb{I}\otimes \rho_E)$, where Y is uniformly distributed and X is such that $\mathrm{H}_{\min}(X|E) \geq k$. We can write

$$\rho_{\mathrm{Ext}(X,Y)YE} \;=\; \sum_{z,y} |z\rangle\langle z| \otimes |y\rangle\langle y| \otimes \rho_{zy}, \qquad \text{with} \qquad \rho_{zy} = 2^{-d} \sum_{x:\, f_y(x)=z} \rho_x.$$

Note that our normalization makes it so that

$$\sum_{z,y} \mathrm{Tr}(\rho_{zy}) \;=\; 2^{-d} \sum_{x,y} \mathrm{Tr}(\rho_x) \;=\; \mathrm{Tr}(\rho) \;=\; 1.$$

Since the state $\rho_{\mathrm{Ext}(X,Y)YE}$ is a ccq state, using the definition of the trace distance we can expand

$$\begin{aligned}D(\rho_{\mathrm{Ext}(X,Y)YE}, 2^{-(d+m)}\mathbb{I}\otimes\rho_E) &= \frac{1}{2}\sum_{z,y} \left\|\rho_{zy} - 2^{-(d+m)}\rho\right\|_1\\ &\leq 2^{\frac{m+d}{2}-1}\Big(2^{-(m+d)}\sum_{z,y}\mathrm{Tr}\left((\rho^{-1/4}(\rho_{zy}-2^{-m}\rho)\rho^{-1/4})^2\right)\Big)^{1/2}\\ &= 2^{\frac{m}{2}-1}\Big(2^{d}\sum_{z,y}\mathrm{Tr}\left(\rho_{zy}\rho^{-1/2}\rho_{zy}\rho^{-1/2}\right) - 2^{-m}\Big)^{1/2},\end{aligned}$$

where for the second line we first applied (5.12) for each (y,z) with $\sigma = \rho_{zy} - 2^{-(d+m)}\rho$ and $\tau = \rho$, and then the usual Cauchy-Schwarz inequality. Do you recognize the expression in the last line? Using the notation from (5.10), we have

$$\mathrm{PGM}(Z|YE) \;=\; 2^d \sum_z \mathrm{Tr}\left(\rho_{zy}\rho^{-1/2}\rho_{zy}\rho^{-1/2}\right),$$

so the sequence of equations above show that

$$D(\rho_{\mathrm{Ext}(X,Y)YE}, 2^{-(d+m)}\mathbb{I}\otimes\rho_E) \leq 2^{\frac{m}{2}-1}\big(\mathrm{PGM}(Z|YE) - 2^{-m}\big)^2.$$

We have thus managed to relate the distance from uniform to the advantage of the pretty good measurement over random guessing (that would succeed with probability 2^{-m}). We can understand this step of the proof as a reduction from arbitrary attacks of an adversary to the extractor, whose optimal success probability is expressed in the first line, to attacks of a very specific form, where the adversary, given a sample (z, y), measures its side information using the pretty-good measurement associated with the family of states $\{\rho_{zy}\}$. The square root factor on the third line expresses the fact that the pretty-good measurement is quadratically far from optimal. What is the point of losing this square root? The pretty-good measurement has a crucial advantage, that we are going to use in the second step of the proof: it has a form of "linearity" in the sense that the PGM operators associated with the family of states $\{\rho_{zy}\}$ can be obtained by summing up PGM operators associated with the states $\{\rho_x\}$. Let's see how this works in our favor.

(ii) A bound on the collision probability. Proceeding exactly as in the case with no side information, we can calculate

$$\begin{aligned}
\mathrm{PGM}(Z|YE) - 2^{-m} &= 2^{-d}\sum_{y,z}\sum_{\substack{x,x':\\ f_y(x)=f_y(x')=z}} \mathrm{Tr}(\rho_x\rho^{-1/2}\rho_{x'}\rho^{-1/2}) - 2^{-m} \\
&= 2^{-d}\sum_{y,z}\Big(\sum_{\substack{x\neq x':\\ f_y(x)=f_y(x')=z}} \mathrm{Tr}(\rho_x\rho^{-1/2}\rho_{x'}\rho^{-1/2}) \\
&\qquad + \sum_{x:f_y(x)=z}\mathrm{Tr}(\rho_x\rho^{-1/2}\rho_x\rho^{-1/2})\Big) - 2^{-m} \\
&= 2^{-m}\sum_{x\neq x'}\mathrm{Tr}(\rho_x\rho^{-1/2}\rho_{x'}\rho^{-1/2}) + \sum_x \mathrm{Tr}(\rho_x\rho^{-1/2}\rho_x\rho^{-1/2}) - 2^{-m} \\
&\leq \mathrm{PGM}(X|E).
\end{aligned}$$

Using the 2-universal hashing property, we have managed to relate the advantage over random of the pretty good measurement in guessing Z, to the success probability of the pretty good measurement to guess X directly. But the last expression is, by assumption, at most $2^{-\mathrm{H}_{\min}(X|E)}$, since the guessing probability achieved from using the PGM cannot be the optimal one. Together with the bound proven in step (i) we finally obtain

$$D(\rho_{\mathrm{Ext}(X,Y)Y}, 2^{-(d+m)}\mathbb{I}) \leq 2^{\frac{m-k}{2}-1},$$

precisely the same bound as when there was no side information at all.

5.4 Solving privacy amplification using extractors

Back to cryptography...how do we use extractors to solve privacy amplification? By now you must have a good idea how this can be done. Let Ext be a (k, ε) strong seeded randomness extractor. Here is a simple protocol:

1. Alice and Bob share a weak secret X, which may be correlated with an eavesdropper holding quantum side information E.
2. Alice choses a random seed Y for the extractor, and computes $R_A = \mathrm{Ext}(X, Y)$. She sends Y to Bob over a public communication channel.
3. Upon receiving Y, Bob sets $R_B = \mathrm{Ext}(X, Y)$.

First note that this protocol is always correct: Alice and Bob output the same string, $R_A = R_B$. Is it secure? Remember the criterion (5.1) we introduced to define security of privacy amplification. Note also that here, at the end of the protocol, Eve has access to her original side information E, but also to any communication exchanged over the public channel: precisely the seed Y. So the condition becomes

$$X : \mathrm{H}_{\min}(X|E)_\rho \geq k \quad \overset{\mathrm{PA}}{\longmapsto} \quad R = \mathrm{Ext}(X, Y) : \rho_{RYE} \approx_\varepsilon \frac{\mathbb{I}_R}{|R|} \otimes \rho_{YE},$$

which is precisely the requirement of a (k, ε) strong extractor! All the pieces have come into place: by instantiating the extractor with the 2-universal extractor based on the 2-universal family of hash functions from Section 5.3.1 you now have a complete construction of a secure one-way protocol for privacy amplification. This will be crucially used in our quantum key distribution protocols.