13 Photons for Quantum Information

This chapter deals with quantum aspects of the transfer of information. It starts with some discussion of basic tools and then describes tasks that cannot be performed classically but can be performed quantum mechanically. It then gives an introduction to some notions and topics in classical and quantum information theory, which may be useful for studying the growing body of research literature in quantum information theory.

13.1 Photons as qubits

13.1.1 Photons

Photons are certainly among the most attractive systems for storing quantum information: they can be generated in (almost) arbitrary quantities, with well defined properties like energy, momentum and polarization. Optical components like mirrors, beam splitters and retardation plates can execute unitary transformations on the photons with high precision. Quantum algorithms can therefore be implemented relatively easily in optical setups that use only linear optics [172, 173, 142]. Readout is relatively straightforward: the state of individual photons can be determined with almost certainty, at least for suitable wavelengths.

The internal degrees of freedom span a 2-dimensional Hilbert space, as required for qubits. The computational basis states are typically chosen as orthogonal polarization states - either linear or circular ones. Photons can be initialized into these states and the polarizations can be controlled by linear (or nonlinear) optical elements such as waveplates. Readout is achieved by sending photons through polarization-selective beam splitters onto single-photon detectors.

In addition to the internal degree of freedom, the position or the path of a photon are also important degrees of freedom. They can serve as qubits, and they are essential for many tasks, such as detection, where the polarization state is converted into different paths for separating the qubit states in space.

These attractive properties have motivated a vast range of applications. The most advanced ones are in (quantum-) communication, information processing and sensing.

For some tasks, it is not necessary to use single photons. Instead, the quantum information can be stored in states of the electromagnetic field that are defined by continuous variables[415]. These systems have infinite-dimensional Hilbert spaces and use use squeezed states [416] or coherent[1] states [418] of light for encode the qubits: in the latter case, the logical states are $|0_L\rangle = |\alpha\rangle$ and $|1_L\rangle = |-\alpha\rangle$, which are almost orthogonal if $\alpha > 2$.

13.1.2 Working with single photons

Major technical challenges for the realization of this type of quantum computer (as well as for quantum communication and similar applications) is the efficient generation and detection

[1] Coherent states [417, 68] are superpositions of harmonic oscillator eigenstates $|n\rangle$, $|\alpha\rangle = \exp\left(-\frac{|\alpha|^2}{2}\right) \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle$, where α is an arbitrary complex number. No two coherent states are orthogonal to each other, but their scalar product decays rapidly with growing distance in the complex plane, $|\langle\alpha|\beta\rangle|^2 = e^{-|\alpha-\beta|^2}$. Coherent states minimize the Heisenberg uncertainty product, and squeezed states enjoy similar quasi-classical properties.

of individual photons [419, 420]. Single photon sources where photons arrive at known but random times have been available for some time [421, 422]. For the purpose of linear optics quantum computing (and many others as well), it would be advantageous to generate "single photons on demand" [423, 424, 425], i.e. at times when they are required by the algorithm.

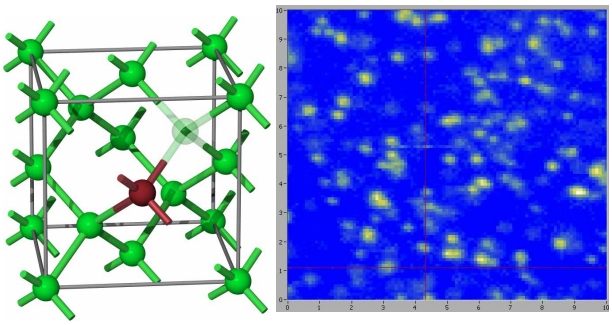

Figure 13.1: Structure of the diamond NV center and confocal scan of a diamond surface showing single emitters.

For scalable systems, solid-state sources may be advantageous. A stable solid-state source is the nitrogen-vacancy (NV)-center in diamond [426]. For wavelengths in the near infrared, quantum dots may become a useful source of single photons. They can be considered as artificial atoms that can be excited by a laser pulse and will re-emit the photon within a short period [427].

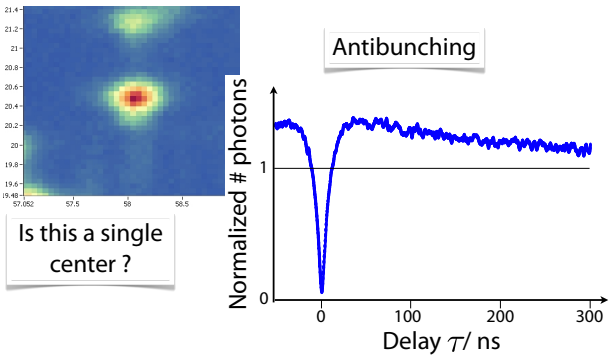

Figure 13.2: Correlation function of photons emitted from a single NV center showing antibunching.

The probability that pairs of photons are generated drops dramatically for single emitters, as shown in Fig. 13.2: the curve shows the probability of detecting a photon at a time τ after the detection of another photon. Clearly, the probability drops to (almost) zero for short delays. This is an immediate consequence of the single emitter: after emitting a photon, the emitter is in the ground state and cannot emit another photon before it is excited again.

Even more useful may be electrically pumped devices, where the electron-hole pairs are created by carrier injection [428, 429].

For applications in linear optics quantum computing or quantum cryptography, it is actually preferable to generate not single photons, but pairs of photons on demand. This was originally achieved by using atomic three-level systems [430]: in this case, the first photon is emitted from a doubly excited atomic state when the atom decays to an intermediate state. If the lifetime of this state is short enough, the system emits a second photon immediately after the first. This type of source has now been replaced by solid-state sources [431]: in these sources, blue or ultra-violet photons are down-converted into pairs of entangled photons with known directions of propagation and polarization. This type of source provides much higher rates of entangled photons than the atomic cascades used earlier. Alternatively, photon pairs can also be produced in semiconductor quantum dots by exciting biexcitons [432, 433]. The resulting photon-pairs have highly correlated polarization.

Detection of single photons is possible using avalanche photo diodes (APDs) or photomultiplier tubes (PMTs). Today, APDs, as solid-state devices, are the preferred devices, offering quantum efficiencies of >90% in the visible. Every detector has a certain dark count rate, i.e. it "detects" photons when none are present. This problem can be reduced by cooling the detector. In the extreme case of a superconducting detector working at a temperature below 1 K, the dark count rate becomes negligible [434, 435]. These devices provide high efficiency, even in the infrared, but have to be operated at cryogenic

temperatures.

While the detection of individual optical photons is possible with good efficiency using commercial devices, lower frequencies present much bigger challenges. Not only are the photon energies lower (i.e. the energy to be detected is smaller), but in addition, thermally excited photons (in the quantum system as well as in the detector) must be avoided [436, 437].

13.1.3 Linear optics quantum computing

An important device for linear optics is the beam splitter (BS). It has 2 input ports and 2 output ports and can be represented as a logical gate operation: It transfers the two input states into two output states. For the symmetric beam splitter, the conversion matrix is

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

which corresponds to a Hadamard gate. If light is incident only on 1 port of the BS, equal intensities exit at the two output ports.

If identical photons are incident on both input ports, they interfere. Their combined state can be written, e.g., as

$$\Psi_{in} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rightarrow \Psi_{out} = U\Psi_{in} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The light therefore exits only on 1 port.

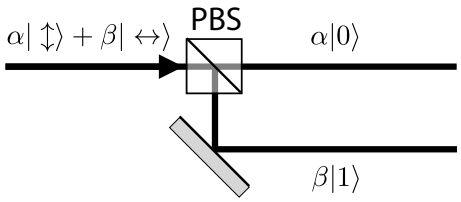

Figure 13.3: Conversion between polarization-mode and spatially encoded mode.

Fig. 13.3 shows how a polarization-encoded photonic qubit can be converted to a spatially encoded qubit of the same photon: The computational basis states of the incident photon are the vertical and horizontal polarization states. The photon is in the state

$$\Psi_{in} = \alpha|\updownarrow\rangle + \beta|\leftrightarrow\rangle.$$

The photon is now incident on an optical device called "polarizing beam splitter" (PBS). This device transmits vertically polarized light but reflects horizontally polarized light. The incident photon is therefore split into two wavepackets with orthogonal polarization and spatially separated paths. If required, we can rotate the polarization of one of the two partial beams, and we can now use the spatially separated paths as computational basis states $|0\rangle$ and $|1\rangle$. The state of the photon behind the beam splitter is then

$$\Psi = \alpha|0\rangle + \beta|1\rangle.$$

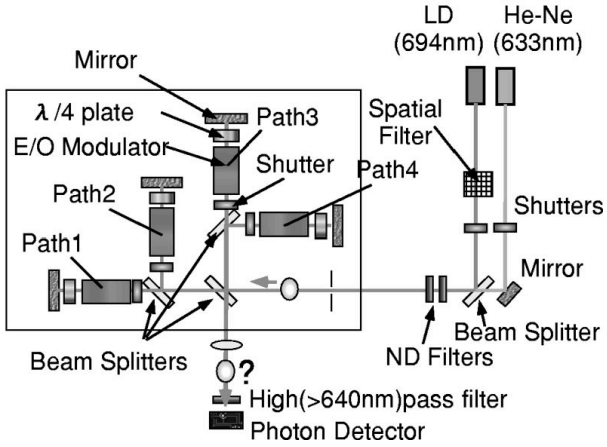

Figure 13.4: Optical implementation of the DJ algorithm [172].

Fig. 13.4 shows, as an example, the optical setup used for implementing the DJ-algorithm [172] for two input qubits and one qubit for storing the function result. It uses only a single photon; the input qubits are implemented as four different optical paths, and the function result is encoded in the polarization state of the photon. The beam splitters split the laser light into four paths with equal amplitude. Each optical path contains an electro-optical modulator. Applying a voltage to this modulator generates a transformation of the photon travelling through it. At

the end of each path, a mirror reflects the photon through the same path until they again hit the beam splitters. At the beam splitters, the partial amplitudes from the different paths interfere and depending on the relative phase, they are reflected or transmitted. The photodetector thus measures different count rates, depending on the relative phases.

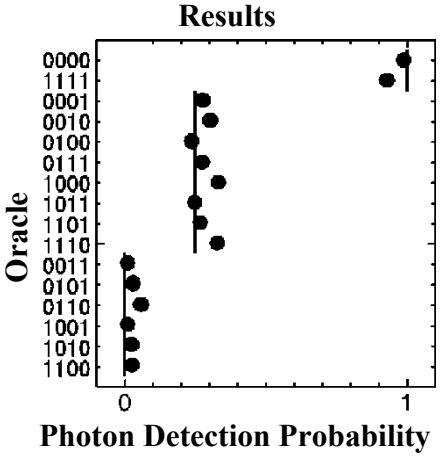

Figure 13.5: Result of the DJ algorithm [172].

Fig. 13.5 shows the observed count rates for the different functions. For the two constant functions, the photon detection probability is maximized, for the balanced values, it vanishes within the experimental uncertainties, and for the other functions, it takes an intermediate value.

Like in classical computing, integration by lithographic processes is rapidly increasing the complexity and capabilities of these devices [438].

13.1.4 Linear optics and measurements

Unfortunately, setups with linear optics cannot be readily extended to larger number of qubits: as the number of qubits increases, one needs either a coupling between different qubits or the number of optical components required increases exponentially with the number of qubits. In vacuum, interactions between photons are extremely weak. In many materials, effective interactions between photons can be mediated by the medium; this is, e.g., the mechanism that allows frequency doubling, parametric amplification, or the Kerr effect. However, while these interactions can be readily observed in high-intensity laser beams, they are usually very weak at the single photon level. Building a quantum computer on this basis appears therefore extremely difficult.

A possible way out was suggested by Knill, Laflamme, and Milburn: they realized that measurements of individual photons represent a nonlinear process that works well enough with single photons and can be used for quantum computing [175, 439]. This linear optics scheme encodes qubits in the mode occupied by the single photon, i.e., two modes are required to encode a logical qubit: $|0_L\rangle = |01\rangle$, $|1_L\rangle = |10\rangle$.

Their scheme differs from the usual network model in that they use measurements, which are clearly nonunitary operations, to process the data. The results of these measurements are fed back into the state of the quantum register by controlled phase shifts. Several steps have been taken towards realizing this scheme, including the construction of a two-qubit gate that is closely related to the CNOT operation [440].

Among the biggest difficulties of this architecture is the necessity for storing qubits. Even if the auxiliary photons used for the measurements can be produced *on demand*, which remains a challenging problem [441, 424, 442], the measurements are inherently probabilistic and have to be repeated several times to ensure success. Until success is assured, the photons have to be kept in a waiting state. While some schemes have been tested to store the quantum state of photons [443, 444, 445], the efficiency of such conversions is still much too low for useful implementations. While these difficulties make it unlikely that such a scheme will be implemented directly, similar proposals have been put forward that may be easier to implement.

13.1.5 Implementation

Experimental work towards this goal is under way. Single-qubit gates are straightforward to be implemented by retardation plates or modulators. Two-qubit gates are significantly more demanding but have been realized as probabilistic gates [446] by interference on a beamsplitter [447]. A probabilistic nondestructive CNOT gate was implemented for two independent photons using only linear optics [448].

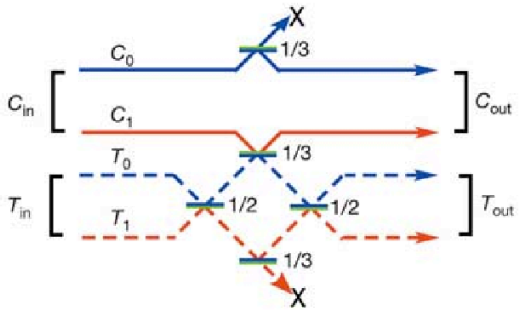

Figure 13.6: Optical CNOT gate [440].

Fig. 13.6 shows a similar scheme, which was also tested experimentally [440]. Each qubit is represented by a single, spatially encoded photon. The control and target qubits must be entered simultaneously. The target qubit goes through an interferometer, which is balanced, so it exits unchanged if the control qubit is not in mode C_1 - except that the two mirrors labeled "1/3" are only partially reflective (1/3). This means that the gate is probabilistic, not all photons reach the output. However, when a single photon is detected in each output (a coincidence count), which occurs with probability $P = 1/9$, the gate has worked correctly.

If the control photon is in mode C_0, it does not interact with the target photon and the target photon is not changed. However, if the control photon is in state C_1, the interaction between control and target photon results in a phase shift of the target photon, which is translated into a spin flip by the second "1/2" beam splitter.

The principle of this approach can also be used in different contexts. Apart from the network model, it can also be applied to the measurement-based approach to quantum-information processing [449] and it can be used for quantum non-demolition detection of photons [450].

13.1.6 Integrated optics

While the basic principles of photonic quantum computing can be demonstrated with discrete devices like mirrors and beam splitters, scalable devices have to rely on miniaturization on the basis of lithographic processing.

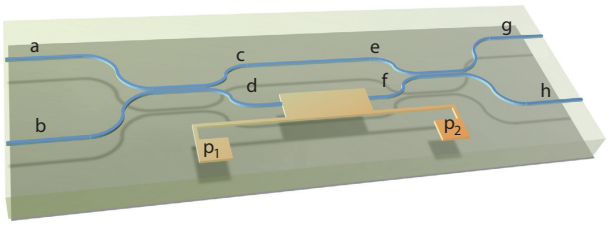

Figure 13.7: Single-qubit gate based on integrated optical waveguides.[451]

Figure 13.7 shows how single-qubit gates can be integrated on photonic chips.

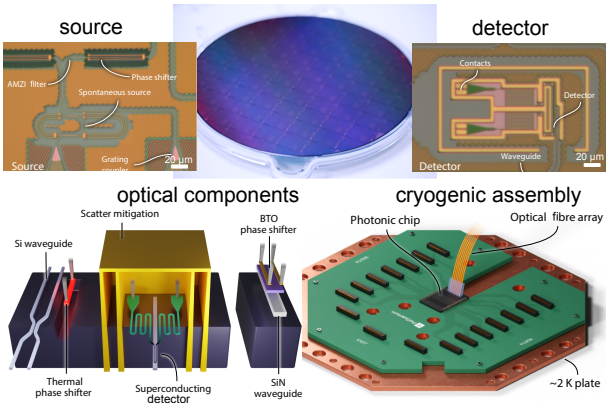

Figure 13.8: Platform for photonic quantum computing.[452]

Figure 13.8 shows some components of a platform for optical quantum processors that was proposed developed by the PsiQuantum[2] team.

[2] https://www.psiquantum.com

[452]. They are developing a general-purpose silicon photonic quantum computer and benchmarked a set of monolithically integrated silicon-photonics-based modules to generate, manipulate, network and detect heralded photonic qubits, demonstrating dual-rail photonic qubits with 99.98% 0.01% state preparation and measurement fidelity, HongOu Mandel (HOM) quantum interference between independent photon sources with 99.50% 0.25% visibility, two-qubit fusion with 99.22% 0.12% fidelity and a chip- to-chip qubit interconnect with 99.72% 0.04% fidelity, conditional on photon detection and not accounting for loss. Their technological basis is low-loss silicon nitride (SiN) waveguides and components to address loss, as well as fabrication-tolerant photon sources, high-efficiency photon-number- resolving detectors (PNRDs), low-loss chip-to-fibre coupling and barium titanate (BTO) electro-optic phase shifters for high-performance fast switching.

13.2 "Quantum only" tasks

Before we start discussing tasks which can only be performed quantum mechanically but not classically, we recall the no-cloning theorem (Section 4.4.10) where it is just the other way round. Any piece of classical information can be copied arbitrarily often and with arbitrary precision, but there is no way to copy an arbitrary quantum state. This inability to copy quantum information is the basis for secure communication by means of quantum key distribution, which we will discuss in Section 13.2.3. Before, in Section 13.2.2, we will show how one qubit may be used to transfer the information of two classical bits by a scheme known as (super-) dense coding. We will start, however, with quantum teleportation.

13.2.1 Quantum teleportation

We may be unable to give a *copy* of a quantum state to a friend, but under certain circumstances we are able to transmit some classical information which allows him or her to prepare precisely the state that we originally had. Our state will then be destroyed, of course, because otherwise we would have been able to violate the no-cloning theorem. The process can then be summarized by the formula

$$|\Psi\rangle_A \otimes |\text{something}\rangle_B \rightarrow |\text{something else}\rangle_A \otimes |\Psi\rangle_B.$$

A necessary resource for this teleportation of an unknown state is entanglement, that is, both partners must share among them two qubits (in the simplest case) in an entangled state. Quantum teleportation was proposed in 1993 by Bennett *et al.* [453] and is surprisingly simple.

We consider the usual characters, Alice and Bob. Let Alice be in possession of a qubit in the state

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle.$$

(Of course she does not *know* α and β, otherwise the problem would be trivial.) Furthermore Alice and Bob share a pair of qubits prepared in one of the Bell states (4.32), often also called an EPR pair,

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

where Alice can manipulate only the first qubit and Bob only the second one. The initial state of the combined three-qubit system is thus

$$\begin{aligned} |\chi\rangle &:= |\psi\rangle|\phi\rangle \\ &= \frac{1}{\sqrt{2}}[\alpha|0\rangle(|00\rangle + |11\rangle) \\ &\quad +\beta|1\rangle(|00\rangle + |11\rangle)]. \end{aligned}$$

Figure 13.9 shows the basic experiment: Alice applies a CNOT(1,2) gate to the two qubits in her possession, followed by a Hadamard gate $\mathbf{H}_1$ acting on the first qubit (the one initially containing $|\psi\rangle$). This entangles the two qubits ini-

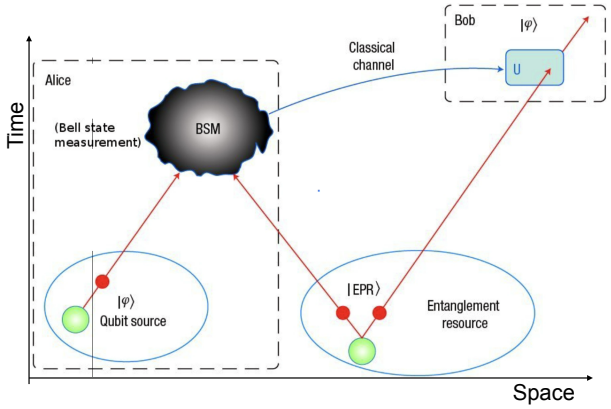

Figure 13.9: Simple quantum teleportation experiment.

tially in the states $|\psi\rangle$ and $|\phi\rangle$ with each other.

$$\begin{aligned}
|\tilde{\chi}\rangle &= \mathbf{H}_1 \text{ CNOT } (1,2)|\chi\rangle \\
&= \mathbf{H}_1 \frac{1}{\sqrt{2}} \left[\alpha|0\rangle(|00\rangle + |11\rangle) + \beta|1\rangle(|10\rangle + |01\rangle)\right] \\
&= \frac{1}{2}\left[\alpha(|0\rangle + |1\rangle)(|00\rangle + |11\rangle)\right. \\
&\quad \left.+\beta(|0\rangle - |1\rangle)(|10\rangle + |01\rangle)\right].
\end{aligned}$$

We rewrite this state in order to bring out clearly what has happened on Bob's end

$$\begin{aligned}
|\tilde{\chi}\rangle &= \frac{1}{2}\left[|00\rangle(\alpha|0\rangle + \beta|1\rangle) + |01\rangle(\alpha|1\rangle + \beta|0\rangle)\right. \\
&\quad \left.+|10\rangle(\alpha|0\rangle - \beta|1\rangle) + |11\rangle(\alpha|1\rangle - \beta|0\rangle)\right] \\
&= \frac{1}{2}\left[|00\rangle|\psi\rangle + |01\rangle\mathbf{X}_3|\psi\rangle + |10\rangle\mathbf{Z}_3|\psi\rangle\right. \\
&\quad \left.+|11\rangle \underbrace{(-i\mathbf{Y}_3)}_{\mathbf{X}_3\mathbf{Z}_3}|\psi\rangle\right] \\
&= \frac{1}{2}\sum_{M_1=0}^{1}\sum_{M_2=0}^{1} |M_1M_2\rangle\mathbf{X}_3^{M_2}\mathbf{Z}_3^{M_1}|\psi\rangle,
\end{aligned}$$

where $\mathbf{X}_3$, $\mathbf{Y}_3$, and $\mathbf{Z}_3$ are the Pauli matrices (4.12–4.14) applied to the qubit 3, that is, Bob's qubit.

Bob now possesses a superposition of four distorted variants of Alice's original state. Alice performs a measurement (in the computational basis) on the two qubits 1,2 to which she has access. She obtains one of the four combinations $|M_1M_2\rangle(M_1, M_2 = 0, 1)$ with equal probabilities. After the measurement the state of the complete system has been projected to

$$|M_1M_2\rangle\mathbf{X}_3^{M_2}\mathbf{Z}_3^{M_1}|\psi\rangle$$

so that Bob *possesses a definite modification of the desired state* $|\psi\rangle$, *but he does not yet know which one*! To let him know, Alice transmits the two measured classical bits (M_1, M_2) through a classical channel. The transmission through the classical channel is limited by the special theory of relativity and prevents superluminal communication, or, as Einstein put it, "spukhafte Fernwirkungen" (spooky actions at a distance). Bob then applies to his qubit the operator

$$\mathbf{Z}_3^{M_1}\mathbf{X}_3^{M_2} = (\mathbf{X}_3^{M_2}\mathbf{Z}_3^{M_1})^{-1}$$

and can enjoy the state $|\psi\rangle$ which is now in his possession, while Alice's original qubit is in the state $|M_1\rangle$.

An important aspect is that in this process neither matter nor energy were transported "explicitly", only two classical bits. Surprisingly enough these two classical bits were sufficient to reconstruct on Bob's side the state $|\psi\rangle$ which requires three real numbers for its complete specification (one amplitude, and two phases, assuming normalization). In a sense, these three real numbers contain infinitely more information than was transmitted; unfortunately (see the following subsection) this information cannot be retrieved completely. Nevertheless, the possibility of teleportation clearly shows how powerful a resource a shared EPR pair is. On the other hand, the necessity to have a shared EPR pair for every qubit (or electron, nucleon) whose state is to be teleported makes it very clear that we are still quite far away from any kind of "beam me up, Scotty" scenario. Nevertheless, single-qubit states have been successfully teleported in more than one laboratory, using optical and NMR techniques. References to those experiments (and to critical comments on them) can be found in [131] and in [40], p. 59.

13.2.2 (Super-) Dense coding

An arbitrary normalized pure single-qubit state is completely specified by three real numbers, for example, the two angles θ and ϕ in the Bloch sphere representation (4.21), plus an overall phase which is usually irrelevant. These real numbers contain much more information than the single binary digit contained in a classical bit, and it is interesting whether that much information can be reliably transmitted by a single qubit. The answer is no, unfortunately. However, transmission capacity can be doubled by using quantum instead of classical bits, as discovered by Bennett and Wiesner in 1992 [454], whose scheme has become known as superdense coding. In a sense, it is the inverse process of teleportation. Alice and Bob share an EPR pair and can transmit two classical bits by a single qubit. The method is difficult to implement and it is not important as a means of practical fast communication. However, it demonstrates one possibility of secure communication, as we shall see.

As in the preceding subsection, Alice and Bob are supposed to share the EPR pair state

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

(By the way, there is no need for any prior direct communication between Alice and Bob: they could have obtained their respective qubits from an "EPR pair distribution agency".) Now, if Alice wants to send the two classical bits (M_1, M_2) to Bob, she applies $\mathbf{X}_1^{M_1}\mathbf{Z}_1^{M_2}$ (to the only qubit accessible to her, that is, qubit 1). This yields one of the four states

$$\begin{aligned}
|\phi_{00}\rangle &:= \mathbf{X}_1^0\mathbf{Z}_1^0|\phi\rangle = |\phi\rangle \\
|\phi_{10}\rangle &:= \mathbf{X}_1^1\mathbf{Z}_1^0|\phi\rangle = \frac{1}{\sqrt{2}}(|10\rangle + |01\rangle) \\
|\phi_{01}\rangle &:= \mathbf{X}_1^0\mathbf{Z}_1^1|\phi\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \\
|\phi_{11}\rangle &:= \mathbf{X}_1^1\mathbf{Z}_1^1|\phi\rangle = -i\mathbf{Y}_1|\phi\rangle = \frac{1}{\sqrt{2}}(|10\rangle - |01\rangle)
\end{aligned} \tag{13.1}$$

Then Alice transmits her qubit to Bob. Note that the four states on the right-hand side are an orthonormal set (the Bell basis which we already encountered in Chapter 4) and thus can be distinguished by an appropriate measurement. Bob might first apply CNOT(1,2) and then measure the target bit 2. This yields

$$\begin{aligned}
\text{CNOT } |\phi_{00}\rangle &\sim |00\rangle + |10\rangle \longrightarrow 0 \\
\text{CNOT } |\phi_{10}\rangle &\sim |11\rangle + |01\rangle \longrightarrow 1 \\
\text{CNOT } |\phi_{01}\rangle &\sim |00\rangle - |10\rangle \longrightarrow 0 \\
\text{CNOT } |\phi_{11}\rangle &\sim |11\rangle - |01\rangle \longrightarrow 1.
\end{aligned}$$

Obviously this yields the first classical bit M_1 transmitted by Alice. The second qubit now has been used up in the measurement. The remaining classical bit M_2 is encoded in the relative sign in the four superpositions (13.1) above. Bob can decode it by applying the Hadamard gate $\mathbf{H} = \frac{1}{\sqrt{2}}(\mathbf{X}+\mathbf{Z})$ to his remaining qubit and then measuring it :

$$\mathbf{H}(|1\rangle + |0\rangle) = \sqrt{2}|0\rangle$$

(for $|\phi_{00}\rangle$ and $|\phi_{10}\rangle$), and

$$\pm\mathbf{H}(|0\rangle - |1\rangle) = \pm\sqrt{2}|1\rangle$$

(for the other two states). Experimentally this has been implemented by both optical and NMR techniques, see [131] for the references.

What about the security of this procedure for information transmission between Alice and Bob? Can a malignant person (usually called Eve, the eavesdropper) intercept the qubit transmitted by Alice and decode the information? Of course she can intercept and measure the qubit, but, regardless of the two classical bits M_1 and M_2 encoded by Alice, Eve will measure $|0\rangle$ and $|1\rangle$ with equal probabilities, so that she obtains no information whatsoever.

→ **Problem** 13.1

The information is encoded in the way the two qubits are entangled, and it can only be decoded by using information on *both* qubits. In the following subsection we will see how entanglement can be employed to generate keys for data encryption.

13.2.3 Quantum key distribution

Secure communication is a field where quantum mechanics may contribute in several ways to create or destroy security. In Section 8.3 we saw how quantum mechanics may help to break classical codes by Shor's algorithm. Here we will discuss how quantum mechanics helps to make secure communication possible by quantum key distribution, one of the central ideas in the field of *quantum cryptography* [455]. Alice and Bob exchange qubits in order to generate a key which can be used later to encrypt a message transmitted by a classical (and public) channel. The encrypted message can only be decrypted by means of the key. Quantum mechanics can be used to make sure that only two persons are in possession of the key. This should be contrasted to old-fashioned techniques such as providing secret agents with "code books" which may get lost, stolen, or copied.

A *key* is a (random) sequence of (classical) bits $\{k_i\}(i = 1, \ldots, N)$ which Alice uses to encrypt the N-bit *message* $\{m_i\}$ and transform it to the *code* $\{c_i\}$ by bitwise addition modulo 2:

$$c_i = k_i \oplus m_i = k_i \text{ XOR } m_i = (k_i + m_i) \bmod 2.$$

Bob can decrypt the code if he possesses the key:

$$m_i = c_i \oplus k_i,$$

as can be easily verified for all four possible combinations (k_i, m_i).

This method of encryption is only safe if the key is used *only once*. If two messages m and m' are encoded with the same key and the codes c and c' are intercepted, the relation

$$c_i \oplus c'_i = m_i \oplus m'_i$$

can be used to eliminate the deliberate irregularities introduced by encoding. Subsequently standard correlation analyses (as available at any secret service) can be applied in an attempt to separate m from m'. Given this situation there is obviously a need to distribute fresh keys among Alice and Bob. Quantum key distribution serves that purpose. There exist several schemes or "protocols" to do this quantum mechanically, see [131]. Here we will discuss only two schemes which are closely related to each other.

13.2.4 BB84

First we discuss the four-state protocol known as BB84 [456]. This protocol uses four pairwise orthogonal states

$$|0\rangle, |1\rangle, |\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$$

(the eigenstates of the Pauli matrices $\mathbf{Z}$ and $\mathbf{X}$, respectively) which can be easily prepared as linearly polarized photons with electric field $\vec{E}$ along $\hat{x}$, $\hat{y}$, and $\hat{x} \pm \hat{y}$, where $\hat{x}$ and $\hat{y}$ are unit vectors along orthogonal coordinate axes.

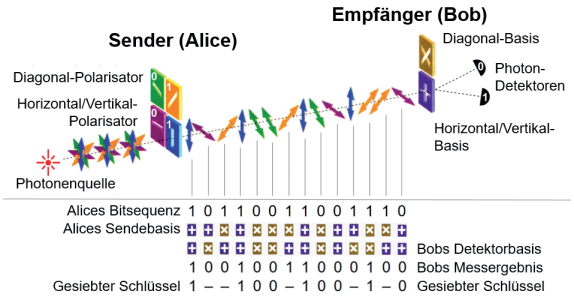

Figure 13.10: Experimental protocol for QKD.

Measurements are performed with linear polarizers along these directions, and detectors. A photon polarized along $\hat{x}$ passes through a polarizer along $\hat{x}$ and is detected, one along $\hat{y}$ is not. To get an unambiguous result the observer must *know that* a photon should be coming along his way and that it is polarized *either* along $\hat{x}$ *or* along $\hat{y}$. A photon polarized along one of the diagonal directions $\hat{x} \pm \hat{y}$ will not yield any information when analyzed with a polarizer along $\hat{x}$, because both possibilities will give a signal in half of all cases.

Alice prepares $2n$ qubits randomly in one of the four states. Each qubit i contains two classical bits, namely:

- $b_{p,i}$ telling which basis, $\{|0\rangle, |1\rangle\}$ or $\{|+\rangle, |-\rangle\}$, was used to prepare the state, and
- $s_{p,i}$ telling which state (1st or 2nd) of the given basis was prepared.

Bob (ideally) receives all these qubits and measures them, randomly switching the basis used for measuring. He also stores two bits for each qubit, namely

- $b_{m,i}$, telling which basis was used to measure the qubit, and
- $s_{m,i}$, telling which state of the given basis was measured.

Alice now (*after* the transmission) tells Bob (over a public channel) the sequence $\{b_{p,i}\}$ which Bob compares to his sequence $\{b_{m,i}\}$. Both parties keep only qubits with $b_{p,i} = b_{m,i}$ and throw away all the others (roughly n), because they do not contain useful information, as discussed above. For the remaining qubits the classical bits $s_{p,i} = s_{m,i}$ are known to both Alice and Bob. They constitute the key.

The security aspects of this procedure become visible if Eve intercepts and measures the qubits. During transmission Eve neither knows which basis Alice uses for preparing the qubits, nor which basis Bob uses for measuring them. Nevertheless she has to supply Bob with qubits resembling those transmitted by Alice, in order not to be discovered immediately. Eve's only possible strategy is to use one of the two measurement bases randomly for each qubit. After each measurement she prepares a fresh qubit in the basis state just measured and passes it on to Bob. After the transmission is complete, Alice and Bob discuss their bases and agree to discard about half of their measurements as useless. (Eve of course listens to the conversation and discards the same measurements.) Let us discuss what effect Eve's attack has on the code, that is, on those qubits which have been measured by Bob in the same basis as used by Alice to prepare them. For about 50 % of these qubits Eve has performed her measurement in the right basis, causing no disturbance. The remaining 50 % of the qubits have been measured in the wrong basis by Eve and then passed on to Bob. The final measurement by Bob (in the right basis) has projected half of these qubits back into the state originally prepared by Alice, so the overall error rate caused by Eve is 25 %.

Alice and Bob can agree to publicly compare a certain share of the key (thereby sacrificing that share, of course), and if they detect no differences they can be pretty certain that no eavesdropping has occurred. (If m bits are compared the probability that they are all correct by chance in the presence of eavesdropping is $\left(\frac{3}{4}\right)^m = 3 \cdot 10^{-13}$ for $m = 100$.) Of course Eve might be clever enough not to intercept *every* qubit, and also there might be errors other than those caused by eavesdropping in a less than perfect transmission line. All these problems have been analyzed and may be overcome, see [41, 131].

13.2.5 Tests

An early demonstration of the BB84 scheme used 23 km of public telecom glass fiber beneath Lake Geneva [457]. In that experiment polarized light pulses with $\lesssim 0.1$ photons per pulse were used: there must be (practically) no pulses with two or more photons because an eavesdropper might intercept just one photon and go unnoticed. The bit error rate was $\sim 1\%$ and the data transfer rate was of the order of MHz instead of the usual (in non-secure communication) GHz. More recent work was reviewed in [455]. The distance over which quantum keys can be distributed was extended beyond 100 km via glass fibers [458] as well as in free space [459]. Significant progress in this field might be achieved by using actual single-photon sources instead of strongly attenuated pulses [460, 461].

Other protocols for secure communication involve entangled states, for example EPR pairs, and it was shown that the Bell inequalities (mentioned in Chapter 4) distinguishing genuine quantum correlations from classical ones can be

used to detect eavesdroppers. An extremely simple scheme involving EPR pairs but no Bell inequalities was suggested by Bennett, Brassard, and Mermin in 1992 [462]. This scheme is essentially equivalent to the BB84 protocol just discussed, as we will see. Alice and Bob share $2n$ EPR pairs

$$|\phi_i\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$$

in the usual way, that is, each qubit is accessible to one person only. Both measure the qubit accessible to them, and thus project it on one of the eigenstates of $\mathbf{X}$ or $\mathbf{Z}$ (at random). They inform each other publicly about the $(\mathbf{X}, \mathbf{Z})$ sequence used, but not about the results of the measurements. They discard all measurements where one has measured $\mathbf{X}$ and the other $\mathbf{Z}$. The remaining measurement results are perfectly anticorrelated and can be used to produce two equal bit strings of length $\sim n$. A part of the key may again be sacrificed to detect eavesdropping. The scheme has an additional advantage: the EPR pairs can be left untouched until just before the key is needed so that the time during which the key is kept in classical storage and can be copied by a thief is minimal. Of course this requires the ability to preserve EPR pairs over long times, but that is a different story.

For further information on quantum cryptography, interested readers are referred to the review [455]. This article treats a broad range of topics, from theoretical foundations to detailed discussions of fiber optical transmission systems. A more recent review treating all aspects of quantum communication is [463].

13.3 A few bits of classical information theory

13.3.1 Measuring Information

Information theory has developed over the past six decades in parallel to computer science. Its roots are in communication theory, that is, in the theory of transmission of information by telephone or radio. Of course, all parts of this book deal with information theory in a wider sense, but as the subfields have developed, questions of computation and algorithm development have been separated from information theory in a narrower sense. In this section we will restrict ourselves to some problems dealing with the *transmission of information*.

The most fundamental questions of course are, what *is* information, or, more precisely, how can it be quantified? These questions were dealt with in the pioneering contributions of Claude Shannon [26] in the late 1940s. The historical (or socio-economic) context was the rapid growth of communication by telephone lines. Consequently the problem was formulated as the problem of effectively transmitting information through a given "channel". The channel, for example a telephone line, may connect two points in space, but it may also connect two points in time, in which case we are dealing with effective data *storage*. As every channel has physical limits, there is an obvious interest in precisely determining these limits and extending them if possible. To do that, a measure of the information content of a communication must be developed and related to the capacity of the channel. That is the content of *Shannon's noiseless channel coding theorem*. Of course channels are always noisy, and questions of error-correction immediately come to mind. Actually there is a large subfield of classical information theory dealing with the development of error-correcting codes. The fundamental limits are fixed by *Shannon's noisy channel coding theorem*.

In contrast to the theory of quantum (or classical) algorithms, here we are not dealing with a small number of (qu-)bits which must be processed, but with large quantities of transmitted data. From the point of view of the communications engineer these data form a random sequence of symbols about which only some statistical properties may be known. It turns out (not unexpectedly) that some key concepts from statistical mechanics, such as entropy, are use-

ful also in information theory, both classical and quantum.

After discussing some notions of classical information theory we will try to generalize the concepts to the quantum regime. Unfortunately it turns out that the use of qubits does not significantly speed up the transmission of classical information (such as this text) through a noiseless channel. Nevertheless it is interesting to study how the notion of classical information may be generalized to quantum information, how strongly quantum information may be compressed (looking for the quantum analogs of Shannon's theorems), and how quantum noise (i.e., *continuous* fluctuations in both amplitude and phase in contrast to mere bit flips) may affect the transmission.

13.3.2 Information content and entropy

The first question is, how to quantify information. Imagine you are told that

$$X = 2.$$

How much information do you gain? That depends on your previous knowledge: if you knew already that X was 2, you learn nothing. If you only knew that X was determined by throwing a die you gain information. The information content of X is a measure of your *ignorance*: how much information would you gain if you learned the value of X? That depends on the number of possible values x of the random variable X and their probabilities $p(x)$. The general formula for the information content of X introduced by Shannon is

$$S(X) \equiv S(\{p(x)\}) = -\sum_x p(x) \log_2 p(x).$$

Since $0 \le p(x) \le 1$, $S(X) \ge 0$. Let us look at more examples to see if this definition makes sense: $p(x) = \delta_{x,2}$ (for integer x) $\Rightarrow S(X) = 0$. (Nothing is learned if we know already that $X = 2$.)

- $p(x) = \frac{1}{N}$ for $x = 1, \ldots, N$ and zero otherwise $\Rightarrow S = \log_2 N$.
  $N = 6 \Rightarrow S = 2.58$ (the fair die)
  $N = 2^m \Rightarrow S = m$: m bits must be specified to convey the information
- $p(6) = \frac{1}{2}, p(1) = \ldots = p(5) = \frac{1}{10} \Rightarrow S = 2.16$ (a loaded die).

The comparison between the fair die and the loaded die shows that the potential information gain decreases if the information about the probability distribution increases. The uniform probability distribution is the one with "maximal ignorance". Obviously S is closely related to the entropy well-known from Statistical Mechanics, and it is indeed often called information entropy or *Shannon entropy*. A simple but important special case is a binary variable ($X = 0$ or 1, say), with $p(1) = p \Rightarrow p(0) = 1 - p$. $S(x)$ is then a function of p only:

$$S(X) = H(p) = -p \log_2 p - (1-p) \log_2 (1-p).$$

The binary entropy function $H(p)$ assumes its maximum value 1 at $p = \frac{1}{2}$; it is zero if X is known with certainty.

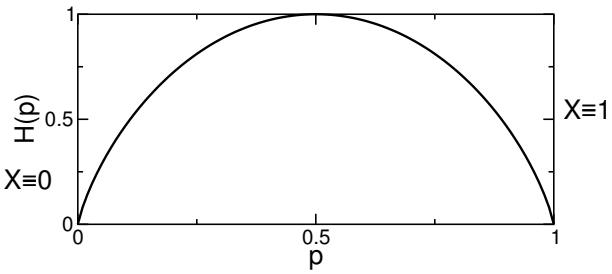

Figure 13.11: The binary entropy function $H(p)$.

13.3.3 Mutual information and the data processing inequality

For two random variables X and Y we can define the *conditional probability* $p(y|x)$ that the random variable Y assumes the value y under the condition that $X = x$, and the *conditional entropy*

$$S(Y|X) = -\sum_x p(x) \sum_y p(y|x) \log_2 p(y|x).$$

(13.2)

Since $-\sum_y p(y|x)\log_2 p(y|x)$ is the information content of Y for given value of X, the conditional entropy $S(Y|X)$ is the average information content remaining in Y if we were to learn the value of X. (Where the average is performed over the possible values of X.) Since the ("simultaneous") probability $p(x,y)$ that $X = x$ and $Y = y$ is given by

$$p(x,y) = p(x)p(y|x), \tag{13.3}$$

we can rewrite (13.2) as

$$S(Y|X) = -\sum_x \sum_y p(x,y)\log_2 p(y|x). \tag{13.4}$$

We now define the *mutual information* content of X and Y as

$$\begin{aligned} I(X:Y): &= \sum_x \sum_y p(x,y)\log_2 \frac{p(x,y)}{p(x)p(y)} \\ &= I(Y:X). \end{aligned} \tag{13.5}$$

If X and Y are independent random variables, that is, $p(x,y) = p(x)p(y)$, the mutual information $I(X:Y) = 0$ and this indicates that $I(X:Y)$ in fact measures how much X and Y "know about each other". We can relate the mutual information to the conditional entropy by noting that

$$\begin{aligned} I(X:Y) &= \sum_x \sum_y p(x,y)\log_2 \frac{p(x,y)}{p(x)} \\ &\quad -\sum_x \sum_y p(x,y)\log_2 p(y) \\ &= \sum_x \sum_y p(x,y)\log_2 p(y|x) \\ &\quad -\sum_y p(y)\log_2 p(y) \\ &= -S(Y|X) + S(Y), \end{aligned}$$

where we have used (13.3), (13.4), and $p(y) = \sum_x p(x,y)$. Due to the symmetry of $I(X:Y)$ we also have

$$I(X:Y) = I(Y:X) = -S(X|Y) + S(X).$$

Defining the information content of the "two-component" random variable (X,Y) by

$$S(X,Y) = -\sum_x \sum_y p(x,y)\log_2 p(x,y) \tag{13.6}$$

and using the normalization conditions $p(x) = \sum_y p(x,y)$ and $p(y) = \sum_x p(x,y)$, we see that

$$I(X:Y) = -S(X,Y) + S(X) + S(Y),$$

where $S(X,Y)$ is the information content of the "vector" random variable (X,Y).

During data processing, information can only decrease. To see this we reconsider the fundamental step (2.1) of data processing from a probabilistic point of view. The register is described by a random variable which is capable of a certain set of states (or values). The set of rules (the program) determining the transition from one state of the register to the next state is encoded in conditional probabilities. In this language, data processing is a stochastic process (a Markov chain). We consider two steps of data processing involving three random variables $X \to Y \to Z$ where successive variables are connected by conditional probabilities $p(y|x)$ and $p(z|y)$ and where the simultaneous probability $p(x,y,z) = p(x)p(y|x)p(z|y)$. Under these conditions the *data processing inequality* says

$$S(X) \geq I(X:Y) \geq I(X:Z),$$

that is, Z cannot know more about X than Y knew which is less than the information content of X. This highly plausible inequality (a corollary to which is the well-known rule "garbage in, garbage out") can be deduced from the properties of the various entropy functions discussed above. (Compare, for example, [40], Chap. 11).

13.3.4 Data compression and Shannon's noiseless channel coding theorem

The basic idea of data compression is very simple and also very old. Determine which sequences of

symbols or words occur most frequently and use abbreviations for them, that is, code these words in short strings of the symbols (bits, for example) used for data transmission. We illustrate this principle with a very simple example. Suppose we wish to transmit information from a source X with a four-letter alphabet with unequal probabilities. Four symbols can be distinguished by using two bits and there is a "natural" (or naïve) way to do this. In the table we show both the naïve code and a "clever" code which we analyze below.

symbol	probability	naïve code	clever code
1	$\frac{1}{2}$	00	0
2	$\frac{1}{4}$	01	10
3	$\frac{1}{8}$	10	110
4	$\frac{1}{8}$	11	111

Note that in the naïve code all symbols are stored in two bits each. The clever code uses bit strings of variable length, but nevertheless the boundaries of the symbols are always well defined: after a "0" or after at most three bits. The average length of the cleverly coded string in bits per symbol then is

$$\frac{1}{2}\cdot 1+\frac{1}{4}\cdot 2+\frac{2}{8}\cdot 3=\frac{7}{4}<2.$$

Let us compare this to the entropy of the source:

$$\begin{aligned} S(X) &= -\frac{1}{2}\log_2\frac{1}{2}-\frac{1}{4}\log_2\frac{1}{4}-\frac{2}{8}\log_2\frac{1}{8} \\ &= \frac{1}{2}\cdot 1+\frac{1}{4}\cdot 2+\frac{2}{8}\cdot 3=\frac{7}{4}. \end{aligned}$$

The fact that the two numbers are equal is no coincidence. Also, no compression scheme can be constructed which works with a smaller number of bits per symbol on average. This is the contents of Shannon's noiseless channel coding theorem.

To illustrate the idea a little more generally (but without going into full generality) we consider a source sending a stream of binary symbols: $X=0,1\ ;\ p(1)=p, p(0)=1-p$ with $p\neq\frac{1}{2}$. (Remember: the central elements of data compression were the fact that not all strings are equally probable, and the use of short codes for frequent symbols.) We will not encode individual symbols but blocks of n symbols with n large. In the typical case such a block will contain np ones and $n(1-p)$ zeros. (Let us postpone for a moment the discussion of what "typical" really means.) There are many blocks of n symbols np of which are ones. The probability of any such sequence of zeros and ones is

$$p_{\text{typ}}=p^{np}(1-p)^{n(1-p)}.$$

Now note that

$$\begin{aligned} \log_2 p_{\text{typ}} &= np\log_2 p+n(1-p)\log_2(1-p) \\ &= -nH(p) \end{aligned}$$

where $H(p)$ is the binary entropy function defined earlier. Thus

$$p_{\text{typ}}=2^{-nH(p)}.$$

As these typical sequences all have equal probability $2^{-nH(p)}$, their total number is $2^{nH(p)}$, and they can be numbered, from 1 to $2^{nH(p)}$. To communicate which one of the $2^{nH(p)}$ possible typical sequences are transmitted, only $nH(p)$ bits are needed, not n bits as in the case where bits are transmitted one by one. It is not possible to distinguish the typical sequences by sending fewer than $nH(p)$ bits, since they are all equally probable, so the compression from n to $nH(p)$ is optimal.

13.3.5 Data compression

So, how typical is typical, and why is the above argument relevant? Why do we really encounter (almost) only typical sequences? It turns out that the answer to these questions is provided by one of the "laws of large numbers" arguments which are possibly familiar from elementary statistical mechanics. There it is shown, for example, that the energy per particle may be allowed

to fluctuate arbitrarily, but nevertheless the total energy of a *large* number of particles practically does not deviate from its mean value. By a similar argument we will now show that, although the individual symbols of a sequence may fluctuate between 0 and 1, a long sequence will never deviate much from the typical number of zeros and ones, that is, np ones and $n(1-p)$ zeros. The probability of finding m ones in a sequence of n symbols is

$$p(m) = \left(\begin{array}{c} n \\ m \end{array} \right) p^m (1-p)^{n-m},$$

the binomial distribution. For fixed p and large n, the binomial distribution is excellently approximated by a Gaussian distribution. To see this we write down $\ln p(m)$, approximating the logarithm of the binomial coefficient with the help of Stirling's formula

$$\ln n! = n \ln n - n + \mathcal{O}(\ln n), \tag{13.7}$$

valid for large n. (We assume that n, m, and $(n-m)$ are all sufficiently large.) We then calculate the first and second derivatives of $\ln p(m)$ which we need for a Taylor expansion. The results are

$$\frac{d}{dm} \ln p(m) = \ln p - \ln(1-p) - \ln m + \ln(n-m) \tag{13.8}$$

and

$$\frac{d^2}{dm^2} \ln p(m) = -\frac{1}{m} - \frac{1}{n-m} = -\frac{n}{m(n-m)}. \tag{13.9}$$

The first derivative of $\ln p(m)$ vanishes if $\frac{p}{1-p} = \frac{m}{n-m}$, or $m = np$, and we see that np is indeed the most probable number of ones in a sequence of length n. A Taylor expansion of $\ln p(m)$ about the value $m = np$ then reads

$$\ln p(m) \approx \ln p(np) - \frac{(m-np)^2}{2} \frac{1}{np(1-p)}. \tag{13.10}$$

This shows clearly that $p(m)$ is a Gaussian

$$p(m) \approx \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(m-np)^2}{2\sigma^2}} \tag{13.11}$$

with standard deviation $\sigma = \sqrt{np(1-p)}$. (We have adjusted the normalization of the Gaussian (13.11) by hand, because we used the crude form (13.7) of Stirling's asymptotic expansion. Taking into account a few more terms in this expansion leads to the correct normalization automatically, but makes the calculation somewhat less transparent.) Note that, while the mean value np grows linearly with the sequence length n, the standard deviation only grows as $\sqrt{n}$. That is, the *relative* fluctuations of the number of ones in a sequence becomes smaller as the sequence grows longer and for long enough sequences we can be pretty sure that almost all sequences are typical.

Thus we only have to transmit $H(p) < 1$ bits per symbol for our binary source. More generally, for a source producing random variables X (capable of d values so that coding the symbols one by one would require $\log_2 d$ bits per symbol) with an information content $S(X)$ we need $nS(X) < n \log_2 d$ bits to communicate n values of X. This fact about the compressibility of data is known as Shannon's noiseless channel coding theorem.

For practical purposes it is of course not always possible to wait until a large number n of symbols have accumulated before starting the transmission. However, there are near-optimal coding schemes for blocks of a few (say, four) symbols only. They are based on the same idea as the example we started with: use shorter transmission codes for the most frequently occurring blocks of symbols. An example for such a scheme is the *Huffman code* (compare [41]).

13.3.6 The binary symmetric channel and Shannon's noisy channel coding theorem

We have to think about signal transmission in the presence of noise, because noise is unavoidable in real-world systems. Depending on the

physical nature of the signal and the transmission channel, different types of noise are possible. We will concentrate on the important and simple case of binary digital transmission (of zeros and ones, that is) and symmetric bit-flip noise. That means that every single bit is flipped with a certain probability p on its way down the channel, regardless of its value (0 or 1) and regardless of the fate of all other bits. Such a channel is called a binary symmetric channel, and we want to know its capacity, measured in (useful) bits out per bit in. It turns out (see [41] for details) that for the maximum information content of the source, $S(X) = 1$ (that is, 0 and 1 are equally probable in the input bit stream) the channel capacity is

$$C(p) = 1 - H(p),$$

where $H(p)$ is again the binary entropy function defined earlier and p is the bit flip probability. For a noisy channel one must use some redundancy, that is, one must employ *error-correcting codes*. Shannon's noisy channel coding theorem tells us that, for any given channel capacity $C(p)$, there exist error-correcting codes which allow for transmission with an arbitrarily small error probability.

Unfortunately the theorem is an existence theorem and does not tell us immediately how such a code may be constructed, but fortunately, a variety of clever error-correcting codes have been constructed (see [131] for some examples), for example, for the transmission of image data from satellites traveling the solar system to Jupiter and beyond.

13.4 A few bits of quantum information theory

13.4.1 The von Neumann entropy

It turns out that a useful quantum analog to Shannon's entropy (information content) for a classical set of probabilities p_i (which characterize the possible values x_i of a random variable X)

$$S(\{p_i\}) = -\sum_i p_i \log_2 p_i$$

is the *von Neumann entropy*

$$S(\boldsymbol{\rho}) = -\mathrm{Tr}\boldsymbol{\rho} \log_2 \boldsymbol{\rho}$$

which is defined for any density operator, that is, any operator $\boldsymbol{\rho}$ with $\boldsymbol{\rho} = \boldsymbol{\rho}^\dagger \geq 0$, $\mathrm{Tr}\boldsymbol{\rho} = 1$. Any such $\boldsymbol{\rho}$ can be decomposed in projectors onto normalized but not necessarily orthogonal pure states,

$$\boldsymbol{\rho} = \sum_i p_i |\phi_i\rangle\langle\phi_i| \quad (p_i \geq 0;\ \sum_i p_i = 1).$$

This is possible in many ways for any given $\boldsymbol{\rho}$, and to any of these possibilities we can assign a (classical) Shannon entropy $S(\{p_i\})$; it can be shown that

$$S(\{p_i\}) \geq S(\boldsymbol{\rho}),$$

with equality if and only if the vectors $|\phi_i\rangle$ are pairwise orthogonal. (Take, for example, the eigenstates of $\boldsymbol{\rho}$.) This inequality has a fairly obvious interpretation in terms of the distinguishability of two quantum states. Imagine a person (Alice) sending a string of classical symbols x_i down a line to another person (Bob), with probabilities p_i. We have learned that the information content of this transmission is $S(\{p_i\})$.

Now let us assume that Alice is a dedicated follower of fashion and goes into the quantum communication business. Instead of sending classical symbols x_i she sends quantum states $|\phi_i\rangle$. While Bob can easily distinguish all possible x_i, he can only distinguish two states with certainty if they are orthogonal to each other. This is also related to the no-cloning theorem: imagine Bob *could* clone arbitrary unknown quantum states. He then could make many copies of the incoming state and perform *many* measurements comparing clones of Alice's state to clones of all possible states and determine Alice's state with high probability.

It is instructive to consider a simple example involving a two-dimensional Hilbert space spanned by the vectors $|\alpha\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $|\gamma\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Let us define a third vector

$$|\beta\rangle := \cos\phi|\gamma\rangle + \sin\phi|\alpha\rangle$$

and the density matrix

$$\begin{aligned} \boldsymbol{\rho} &:= p|\alpha\rangle\langle\alpha| + (1-p)|\beta\rangle\langle\beta| \\ &= \begin{pmatrix} p + (1-p)\sin^2\phi & (1-p)\cos\phi\sin\phi \\ (1-p)\cos\phi\sin\phi & (1-p)\cos^2\phi \end{pmatrix}. \end{aligned}$$

The easiest way to calculate the von Neumann entropy $S(\boldsymbol{\rho})$ is via the eigenvalues λ_i of $\boldsymbol{\rho}$:

$$S(\boldsymbol{\rho}) = -\sum_i \lambda_i \log_2 \lambda_i.$$

The eigenvalues of the above density matrix are

$$\lambda = \frac{1}{2} \pm \sqrt{\frac{1}{4} - p(1-p)\cos^2\phi}.$$

For $\phi = 0$ the states $|\alpha\rangle$ and $|\beta\rangle$ are orthogonal, the eigenvalues of $\boldsymbol{\rho}$ are $\lambda = p$ and $\lambda = 1-p$ and thus $S(\boldsymbol{\rho}) = H(p)$ (the binary entropy function), whereas for $\phi \neq 0$ $|\alpha\rangle$ and $|\beta\rangle$ cannot be distinguished with certainty, and $S(\boldsymbol{\rho})$ is strictly smaller than $H(p)$, as seen in the figure.

The quantum entropy has some non-classical properties. Whereas classical random variables X, Y always fulfill

$$S(X) \leq S(X,Y),$$

that is, the entropy of a subsystem is never greater than that of the total system, this *is* possible for a quantum system. Consider two qubits A, B in the (pure!) state

$$\begin{aligned} |\phi\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \\ \boldsymbol{\rho}_{AB} &= |\phi\rangle\langle\phi| \Rightarrow S(\boldsymbol{\rho}_{AB}) = 0. \end{aligned}$$

However, the reduced density matrix of subsystem A (obtained from $\boldsymbol{\rho}_{AB}$ by performing the trace over the Hilbert space of B) is $\boldsymbol{\rho}_A = \frac{1}{2}\mathbf{1} \Rightarrow S(\boldsymbol{\rho}_A) = 1$.

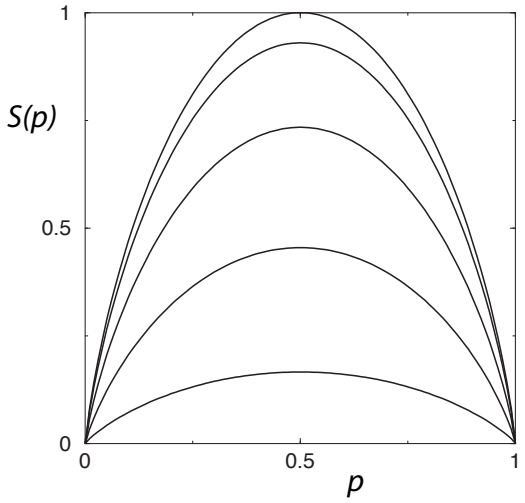

Figure 13.12: The von Neumann entropy for a simple two-dimensional density matrix. Curves are for $\phi = 0, 0.1\pi, 0.2\pi, 0.3\pi$, and 0.4π, respectively (top to bottom). See text for details.

Evidently this is related to the entanglement between A and B. In any pure state $\boldsymbol{\rho}_{AB}$, the subsystems A and B are entangled if and only if

$$S(\boldsymbol{\rho}_{AB}) < S(\boldsymbol{\rho}_A) \text{ (or } S(\boldsymbol{\rho}_B)),$$

where, of course, $\boldsymbol{\rho}_A$ is again the reduced density matrix.

Most theorems concerning entropy, which are relevant to quantum information theory, can be derived from a few fundamental properties, which are discussed, proved and applied in [40] and which we just quote here for the sake of completeness:

i) *Concavity*

$$\lambda_1 S(\boldsymbol{\rho}_1) + \lambda_2 S(\boldsymbol{\rho}_2) \leq S(\lambda_1\boldsymbol{\rho}_1 + \lambda_2\boldsymbol{\rho}_2)$$

$(\lambda_{1,2} \geq 0, \lambda_1 + \lambda_2 = 1)$. In statistical mechanics, the concavity of the entropy is related to thermodynamical stability.

ii) *Strong subadditivity*

$$S(\boldsymbol{\rho}_{ABC}) + S(\boldsymbol{\rho}_B) \leq S(\boldsymbol{\rho}_{AB}) + S(\boldsymbol{\rho}_{BC}).$$

iii) *Triangularity*

$$|S(\boldsymbol{\rho}_A) - S(\boldsymbol{\rho}_B)| \leq S(\boldsymbol{\rho}_{AB} \leq S(\boldsymbol{\rho}_A) + S(\boldsymbol{\rho}_B)$$

All of these inequalities also hold (in appropriately modified form) for the (classical) Shannon entropy, except the first one in iii).

13.4.2 The accessible information and Holevo's bound

We are still dealing with the transmission of classical data through a quantum channel. Let Alice have a classical information source X, that is, a random variable with values x_i and probabilities $p_i (i = 0, \ldots, n)$. According to the value x_i to be transmitted, Alice prepares one quantum state $\boldsymbol{\rho}_i$ from a fixed set of (mixed, in general) states $\boldsymbol{\rho}_0, \ldots, \boldsymbol{\rho}_n$ and gives it to Bob who measures the state and gets a result that can be treated as a classical random variable Y capable of values $y_0, \ldots, y_m$. Let us discuss Bob's measurement a little more precisely. Bob has a set of *measurement operators* $\mathbf{M}_i (i = 0, \ldots, m)$ which he can apply to any incoming state vector $|\psi\rangle$ (and also, with appropriate changes in notation, to mixed states). The probability of finding the result i is

$$p_i = \langle\psi|\mathbf{M}_i^\dagger \mathbf{M}_i|\psi\rangle$$

and the state immediately after the measurement is

$$\frac{\mathbf{M}_i|\psi\rangle}{\sqrt{\langle\psi|\mathbf{M}_i^\dagger \mathbf{M}_i|\psi\rangle}}.$$

The operators $\mathbf{E}_i := \mathbf{M}_i^\dagger \mathbf{M}_i$ are positive, and if $\sum_{i=0}^m \mathbf{E}_i = \mathbf{1}$ they are called POVM elements (positive operator valued measure elements). (If the sum is smaller than one, Bob's measurement misses some possibilities of the incoming $|\psi\rangle$.) An extremely simple example for a set of POVM elements are the projectors $\mathbf{P}_i$ on the states of a basis.

Turning back to the result Y of Bob's measurement (described by the POVM elements $\mathbf{E}_0, \ldots, \mathbf{E}_m$), it is clear that what Bob can learn about Alice's message is $I(X : Y)$, the mutual information, which depends on the cleverness of his measurement strategy. The *accessible information* is the maximum of $I(X : Y)$ over all measurement strategies. There is no prescription to calculate the accessible information, but there is a *bound* by Holevo (also often spelled Kholevo). Under the conditions described above, and with $\boldsymbol{\rho} := \sum_i p_i \boldsymbol{\rho}_i$, we have

$$I(X : Y) \leq S(\boldsymbol{\rho}) - \sum_i p_i S(\boldsymbol{\rho}_i) =: \chi,$$

where χ is sometimes called the Holevo information. (For the simplest possible example compare Section 12.1.2 of [40].)

13.4.3 Schumacher's noiseless channel coding theorem

Consider a "quantum alphabet" of states $|\phi_i\rangle$ (not necessarily orthogonal to each other) with probabilities p_i. Such an alphabet can be described by a density operator

$$\boldsymbol{\rho} = \sum_{i=1}^{|A|} p_i |\phi_i\rangle\langle\phi_i|.$$

A message is a sequence of n "quantum characters": $|\phi_{i_1}\rangle|\phi_{i_2}\rangle \ldots |\phi_{i_n}\rangle$. The ensemble of n-symbol messages is described by the density operator $\boldsymbol{\rho}^{\otimes n}$ which lives in a Hilbert space $\mathfrak{H}^{\otimes n}$ of dimension

$$|A|^n = 2^{n \log_2 |A|}$$

(or smaller, if the alphabet states are not linearly independent).

Is it possible to compress the information contained in $\boldsymbol{\rho}^{\otimes n}$? Schumacher's 1995 theorem provides an affirmative answer. For sufficiently large

n, $\boldsymbol{\rho}^{\otimes n}$ is compressible to a state in a Hilbert space of dimension $2^{nS(\boldsymbol{\rho})}$ (that is, in $nS(\boldsymbol{\rho})$ qubits) with a fidelity (probability that after decompression the original state is recovered) approaching 1. This means that $S(\boldsymbol{\rho})$ is the number of qubits of essential quantum information, per character of the alphabet.

The proof rests on the same ideas as that of Shannon's noiseless channel coding theorem, namely typical sequences and the laws of large numbers. The density operator $\boldsymbol{\rho}$ can be decomposed into its eigenstates $|x\rangle$ (which are orthonormal), with eigenvalues $p(x)$:

$$\boldsymbol{\rho} = \sum_x p(x)|x\rangle\langle x|.$$

Then the von Neumann entropy is equal to the Shannon entropy

$$S(\boldsymbol{\rho}) = S(\{p(x)\}).$$

We can then define a typical sequence

$$x_1, x_2, \ldots, x_n$$

of classical symbols x_i and associate with it a typical state

$$|x_1\rangle|x_2\rangle \ldots |x_n\rangle$$

in the Hilbert space $\mathfrak{H}^{\otimes n}$. The typical states span the *typical subspace* and by the laws of large numbers a few facts can be shown about the typical subspace for sufficiently large n which are very similar to the properties of the typical sequences leading to Shannon's noiseless channel coding theorem. (See [40] for a nice parallel treatment of both theorems.)

- $\boldsymbol{\rho}^{\otimes n}$ has almost all of its weight in the typical subspace:

  $$\mathrm{Tr}\mathbf{P}(n)\boldsymbol{\rho}^{\otimes n} \geq 1 - \delta \quad (\delta \to 0),$$

  where $\mathbf{P}(n)$ is the projector on the typical subspace.

- The dimension of the typical subspace is asymptotically $2^{nS(\boldsymbol{\rho})}$:

  $$\mathrm{Tr}\mathbf{P}(n) \approx 2^{nS(\boldsymbol{\rho})},$$

  implying that compression is possible.

- The weight of $\boldsymbol{\rho}^{\otimes n}$ in any *smaller* subspace is negligible. Let $\mathbf{Q}(n)$ be a projector on any subspace of $H^{\otimes n}$ of dimension at most 2^{nR} with $R < S(\boldsymbol{\rho})$. Then for any $\delta > 0$ and n sufficiently large

  $$\mathrm{Tr}\mathbf{Q}(n)\boldsymbol{\rho}^{\otimes n} \leq \delta$$

  implying that compression is limited: if one tries to press too hard, data will be lost.

13.4.4 Classical information over noisy quantum channels

This is a subject of ongoing research (as is, even more so, the subject of quantum information over noisy quantum channels). The usage of quantum states for information transfer offers many possibilities which do not exist classically. Many of these possibilities are related to entanglement. For example, two or more successive qubits transmitted may be entangled, and there may also be entanglement between transmitter and receiver. (This leads to the fascinating possibilities of quantum cryptography and teleportation discussed in the first part of this chapter.) Many of the schemes involving entanglement between the transmitted qubits are not explored very well. The simplest case is that of product state transmission, that is, the n-symbol quantum message is just a product state of n factors (*no* entanglement). For that case an analogy of Shannon's noisy channel coding theorem has been shown which gives a *lower bound* for the capacity of a noisy quantum channel. That lower bound is known as the Holevo–Schumacher–Westmoreland (HSW) bound. Some researchers suspect that the bound is in fact the exact value of the capacity, but this has not yet been proved. Details on the HSW theorem, together with some simple examples, can be found in [40].

Problem

Calculate the reduced density operators for the states (13.1) and show that they do not depend

on the classical bits to be transmitted.