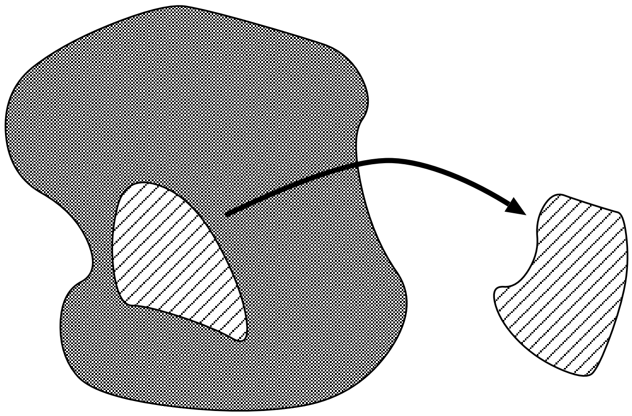
Lossless Quantum Data Compression and Secure Direct Communication

by Kim Boström



Potsdam, Germany, January 2004

Contents

Introduction

Quantum information theory is the combination of quantum mechanics and information theory. The profit is on both sides: quantum mechanics gains valuable aspects concerning the physical interpretation of the theory, and information theory gains enhanced capabilities of information processing and communication. Classically, it is the logical yes/no decision, the *bit*, which forms the elementary unit of information. A sequence of bits forms the basic object of information theory, the *message*. A message can be composed and read out by addressing each bit individually. In quantum information theory the elementary unit of information is the *qubit*, which represents the linear *superposition* of yes and no. A sequence of qubits forms the *quantum message* and here another phenomenon shows up: *entanglement*. A quantum message consisting of entangled qubits contains "non-local" information, which means that the information cannot be stored and read out by addressing each qubit individually, but only by performing a joint operation on all qubits. There is no analogon for this in classical information theory.

The extension of information theory to quantum information theory enables us to search for new algorithms and communication protocols. Shor's factoring algorithm [1] shows that the capabilities of a quantum computer exceed those of a classical computer. The Shor algorithm represents the only known efficient algorithm for prime number factorization. Here, "efficient" means that the time needed for the computation is a polynomial function of the input length. Any superpolynomial relation between input length and computation time implies that the algorithm is "inefficient". The origin for the gap of efficieny between classical and quantum computers is the fact that it is impossible to efficiently simulate a quantum computer on a classical computer. The input of a quantum computer is a sequence of qubits, say of length N. The corresponding Hilbert space is of dimension 2^N, where each dimension represents one degree of freedom. A classical computer has to carry out calculations by addressing any of these 2^N degrees of freedom. In contrast to that, the quantum computer performs his algorithms by addressing the N qubits only. The quantum processor does not have to know about linear algebra: the laws of quantum mechanics do the job for free.

If we would be able to build a quantum computer with enough resources, then by using the Shor algorithm the most used public key cryptosystem, the RSA protocol [2], would no longer provide security. On one hand quantum information theory dooms the RSA cryptosystem, on the other hand it promotes a potential successor: the BB84 protocol [3]. This communication protocol can be used to distribute a secret random key between two parties. The secret key then enables the parties to communicate in a perfectly secure manner. Using a beam of single polarized photons, a random sequence of bits can be distilled which is completely unavailable to any external non-authorized person. In its present stage the protocol is implemented over distances of 67 kilometres using an optical fibre [4]. A free-space connection has been established up to 10 km [5]. If an earth-based station could be connected through the air with a satellite system using BB84, then a perfectly secure communication of messages all over the world would become possible. But already the mere existence of the BB84 protocol solves a crucial theoretical problem, namely the so-called "key distribution problem" which has puzzled the crypto-scientists for decades: Perfectly secure classical communication protocols rely on the existence of a shared secret key. How can this key be established without having an unauthorized person potentially eavesdropping it? The BB84 protocol now offers a method to distribute a random key between two parties which is secure against eavesdropping. Based on the fact that any measurement unavoidably disturbs the measured system, the presence of an external eavesdropper can be revealed, in which case the protocol is aborted, so the eavesdropper is left with a useless random string.

As one can see by these examples, the field of quantum information theory represents an exciting research field with significant theoretical and practical implications. My contribution to the field concerns the encoding and transmission of information via a quantum channel. I have studied the implementation of so-called "variable-length codes" on quantum systems [P1], a topic which is quite new to quantum information theory. Using variable-length codes it is possible to compress and decompress quantum data without any loss of information. Such cannot be provided by use of the traditional block compression scheme also known as *Schumacher compression* [6], which is lossless only in the asymptotic limit of infinitely long messages. Furthermore, I have developed and investigated a novel quantum cryptosystem – the *ping-pong protocol* – which enables the secure direct communication of classical messages from one party to another without the need to establish a shared secret key [P2].

The thesis is divided into three parts. The first part presents the basic concepts of classical information theory and shows how they are transferred to the quantum domain. The second part is concerned with the question how to compress quantum data, especially if the compression takes place without any loss of information. Methods to implement variable-length codes on a quantum system are proposed and discussed and the limitations of these codes are investigated. The third part deals with quantum cryptography. After a brief review of the issue, the ping-pong protocol, which is a secure communication protocol based on an entangled pair of qubits, is presented and its security and practicality are discussed.

Main Results

Lossless quantum data compression

The first main result is the development of a suitable formalism to describe and categorize *variable-length quantum codes*, which represent the basis for *lossless quantum data compression* [P1, P3, P4]. In the classical case, there are two different types of compression codes: *Lossy* and *lossless* codes. Lossy codes (like the JPEG algorithm for images) reach very high compression rates at the cost of losing information. Lossless codes (like the GIF algorithm for images), reach in most cases high compression rates and preserve the information. It is clear by a simple counting argument that not all messages can be compressed in a lossless way. There are always messages (in fact the vast majority) that are *expanded* by the algorithm. Luckily, these messages are mostly useless (e.g. images of white noise) and very unlikely to appear in everyday's use. In the quantum domain, there is no much variety of compression codes. The authoritative quantum compression code, the *Schumacher code* [7, 6, 8], is modeled along the line of Shannon's source coding theorem [9] using a fixed-length block code. However, such code is, by construction, lossy for a finite number of letters and is very difficult to implement. So the development of novel quantum compression codes that are effective and easy to implement represents an important task [10, 11].

As is known from classical information theory, one can only compress data without loss of information if one maps the source letters to codewords of variable length. Source letters appearing with high frequency are mapped to shorter codewords than letters of rare use. In the quantum domain, superpositions of letters are allowed, so one in general obtains a superposition of codewords of distinct length. Such states show a high degree of entanglement, not only with respect to the letter values but also with respect to the *number* of letters in the message, which is measured by a quantum mechanical *length operator* acting on a Fock space. The best known classical example of a lossless code is the *Huffman code* [12] which has been transferred to the quantum domain firstly in [10]. A further development of quantum variable-length codes has been made in [11]. The papers [P1, P3, P4] put the formalism of quantum variable-length coding on a generalized basis and derive some important theorems. In particular, it has been shown that it is impossible to compress an unknown quantum message without loss of information. Instead, the sender must have apriori[1] knowledge about the source message, and then he uses a classical side-channel to store the length information about

[1] Here and in the following I use "apriori" as an anglified version of the latin word "a priori" which means "beyond experience". The anglified version can also be transformed into the adverb form "apriorily". The same applies to "aposteriori". Such anglifications might be unsatisfactory for language purists, but I prefer them for the sake of practicality.

the encoded message. As a pleasant side effect, the amount of information transferred through the quantum channel can occasionally be compressed below the von-Neumann entropy of the source message ensemble. As could be shown, the *sum* of quantum and classical information is bounded from below by the von-Neumann entropy. This theorem represents a quantum analog of Shannon's source coding theorem for lossless codes. An explicit algorithm has been proposed that realizes lossless quantum data compression for any given source ensemble.

Secure direct communication

The second main result concerns the construction and discussion of a novel quantum cryptographic protocol. This so-called "ping-pong protocol" developed in [P2] uses entanglement as a resource for secure communication. Similiar to the E91 protocol proposed by Ekert [13] and simplified by Bennet et al. [14], the protocol is based on an entangled pair of qubits. In contrast to E91, the entanglement is not destroyed during the transmission. Instead, one of the photons, the *travel qubit*, is transmitted to Alice and the other one, the *home qubit*, is stored by Bob. Alice encodes one bit of information by applying one of two unitary transformations to the travel qubit, and then she sends the qubit back to Bob. By performing a Bell measurement, Bob decodes the stored information. With a certain control probability, Alice and Bob perform a synchronized measurement, which reveals the presence of an external eavesdropper with nonzero probability. It can be proven that the security against arbitrary eavesdropping attacks is provided for the idealized case of a perfect quantum channel. The scheme is designed in the spirit of the *quantum dense coding scheme* raised by Bennett and Wiesner [15]. In contrast to their conclusion that such protocol can only be used for key distribution and is thus equivalent to the BB84 scheme, the ping-pong protocol represents an improvement over BB84 in the following way. First, it is *deterministic*, so Alice is really able to send a particular non-random message to Bob. Each letter of the message corresponds to a quantum state which is prepared by Alice, sent through the quantum channel and is then decoded by Bob. The protocol is more *effective*, because no qubits have to discarded. The protocol is *instantaneous*, that is, Bob is able to decode the message while receiving it. This feature distinguishes the ping-pong protocol from other deterministic quantum cryptographic protocols [16]. Because of its deterministic nature, one can use the ping-pong protocol for *quasi-secure direct communication*, that is, the message can be sent directly with a considerable degree of security. For example, the probability to eavesdrop the word "Hello" (40 bits) without being detected is about 10^{-7}, which is of the same magnitude as the chance to get killed by a lightning flash. If the protocol is used for the transmission of a random key, then one obtains *perfect security*, which means that a key of infinite length can only be eavesdropped with zero probability. It remains an important issue to proof the security of the ping-pong protocol also for the case of an *imperfect* quantum channel. Wojcik [17] has studied an attack scenario on the ping-pong protocol that exploits channel losses in order to eavesdrop information without being detected. The author could then show that a slight modification of the control mechanism closes this security hole. Cai [18] has pointed out that it is possible to *disturb* the communication between the two parties without being detected. He also pointed out, however, that an additional classical method of message authentification suffices to overcome the problem. The ping-pong protocol is experimentally realizable with relatively small effort. The collaboration with an experimentalist group at the University of Potsdam has already started. The aim is to realize the ping-pong protocol by experiments and to explore the potential of future commercial applications.

Publications

Papers

[P1] K. Bostroem and T. Felbinger. Lossless quantum data compression and variable-length coding. *Phys. Rev. A* **65**, 032313 (2002).
LANL e-print quant-ph/0105026.

[P2] K. Bostroem and T. Felbinger. Deterministic secure direct communication using entanglement. *Phys. Rev. Lett.* **89**, 187902 (2002).
LANL e-print quant-ph/0209040.

Preprints

[P3] K. Bostroem. Concepts of a quantum information theory of many letters. (2000).
LANL e-print quant-ph/0009052.

[P4] K. Bostroem. Lossless quantum coding in many-letter spaces. (2000).
LANL e-print quant-ph/0009073.

[P5] K. Bostroem. Quantizing Time. (2003).
LANL e-print quant-ph/0305128.

[P6] K. Bostroem. Time symmetry in Rigged Hilbert Spaces. (2001).
LANL e-print quant-ph/0005024.

Selected Talks

[T1] "The Ping-Pong Protocol". Invited talk, University of Camerino, Italy (Dec 2003).

[T2] "Quantizing Time". Invited talk, Max-Planck Institut Dresden, Germany (May 2003).

[T3] "Quantum Codes". DFG project presentation, Hünfeld, Germany (Jan 2003).

[T4] "Secure direct communication using entanglement". Invited talk, Ludwig-Maximilians-University Munich (April 2002).

[T5] "Time in quantum mechanics". DPG seminar talk, Osnabrück, Germany (March 2002).

[T6] "Lossless Quantum Data Compression and variable-length Coding". Invited talk, Imperial College, London (Feb 2002).

[T6] "Digging in the Mud – A Time Operator in Quantum Mechanics". A2 seminar talk, TU Berlin, Germany (Dec 2001).

Posters

[Po1] "Quantum Codes". DFG meeting, Hünfeld, Germany (Jan 2003).

[Po2] "Lossless Quantum Data Compression and variable-length Coding". DPG meeting, Bad Honnef (May 2001).

Part I

Concepts of Information

Chapter 1

Classical Information

1.1 Communication

Communication is the transmission of messages – text, thoughts, ideas, pictures, sounds, speech, whatever – from one party to another. These parties are usually referred to as 'Alice' and 'Bob', but they do not need to be human beings. In fact, Alice and Bob can be any kind of information processing systems, though it does not hurt to imagine them as human beings. Alice wants to transmit a particular message to Bob. How can this be accomplished, or better: What do we actually mean by "transmission"? Let us agree that the transmission is successful if Bob has *reproduced* the message that Alice wants to transmit. For example, if the message is a *thought* in Alice's mind then the transmission is successful if eventually Bob has the same thought in his mind. Since a thought cannot be delivered like a parcel, there must be another way to transmit it. The crucial trick is to *encode* the message, i.e. to translate it into a *signal*. If the signal is a continuous function of continuous parameters, then we speak of an *analog signal*. If the signal is a discrete sequence of finitely many signal elements, then it is a *digital signal*. (For example, the tracks on a record represent an analog signal, while the tracks on a CD represent a digital signal.) Once Alice has encoded her message into a signal by use of an *encoder*, this signal can be sent to Bob using a *channel*. A channel is a physical system which is able to transport signals of a certain kind. When the signal has reached Bob's side then he must *decode* it by use of a *decoder*. The decoder takes the incoming signals and applies the inverse encoding operation, so that he recovers the source message from the signal. In order for the transmission to work, the encoder and decoder must be adapted to the channel. The signal can be transmitted by the channel in space and in time. (For example, a telephone line transmits signals mainly in space, while a floppy disk transmits them mainly in time.) Each message has been *formed into* the channel by modification of the channel's state, so the "*information*" in the channel is the message. The channel is also called a "carrier of information" or "medium". Any physical system can act as a channel and the nature of the channel determines the nature of the information. In particular, if the channel is a *quantum* system then the information that is stored on this medium is *quantum information*. (Strictly speaking, every system is a quantum system, though if the carrier system is macroscopic enough then its quantum nature does not show up, so the information is effectivelly classical.)

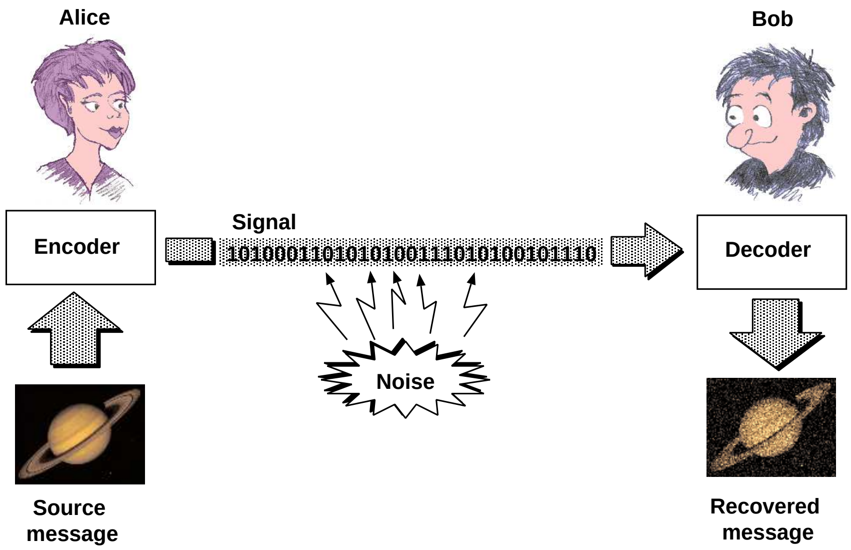


Figure 1.1: The general model of communication introduced by Shannon. Alice encodes a source message (here a picture of saturn) into a signal by use of an encoder. Here the channel is a binary channel, so the signal is a sequence of zeroes and ones. The signal is transmitted through a channel to Bob. In general the channel is exposed to noise, so the signal is distorted. Bob receives the signal and recovers the source message by use of a decoder.

If encoding, transmission and decoding succeeds then Alice has communicated a message to Bob. Though there are some obstacles to a successful communication. First, the code might be "lossy", i.e. distinct source messages are not all mapped to distinct signals, so the decoding procedure will not always reproduce the correct source message. (By far most codes are of the lossy type, especially if the channel is a digital one. For example, it is impossible to map the sound of a violin to a sequence of binary digits without losing information. It is also clear that language is a quite lossy code for thoughts, as many misunderstandings among people impressingly show.) Second, there can be *noise* on the channel. A *noisy channel* transforms with a certain probability one signal into another so that the decoding procedure will fail to reproduce the correct source message. All real-world channels are noisy, but in order to describe the principle mechanism of the channel it is convenient to consider the noiseless version. Now this is roughly sketched the general model of communication as it has been introduced 1948 by Claude Shannon [9], and it is illustrated in Fig. 1.1.
Let us comprise the main notions in this section:

> *A channel is a physical system that is used for the transmission of messages from one party to another. A message is an arbitrary object that can be encoded into a signal which is then transmitted through the channel. The transmission is successful if the message is reconstructed on the receiver's side of the channel.*

In the following we will restrict ourselves to *digital channels*, i.e. channels that transmit only *digital signals*. These are discrete sequences of symbols taken from an *alphabet*.

The symbols are also called *letters* and the sequences are also called *messages*. In order to distinguish the message that Alice wants to communicate with the message that is transmitted through the channel in form of a signal, we call them the *source message* and the *code message*, respectively.

1.2 Codes and messages

Codes are everywhere. There is a code for bank accounts, telephone devices and inhabitants of a country, there even is a code for living beings: the genetic code. Writing is a code for language, which is a code for thoughts, which is a code for imaginations, considerations, emotions and sensations. The reason why there are so many codes is that there is no communication without codes. One could go so far as to say that *reality* is a product of our communication with nature, and that our description of reality is merely a description of the codes that we can handle. These codes are determined by our channels of *perception* (eyes, ears, etc.) and also by our channels of *reflection* (neurons, axions, etc.).

In the context of digital commmunication, a *code* maps source messages of any kind to messages consisting of a discrete sequence of symbols taken from a finite alphabet. The source message can be a text, a picture, a thought, a sound or whatever, and it is chosen by the sender from a message set $\mathcal{M}$. A (classical digital) *message* is an ordered sequence $\boldsymbol{x} = x_1x_2x_3\cdots$ of symbols x_n taking from an alphabet $\mathcal{A} = \{a_1, \ldots, a_K\}$ of finite size $|\mathcal{A}| = d$. A finite message of length N is denoted by $\boldsymbol{x} = x^N$ and the set of all messages of length N is denoted by $\mathcal{A}^N$. The *empty message* is denoted by $x^0 = \Box$ and the corresponding set by $\mathcal{A}^0 = \{\Box\}$. The set of all finite messages that can be composed from the alphabet $\mathcal{A}$ is denoted by

$$\mathcal{A}^+ := \bigcup_{N=0}^{\infty} \mathcal{A}^N. \tag{1.1}$$

A *code* or *source code* is a map $c : \mathcal{M} \to \mathcal{A}_c^+$ from a set $\mathcal{M}$ of source messages into the set $\mathcal{A}_c^+$ of finite messages composed from a code alphabet $\mathcal{A}_c$. With the size of the alphabet given by $|\mathcal{A}_c| = K$ the code c is a *K-ary code*. The set $\mathcal{C} = c(\mathcal{M})$ is called the *codebook* and each member is a *code message* or *codeword*. If the codebook only contains messages of some fixed length N then c is a *block code*, otherwise it is a *variable-length code*. A code is *lossless* or *non-singular* if there are distinct codewords for distinct messages, i.e.

$$\forall x, y \in \mathcal{M} : \; x \neq y \Rightarrow c(x) \neq c(y). \tag{1.2}$$

In case of a *lossy* or *singular* code, the above condition is hurt, so some messages are mapped to the same codeword. The decoder then can map the codeword back to a representative source message being in the equivalence class of source messages that are mapped to the same codeword, or he may return an error or an empty message. Lossy codes are useful when it is more important to reduce the size of the message than to ensure the correct decoding (a good example is the MP3 code for sound data).

In practice one is often interested in translating one message into another message. This special type of code is based on the concept of a *symbol code*. A symbol code

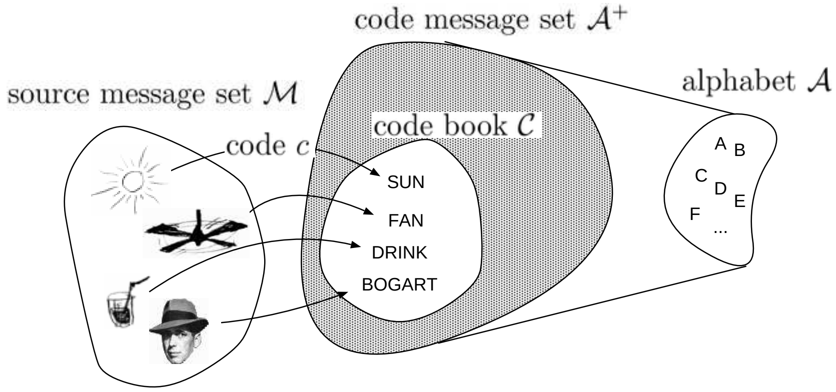


Figure 1.2: A code is a map from a set of source messages into a set of code messages composed from an alphabet. For block codes the length of the code messages is fixed, for variable-length codes it varies.

c translates each symbol from a source alphabet $\mathcal{A}$ into a codeword over some code alphabet $\mathcal{A}_c$, i.e. $c : \mathcal{A} \to \mathcal{A}_c^+$. The *extension* c^+ of a symbol code c is a map from the set of all finite messages over the source alphabet to finite messages over the code alphabet by concetenation of the individual codewords, $c : \mathcal{A}^+ \to \mathcal{A}_c^+$, where

$$c^+(x_1 \cdots x_N) := c(x_1) \cdots c(x_N). \tag{1.3}$$

A symbol code is called *uniquely decodable* if its extension is non-singular. Therefore, uniquely decodability is stronger than non-singularity. A symbol code might be lossless, i.e. non-singular, on the source alphabet, but its extension might fail to be lossless because the concetenation of codewords lead to ambiguities. If we are given a non-singular symbol code, can we decide if it is also uniquely decodable? There is an important theorem called the *McMillan theorem* which yields a criterion for the uniquely decodability of symbol codes.

Theorem 1 (MacMillan) If the symbol code $c : \mathcal{A} \to \mathcal{A}_c^+$ is *uniquely decodable* then the codeword lengths $l_c(x)$ fulfill the *Kraft inequality*

$$\sum_{x \in \mathcal{A}} |\mathcal{A}_c|^{-l_c(x)} \leq 1. \tag{1.4}$$

Conversely, given a set of codeword lengths that satisfy the Kraft inequality then there exists a uniquely decodable code with these codeword lengths.

A proof of this theorem can be found in [12], pp.90. A uniquely decodable code which satisfies the Kraft inequality with *equality* is called a *complete code*. Uniquely decodability means that distinct source messages are mapped to distinct code messages, but one might have to look at the *entire* code message to decode the source message. In practice this means that one would have to wait until the transmission is completed before one can start the decoding process. A code which avoids this difficulty is the *prefix code*. A prefix code is a symbol code where no codeword is the prefix of any other codeword. Prefix codes are *instantaneous*, which means that they can be decoded on the fly, i.e. letter by letter. Block codes are also prefix codes. A prefix code is *self-separating*, as an example consider the variable-length code

$$c(\mathtt{A}) = \mathtt{0} \tag{1.5}$$
$$c(\mathtt{B}) = \mathtt{10} \tag{1.6}$$
$$c(\mathtt{C}) = \mathtt{110} \tag{1.7}$$
$$c(\mathtt{D}) = \mathtt{111}. \tag{1.8}$$

This code is obviously a prefix code. The message 01011111010 is decoded as

$$\mathtt{01011111010} \to \mathtt{0\ \ 10\ \ 111\ \ 110\ \ 10} \to \mathtt{ABDC}. \tag{1.9}$$

Although prefix codes form a proper subset of uniquely decodable codes, the MacMillan theorem can also be formulated for prefix codes:

Theorem 2 (Prefix codes and Kraft inequality) If the symbol code $c : \mathcal{A} \to \mathcal{A}_c^+$ is a prefix code then the codeword lengths $l_c(x)$ fulfill the Kraft inequality (1.4). Conversely, given a set of codeword lengths that satisfy the Kraft inequality then there exists a prefix code with these codeword lengths.

Let us not ignore another type of code which is frequently used in practice but rarely disussed in the theory, because it is not so effective. It is a code which reserves a special symbol in the code alphabet as a *separator* between the codewords. Let us use the name *separator code* and define it by

$$C_\odot(x_1 \cdots x_N) := c(x_1) \odot c(x_2) \odot \cdots \odot c(x_N), \tag{1.10}$$

where "$\odot$" represents the separator symbol in the code alphabet $\mathcal{A}_c$. A famous example for a separator code is the english writing: Sequences of grammatical elements are translated into codewords composed from the english alphabet and the codewords are separated by a blank space. Another example is the Morse code, where a pause in the stream of dots and dashes serves as a codeword separator. Needless to say, also separator codes are instantaneous, because the decoder can separate the incoming stream of symbols whenever he receives the separator symbol.

In the context of quantum compression we will face yet another possibility to separate the codewords, namely by use of a *side channel* that stores the length information of each codeword. The decoder receives the length information through the side channel and then separates the incoming stream of codewords in the main channel at the right places. Also this kind of code is instantaneous.

Still another class of codes is of interest. Instead of encoding each symbol in the source message separately we can devide the source message into blocks of length k and consider each block as a *supersymbol* from an alphabet $\mathcal{A}^k = \{x_1 \cdots x_k \mid x_n \in \mathcal{A}\}$ which we then assign a codeword. Let us call such code a *supersymbol code* c_k. Its extension c_k^+ is then defined by

$$c_k^+(x_1 \cdots x_N) := c_k(x_1 \cdots x_k) \cdots c_k(x_{N-k+1} \cdots x_N), \tag{1.11}$$

where k must be a divisor of N. The "biggest" supersymbol code would be a code with source block length $k = N$, i.e. every possible source message of length N is assigned to its own codeword. Since the number of possible source messages of length N is *vast* for N reasonably large, such a code is of no practical interest. In a theoretical context, however, it can be quite useful, as we will see when proving Shannon's source coding theorem.

1.3 Information content

It is not so easy to define the "information content" of a message and there are many sophisticated approaches. However, with the help of Shannon's communication model it is possible to define the information content of a message in a straightforward way:

> *The information content of a message is a measure for the effort of communicating it.*

This effort of communication is determined by the *code* that is used for representing the message. So let us define the *code information content* or *size* of an individual message $m \in \mathcal{M}$ for a given code c over the alphabet $\mathcal{A}_c$ by

$$\mathcal{I}_c(m) := \log_2 |\mathcal{A}_c| \cdot L_c(m), \tag{1.12}$$

where $L_c(m)$ is the length of the codeword for m. The above expression can be interpreted as a measure for the effort of communicating the message m by use of the code c, and therefore as its *size*, i.e. the space it would occupy on a hard disk. The factor $\log_2 |\mathcal{A}_c|$ accounts for the resources which are occupied by encoding and decoding the message and the factor $L_c(m)$ accounts for the effort of transporting the codeword through the channel. The advantage of the code information measure $\mathcal{I}_c(m)$ is that it is defined for *individual messages* and that it is *independent of statistics*. It only depends on the *code*, reflecting the philosophy that there is no information contained in an object without a code giving it some *meaning*. (For example, the codeword "`XWF$%&$FggHz`" may be a random message of symbols or may in a certain code represent the first digits of π or in another code the beginning of a Mozart symphony.) If c^+ is an extension of the symbol code c then the code information content of a message is equal to the sum of the information contents of the individual symbols:

$$I_{c^+}(x_1 \cdots x_N) = \mathcal{I}_c(x_1) + \cdots + \mathcal{I}_c(x_N). \tag{1.13}$$

1.4 Compression

Compression means reducing the effort of communication *on average*. With respect to a given fixed code alphabet, compression means *reducing the expected length of the encoded message*. In order to find out if a code is compressive or not, we need *statistical* knowledge about the source message ensemble.

Assume that we are given a discrete set $\mathcal{M}$ of possible source messages and a *probability distribution* $p : \mathcal{M} \to [0, 1]$. The message $m \in \mathcal{M}$ is sent with "*apriori probability*"

$p(m)$ where

$$\sum_{m \in \mathcal{M}} p(m) = 1. \tag{1.14}$$

The pair $M = (\mathcal{M}, p)$ is a *statistical ensemble* and any function $f : \mathcal{M} \to \mathbb{R}$ is a *random variable*, symbolized by $f(M)$. The *expectation value* of the random variable $f(M)$ is given by

$$\langle f(M) \rangle := \sum_{m \in \mathcal{M}} p(m)\, f(m), \tag{1.15}$$

and the *uncertainty* of $f(M)$ is defined as

$$\Delta f(M) := \sqrt{\langle f^2(M) \rangle - \langle f(M) \rangle^2}. \tag{1.16}$$

Since the code information I_c is a real-valued function on $\mathcal{M}$, it represents a random variable, and we define the *code information content* $\mathcal{I}_c(M)$ or *size* of the message ensemble M using the code c as the expectation value of $\mathcal{I}_c(M)$:

$$\mathcal{I}_c(M) := \langle \mathcal{I}_c(M) \rangle = \log_2 |\mathcal{A}_c| \sum_{m \in \mathcal{M}} p(m)\, L_c(m) \tag{1.17}$$

$$= \log_2 |\mathcal{A}_c| \cdot \langle L_c(M) \rangle. \tag{1.18}$$

Let us understand a *raw code* as a code that assigns one symbol for each message in $\mathcal{M}$. Hence, the code alphabet is of the same size as the source message set, $|\mathcal{A}_c| = |\mathcal{M}|$, and the codeword length is $L_c(m) \equiv 1$, so the ensemble information for a raw code is equal to

$$\mathcal{I}_0(M) := \log_2 |\mathcal{M}|, \tag{1.19}$$

which we call the *raw information content* or *raw size* of the ensemble M. The raw information content represents the *uncompressed size* of a message ensemble and is a useful notion that serves as a reference for the *effectivity* of a code, which we define as

$$\eta_c(M) := \frac{\mathcal{I}_0(M)}{\mathcal{I}_c(M)}. \tag{1.20}$$

A code c is *compressive* on the ensemble M exactly if it fulfills

$$\eta_c(M) > 1, \tag{1.21}$$

or equivalently

$$\mathcal{I}_c(M) < \mathcal{I}_0(M). \tag{1.22}$$

A code with an effectivity smaller than 1 is *expansive*. Expansive codes add *redundancy* to the messages, which can be very useful especially in the presence of noise. In this context the *error correcting codes* are of particular interest.

The task of compression is thus bringing the size of the message ensemble, given by $\mathcal{I}_c(M)$, below the raw information content $\mathcal{I}_0(M)$. There are two ways of compressing a message ensemble M: *block compression* and *variable-length compression*.

Block compression

If we try for block compression, then all the codewords must have the same length L. Thus with $\mathcal{A}_c$ being the code alphabet we have $|\mathcal{A}_c|^L$ distinct codewords. If we want a *lossless* code, then we have to encode all possible messages, which requires

$$|\mathcal{A}_c|^L \geq |\mathcal{M}|. \tag{1.23}$$

Taking the binary logarithm of the above relation we obtain

$$L \cdot \log_2 |\mathcal{A}_c| \geq \log_2 |\mathcal{M}| = \mathcal{I}_0(M). \tag{1.24}$$

The code information content of the block code reads

$$\begin{aligned}
\mathcal{I}_c(M) &= \log_2 |\mathcal{A}_c| \sum_{m\in\mathcal{M}} p(m)\, L_c(m) &(1.25)\\
&= L \cdot \log_2 |\mathcal{A}_c| \sum_{m\in\mathcal{M}} p(m) &(1.26)\\
&= L \cdot \log_2 |\mathcal{A}_c| &(1.27)\\
&\geq \mathcal{I}_0(M), &(1.28)
\end{aligned}$$

where in the last line we used (1.24). Thus there is no lossless compression possible when using a block code. If we anyway want to compress the message ensemble using a block code, we have to choose a *lossy* code. Let us assume that only messages in some set $T \subset \mathcal{M}$ are encoded, then the loss of information is measured by the *fidelity* of the code c as the *probability of successful decoding*, given by

$$F_c = \sum_{m\in T} p(m). \tag{1.29}$$

The fidelity is connected with the *probability of failure*, ϵ, of decoding the correct source message by

$$F_c = 1 - \epsilon. \tag{1.30}$$

Now assume that the set T is a set of *typical messages* which contains almost all the probability,

$$\sum_{m\in T} p(m) \approx 1, \tag{1.31}$$

i.e. ϵ is very small. Since T is assumed to be smaller than the source message set, $|T| < |\mathcal{M}|$, the code fulfills the condition

$$|T| \leq |\mathcal{A}_c|^L < |\mathcal{M}|, \tag{1.32}$$

which implies

$$L \cdot \log_2 |\mathcal{A}_c| < \log_2 |\mathcal{M}|. \tag{1.33}$$

The code information content now reads

$$\begin{aligned}
\mathcal{I}_c(M) &= \log_2 |\mathcal{A}_c| \sum_{m\in T} p(m)\, L_c(m) &(1.34)\\
&= L \cdot \log_2 |\mathcal{A}_c| \sum_{m\in T} p(m) &(1.35)\\
&\approx L \cdot \log_2 |\mathcal{A}_c| &(1.36)\\
&< \mathcal{I}_0(M), &(1.37)
\end{aligned}$$

where in the last line we used (1.33). Thus by neglecting irrelevant information (the messages outside of T) we can achieve compression by using a block code.

Variable-length compression

The second kind of compression is *variable-length compression*. Here the lengths of the codewords are allowed to vary. In case of a *lossless* code there must be a codeword for every message in $\mathcal{M}$. As we can see by definition (1.12), in order to compress the messages without losing information we have to reduce the *expected length* $L = \langle L_c(M)\rangle$ of the codeword while encoding every message. Consequently, highly probable messages obtain short codewords and less probable messages obtain longer codewords. This is the general strategy in variable-length compression.

1.5 Random messages

A type of message ensemble of particular interest is the *random message* $\boldsymbol{X} = X_1X_2\cdots$, which is a sequence of statistically independent ensembles X_n of symbols x_n drawn from an alphabet $\mathcal{A}$ with a certain apriori probability distribution $p(x_n)$. A random message of length N is denoted by $\boldsymbol{X} = X^N$ and the probability distribution on the resulting set of possible sequences is given by

$$p(x_1\cdots x_N) = p(x_1)\cdots p(x_N). \tag{1.38}$$

The random message can be regarded as a first approximation to a real-world message. For example, each language has its own characteristic apriori probability distribution of symbols in a written text. By determining the relative frequencies of the symbols appearing in a given text one can estimate the corresponding apriori probabilities and gain information about the corresponding language.
Let c^+ be the extension of a symbol code c on the source alphabet $\mathcal{A}$, then the length of the code sequence for a source sequence $x^N = x_1\cdots x_N$ is equal to the sum of the lengths of the individual codewords,

$$L_{c^+}(x^N) = \sum_{n=1}^{N} l_c(x_n). \tag{1.39}$$

The average length of a code sequence is thus given by

$$\langle L_{c^+}(X^N)\rangle = \langle\sum_{n=1}^{N} l_c(X_n)\rangle = N\cdot\langle l_c(X)\rangle. \tag{1.40}$$

As a consequence, for symbol codes the code information content of the sequence ensemble X^N is N times the code information content of the individual symbol ensemble,

$$\mathcal{I}_{c^+}(X^N) = \langle\mathcal{I}_c(X^n)\rangle = N\log_2|\mathcal{A}_c|\cdot\langle\mathcal{I}_c(X)\rangle \tag{1.41}$$

$$= N\cdot\mathcal{I}_c(X), \tag{1.42}$$

or equivalently

$$\mathcal{I}_c(X) = \frac{1}{N}\mathcal{I}_{c^+}(X^N). \tag{1.43}$$

The same goes for the raw information content,

$$\mathcal{I}_0(X^N) = \log_2(|\mathcal{A}|^N) = N \log_2 |\mathcal{A}| = N \cdot \mathcal{I}_0(X). \tag{1.44}$$

1.6 Shannon's source coding theorem

Shannon's famous source coding theorem [9] establishes a link between the average information per symbol and the Shannon entropy of the symbol ensemble in case of asymptotically faithful codes. A code is *asymptotically faithful* if its decoding fidelity reaches unity in the limit of infinitely long source messages,

$$F_c \to 1 \quad \text{for } N \to \infty. \tag{1.45}$$

With the help of the code information defined by (1.17), the theorem can be formulated as follows.

Theorem 3 (Source coding theorem) *Given a symbol ensemble $X = (\mathcal{A}, p)$. There is a an asymptotically faithful code c on the random message X^N such that the code information content per source symbol approaches for long messages the Shannon entropy, i.e.*

$$\frac{1}{N}\mathcal{I}_c(X^N) \to H(X) \quad \textit{for } N \to \infty \tag{1.46}$$

where

$$H(X) := -\sum_{x \in \mathcal{A}} p(x) \log_2 p(x) \tag{1.47}$$

is the Shannon entropy *of the ensemble X. It is impossible to faithfully compress below $H(X)$ bits per symbol, i.e.*

$$\frac{1}{N}\mathcal{I}_c(X^N) \geq H(X), \tag{1.48}$$

for any asymptotically faithful code c.

Shannon's theorem thus states that it is asymptotically possible to compress messages down to $H(X)$ bits per symbol. If we compress the messages below $H(X)$ then we will lose all information in the asymptotic limit.

1.6.1 Block compression

One way to derive the source coding theorem is to use *block compression*. As we have seen above, this requires that the code is *lossy*, i.e. we cannot encode all source messages. Shannon's idea was to find a suitable set of *typical* messages. Since the typical messages form a *tiny* subset of all possible messages, one needs much less resources to encode them. Shannon showed that the probability for the occurence of non-typical messages tends to zero in the limit of large message lengths. Thus we have the paradoxical situation that although we "forget" to encode most messages, we lose no information in the limit of very long messages. In fact, we make use of *redundancy*, i.e. we do not encode "unnecessary" information represented by messages which almost never occur.

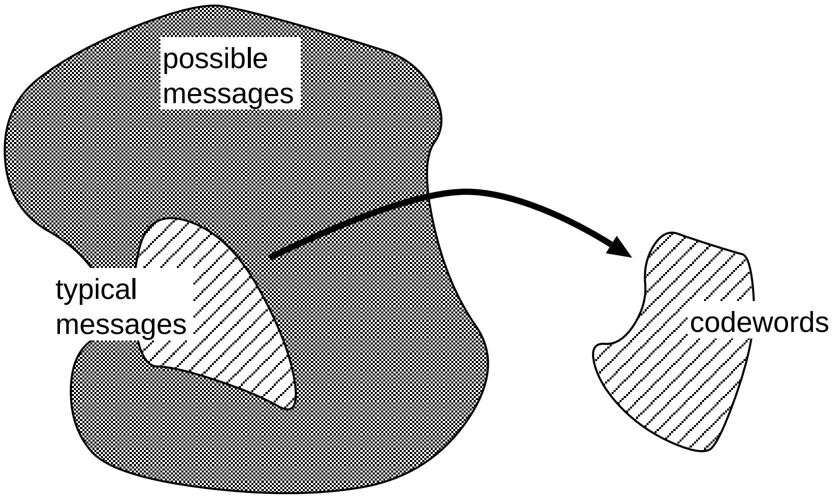


Figure 1.3: The concept of Shannon's block compression. Only typical messages are encoded into a message. Because the set of typical messages is much smaller than the set of all possible messages, the effort of communicating only typical messages is reduced. For very long messages almost every message is a typical message so the probability of a decoding error tends to zero in the limit of infinitely long messages.

Let us now give a hand-waving proof of Shannon's source coding theorem for block codes. Consider a message $\boldsymbol{x} = x^N$ with N very large. Typically, the symbol a_i will appear with the frequency $N_i \approx Np_i$. Hence, the probability of such a *typical message* is roughly

$$p_{typ} = p_1^{N_1} \cdots p_K^{N_i} = \prod_{i=1}^{K} p_i^{Np_i}. \tag{1.49}$$

The set T of typical messages contains almost all the probability,

$$\sum_{x^N \in T} p(x^N) \approx 1. \tag{1.50}$$

Because $p(x^N \in T) \approx p_{\text{typ}}$, we have

$$\sum_{x^N \in T} p(x^N) \approx \sum_{x^N \in T} p_{\text{typ}} = |T| \cdot p_{\text{typ}} \approx 1, \tag{1.51}$$

so the set T of *typical sequences* is roughly of the size

$$|T| \approx \frac{1}{p_{typ}}. \tag{1.52}$$

If we encode each member of T by a binary message we need approximately

$$\begin{aligned} I_N &= \log_2 |T| = -\log_2 p_{\text{typ}} && (1.53)\\ &= -\log_2 \prod_{i=1}^{K} p_i^{Np_i} = -N \sum_{i=1}^{K} p_i \log_2 p_i && (1.54)\\ &= N \cdot H(X) && (1.55) \end{aligned}$$

bits. Thus for very long messages the average number of bits per symbol reads

$$I = \frac{1}{N} I_N = H(X). \tag{1.56}$$

Thus we find that in the limit of infinitely long messages the information per symbol approaches the Shannon entropy of the symbol ensemble, (1.46). A good review on the issue can also be found in [19, 20], also including a rigorous proof of the source coding theorem.

1.6.2 Variable-length compression

Is Shannon's source coding theorem also valid for variable-length codes or do we need more or less resources? The answer is that the source coding theorem is also valid in this case. This thesis has a focus on variable-length codes, so let us get a little bit more into detail.
Since the code c^+ on the set $\mathcal{A}^N$ of source messages of length N shall be lossless, the corresponding symbol code c on the alphabet $\mathcal{A}$ must be uniquely decodable. Consider a uniquely decodable symbol code $c : \mathcal{A} \to \mathcal{A}_c$ where the length of the codeword for the symbol $x \in \mathcal{A}$ is given by $l_c(x)$. Define $d = |\mathcal{A}|$ then c is a K-ary code. Because c is assumed to be uniquely decodable, it follows from the MacMillan theorem that the codeword lengths satisfy the Kraft inequality (1.4),

$$Q := \sum_{x \in \mathcal{A}} K^{-l_c(x)} \leq 1, \tag{1.57}$$

which implies that

$$\log_b Q \leq 0, \tag{1.58}$$

for any logarithmic basis b. Now define *implicit probabilities* $q(x)$ by

$$q(x) := \frac{1}{Q} K^{-l_c(x)}, \tag{1.59}$$

then by taking the K-ary logarithm the above equation can be formed into

$$l_c(x) = -\log_K q(x) - \log_K Q. \tag{1.60}$$

Thus the expected length of the codeword ensemble is given by

$$\begin{aligned} L := \langle l_c(X) \rangle &= \sum_{x \in \mathcal{A}} p(x)\, l_c(x) && (1.61) \\ &= -\sum_{x \in \mathcal{A}} p(x) \log_K q(x) - \log_K Q. && (1.62) \end{aligned}$$

The *Gibbs inequality* (1.119) states that

$$\sum_x p(x) \log_K q(x) \leq \sum_x p(x) \log_K p(x), \tag{1.63}$$

for any two probability distributions $p(x), q(x)$, therefore the expected length (1.62) is bounded from below by

$$L \geq -\sum_{x\in\mathcal{A}} p(x)\log_K p(x), \tag{1.64}$$

where we have used (1.58). Because $\log_K x = \log_2 x/\log_2 d$ the above inequality implies that

$$\log_2 K \cdot L \geq H(X). \tag{1.65}$$

By definition (1.12) the code information content equals

$$\mathcal{I}_c(X) = \log_2 K \cdot L, \tag{1.66}$$

so inequality (1.65) can be written as

$$\mathcal{I}_c(X) \geq H(X). \tag{1.67}$$

Equality is achieved if and only if the individual lengths satisfy

$$l_c(x) = -\log_K p(x). \tag{1.68}$$

In other words:

> *The code information of the source symbol ensemble is for any uniquely decodable code bounded from below by the Shannon entropy of the ensemble. The bound can be reached at least in the asymptotic limit.*

This is a very important result and it justifies once more the central role of the Shannon entropy as a measure for the information content of a message ensemble.
Now let us restrict to the case of a *binary* code. Here we have $\mathcal{I}_c(X) = L$, and therefore

$$L \geq H(X). \tag{1.69}$$

There is a code which has been proposed by Shannon [9] and is therefore referred to as the *Shannon code*. Here it is:

1. Arrange the symbols $a_1, \ldots, a_K$ in the source alphabet $\mathcal{A}$ in order of decreasing probability, so that $p_1 \geq \cdots \geq p_K$.
2. Let $P_k = \sum_{i=1}^{k-1} p_i$ be the cumulative probability for the symbol a_k. The codeword for the symbol a_k is obtained by expanding P_k as a binary number up to the length

$$l_k = \lceil -\log_2 p_k \rceil. \tag{1.70}$$

This code is also known under the name *Shannon-Fano code*, because R.M. Fano had a similiar idea (which was acknowledged by Shannon [9]). By construction, the Shannon code maps each symbol $x \in \mathcal{A}$ to a binary codeword of the length

$$l_c(x) = \lceil -\log_2 p(x) \rceil, \tag{1.71}$$

therefore the codeword lengths fulfill

$$-\log_2 p(x) \leq l_c(x) \leq -\log_2 p(x) + 1. \tag{1.72}$$

Multiplying the above relation by $p(x)$ and summing over x yields

$$H(X) \leq L \leq H(X) + 1. \tag{1.73}$$

Now we still have an overhead of at most 1 bit. This overhead can be reduced by use of a *supersymbol code* (1.11), which considers each block of length k as a supersymbol from the alphabet $\mathcal{A}^k$ and encodes it into a binary sequence. The block length k must be a divisor of N. Let us choose $k = N$, then the same arguments as above lead to

$$H(X^N) \leq L \leq H(X^N) + 1, \tag{1.74}$$

where now

$$L = \langle l_c(X^N) \rangle. \tag{1.75}$$

Because the sequence X^N consists of N independent ensembles X, we have

$$H(X^N) = N \cdot H(X), \tag{1.76}$$

and thus after dividing by N the inequality (1.74) transforms into

$$H(X) \leq \frac{1}{N} L \leq H(X) + \frac{1}{N}. \tag{1.77}$$

Therefore the *expected codeword length per symbol*,

$$L_N := \frac{1}{N} L \tag{1.78}$$

obeys the inequality

$$H(X) \leq L_N \leq H(X) + \frac{1}{N}. \tag{1.79}$$

If we let $N \to \infty$ then we find that

$$L_N \to H(X), \quad N \to \infty, \tag{1.80}$$

so the Shannon code is a lossless code that is maximally compressing in the asymptotic limit. This proves the source coding theorem for the case of uniquely decodable symbol codes.

Although this is a brilliant result, the Shannon code is not an *optimal* code, i.e. a code that *minimizes* the expected length of a message ensemble. The Shannon code is only *asymptotically optimal*.

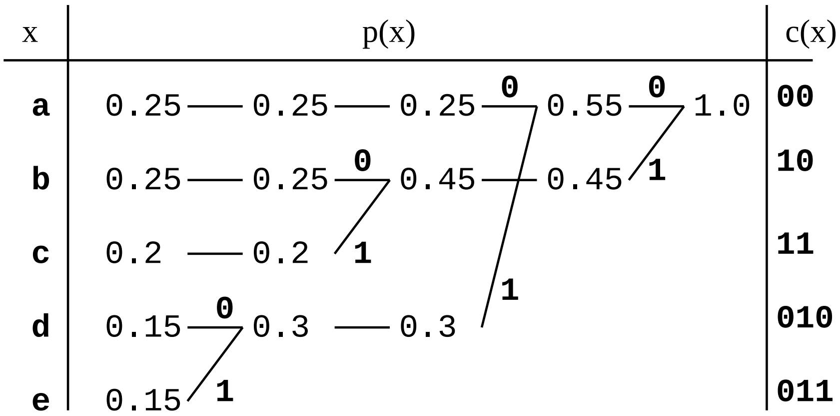


Figure 1.4: The algorithm for generating the Huffman code, operating on a particular alphabet.

The Huffman code

An optimal prefix code is given by the *Huffman code* and it is generated by the following algorithm [19]:

1. Choose two of the least probable symbols. These two symbols will be given the longest codewords, which will have equal length, and differ only in the last digit.
2. Combine these two symbols into a single symbol, and repeat.

As an example, consider the alphabet $\mathcal{A} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{K}, \mathtt{e}\}$ with the respective probabilities $\mathcal{P} = \{0.25, 0.25, 0.2, 0.15, 0.15\}$. The algorithm that generates the Huffman code is depicted in Fig. 1.4 and yields the codewords $\mathcal{C} = \{\mathtt{00}, \mathtt{10}, \mathtt{11}, \mathtt{010}, \mathtt{011}\}$. In contrast to the Shannon code, the lengths of the codewords are generally *not* equal to $\lceil -\log_2 p(x) \rceil$, instead they can be smaller or bigger. The *average* length of the Huffman codewords is smaller or equal to that of the Shannon codewords. Although the Huffman code compresses better than the Shannon code, in the context of a proof it might be more practical to use the Shannon code, right because its codeword lengths are explicitely given by $\lceil -\log_2 p(x) \rceil$.

1.7 Channels

A channel is a physical system that is able to transport the encoded message through space and time. More precisely: A source message is encoded into a code message which is transported via the channel, then the code message is decoded back into the source message. The entire procedure represents the *transmission* of the source message. We can also speak of the *storage* of the message in the channel. The message is *stored* (in an encoded form) in the channel by the sender and is *read out* by the receiver. These are all equivalent formulations.

So far we have only considered the encoding and decoding of a message, but not the *transport* of the corresponding code message through the communication channel. As long as the channel transports the message in an absolutely faithful manner, it does not play an important role indeed. But what if there is *noise* on the channel which influences the transmission?

1.7.1 Probabilities

The probability that the decoded message coincides with the source message is called the *fidelity* of the transmission. The fidelity is influenced on one hand by the nature of the *code*, e.g. if it is a lossy code, and on the other hand by the properties of the *channel*, e.g. if there is noise on the channel. The influence of the channel is described by its probability to transform one message into another. The transformation of one message x into another message y is a *transition* and is symbolized by $x \to y$. This transition can also be regarded as a *conditional event*, namely the event of receiving y when x has been sent, which is symbolized by $y|x$. Thus a conditional event $y|x$ must be read from right to left when interpreted as a transition,

$$y|x \equiv x \to y. \tag{1.81}$$

The probability for this to happen is given by the *transition probability* or *conditional probability*

$$p(x \to y) \equiv q(y|x) \in [0,1]. \tag{1.82}$$

The channel is completely described by the transition probability $q(y|x)$, which has to fulfill

$$\sum_{y \in \mathcal{Y}} q(y|x) = 1, \quad \forall x \in \mathcal{X}, \tag{1.83}$$

where $\mathcal{X}$ is the set of all possible source messages and $\mathcal{Y}$ is the set of all receivable messages. Condition (1.83) means that there must be *some* message coming out of the channel. If we wish to include *channel losses*, i.e. the case where no message is received, then we have to include the *empty message* $y^0 = \square$ in the set $\mathcal{Y}$. The *joint probability* that x is sent and y is received is then defined by

$$p(x,y) := q(y|x)p(x), \tag{1.84}$$

where $p(x)$ is the *apriori probability* that x is chosen by the sender. If $p(x) \neq 0$ then one has

$$q(y|x) = \frac{p(x,y)}{p(x)}. \tag{1.85}$$

The reader may wonder why here the joint probability is defined on the basis of the conditional probability and not vice versa. The reason is that $q(y|x)$ is determined by the channel alone, whose properties are completely independent from the apriori probabilities $p(x)$ that a particular message is chosen by the sender. From this point of view it does not make sense to consider the joint probability $p(x,y)$ as more fundamental than the transition probability $q(y|x)$. Moreover, the right-hand side of (1.85) is *undefined* for $p(x) = 0$, while there is no reason for $q(y|x)$ to be apriorily undefined, because it represents an intrinsic property of the channel and does not depend on the sender's individual decision to send or not to send particular messages. The transmission process consists of two acts, namely storing the message x in the channel and reading out the message y from the channel. The joint probability for both events, storing x and reading out y, is then given by the product of the probabilities of these two acts, namely $p(x,y) = q(y|x)p(x)$. Condition (1.83) implies for the joint probability that

$$\sum_y p(x,y) = \sum_y q(y|x)p(x) = p(x), \tag{1.86}$$

as desired. The probability $q(y)$ that the message y is received, regardless of what message is sent, is defined by

$$q(y) := \sum_x p(x,y). \tag{1.87}$$

In this context, $q(y)$ is no apriori probability, because it is *determined* by the apriori probability $p(x)$ and the transition probability $q(y|x)$. In order to avoid such confusion, let us call $p(x)$ the *input probability* and $q(y)$ the *output probability*. Finally we define the *aposteriori probability* $p(x|y)$ that the message x has been sent *provided that* the message y is received, by

$$p(x|y) := \frac{p(x,y)}{q(y)}, \tag{1.88}$$

which is thus also a conditional probability, just like $q(y|x)$. Using definition (1.84) we find the relation

$$p(x|y) = \frac{q(y|x)p(x)}{q(y)}, \tag{1.89}$$

which is also known as *Bayes' rule*, and which can be rewritten as

$$p(x|y) = \frac{q(y|x)p(x)}{\sum_x q(y|x)p(x)}. \tag{1.90}$$

Noiseless channel

A channel with the transition probability

$$q(y|x) = \delta(x,y), \tag{1.91}$$

is called a *noiseless channel*, where the discrete δ-function is straightforwardly defined for messages of any kind by

$$\delta(x,y) := \begin{cases} 1 & ; x = y \\ 0 & ; x \neq y \end{cases}. \tag{1.92}$$

It follows:

$$p(x,y) = q(y|x)p(x) = \delta(x,y)p(x), \tag{1.93}$$

and

$$q(y) = \sum_x p(x,y) = \sum_x \delta(x,y)p(x) = p(y), \tag{1.94}$$

and

$$p(x|y) = \frac{p(x,y)}{q(y)} = \frac{\delta(x,y)p(x)}{p(y)} = \delta(x,y) = q(y|x). \tag{1.95}$$

The action of a noiseless channel can be completely disregarded, just as if the message would directly go from Alice to Bob. Any other channel which does not obey (1.91) is called a *noisy channel*.

Memoryless channel

Mostly one considers messages that are finite sequences $x = x_1 \cdots x_N$ of N symbols x_n taken from some alphabet $\mathcal{A}$, and it is these sequences that can be received, i.e.

$$\mathcal{X} = \mathcal{Y} = \mathcal{A}^N. \tag{1.96}$$

The apriori probabilities are mostly that of a random message

$$p(x_1 \cdots x_N) = p(x_1) \cdots p(x_N), \tag{1.97}$$

and one says that the message is emitted from a *memoryless source*. Yet mostly, the channel is a *memoryless channel*, i.e. a channel whose transition probabilities have the form

$$p(y_1 \cdots y_N | x_1 \cdots x_N) = p(y_1|x_1) \cdots p(y_N|x_N), \tag{1.98}$$

which means that a memoryless channel transforms every single symbol while "forgetting" the other symbols. The memoryless channel is completely described by the individual transition probabilities $q(y|x)$. One sometimes adds that the above defined channel is a channel *without feedback*, which means that the input symbols do also not depend on the past output symbols.

1.7.2 Entropies

The Shannon entropy $H(X)$ of an ensemble $X = \{\mathcal{X}, p\}$ is, as we have seen, a good measure for the information content of X if it is interpreted as an ensemble of messages that are sent with a certain probability. One can faithfully compress the messages to $H(X)$ bits on average, but not below. Because it is so beautiful, here once more the Shannon entropy:

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x). \tag{1.99}$$

The Shannon entropy is non-negative,

$$H(X) \geq 0, \tag{1.100}$$

where equality is reached only in case of $p(x) = \delta(x, x')$ for some $x' \in \mathcal{X}$. (We have to be a bit careful because $\log 0$ is undefined, but a closer analysis shows that $p \log p \to 0$ for $p \to 0$.) As we see, $H(X)$ is based on the probability $p(x)$ that Alice selects a message $x \in \mathcal{X}$ and sends it to Bob through the channel. Shannon's source coding theorem showed that the optimal length of a binary codeword for the message x is $l_c(x) = -\log_2 p(x)$, therefore we are entitled to consider

$$I(x) := -\log_2 p(x) \tag{1.101}$$

as the information content of the *individual* message x from the ensemble X. We can also interpret $I(x)$ as the "surprise value" that is connected with learning the message x. The less probable the message, the more surprise when learning it. If the probability is 1, then $\log_2 1 = 0$, so there is no surprise at all. Now $I(X) = -\log_2 p(X)$ represents a random variable, so we can understand $H(X)$ as the *expectation value* of this random variable,

$$H(X) = \langle I(X) \rangle = \langle -\log_2 p(X) \rangle. \tag{1.102}$$

In other words, $H(X)$ is the *expected suprise* when learning the value of X. (A peculiar combination of words: a "surprise" that is "expected"? Probably "average surprise" would do better here.) In yet another interpretation, $H(X)$ is Bob's *apriori ignorance* about what Alice has sent. By learning the value of X this ignorance is reduced to zero. As we can see, there is a lot of approaches to information, but all of them lead to Shannon entropy.

Now there is a channel between Alice and Bob which determines the transition probability $q(y|x)$ that on Bob's side of the channel it is the message $y \in \mathcal{Y}$ that he receives when Alice has sent x. Knowing both probabilities $p(x)$ and $q(y|x)$ Bob can infer, by use of Bayes' rule (1.89), the aposteriori probability $p(x|y)$ which represents the probability that x has been sent given that y is received. This is a valuable function, because it reflects Bob's *uncertainty* about the message that Alice has sent. (It does not reflect his *ignorance*, because his ignorance is already connected to the apriori probability $p(x)$.) Since $p(x|y)$ is, for any fixed y, a probability distribution with respect to x, it is straightforward to consider the function

$$H(X|y) := -\sum_x p(x|y) \log p(x|y), \tag{1.103}$$

which is just the entropy of the ensemble $X|y$ of send events, which are *conditional* on the fact that y is received. (One might think of calling it the "conditional entropy" but this notion is reserved for another expression given below.) The function $H(X|y)$ can be interpreted as Bob's *specific uncertainty* about the original source message when he receives y. For example, in case of a noiseless channel we have $q(y|x) = \delta(x, y)$, which implies $p(x|y) = \delta(x, y)$, so the specific uncertainty vanishes. The average of the specific uncertainty over the ensemble Y of receive events is then given by the function

$$H(X|Y) := \sum_{x,y} q(y) H(X|y), \tag{1.104}$$

which is called the *conditional entropy* or *equivocation*. Because for each y the entropy $H(X|y)$ is non-negative, the conditional entropy is also non-negative,

$$H(X|Y) \geq 0. \tag{1.105}$$

Using the definition of the joint probability (1.84) we find that

$$H(X|Y) = \sum_{x,y} q(y) H(X|y) \tag{1.106}$$

$$= -\sum_{x,y} q(y) p(x|y) \log_2 p(x|y) \tag{1.107}$$

$$= -\sum_{x,y} p(x,y) \log_2 p(x|y), \tag{1.108}$$

which is nothing but the average of the random variable $f(X,Y) = -\log_2 p(x|y)$,

$$H(X|Y) = \langle -\log_2 p(x|y) \rangle. \tag{1.109}$$

The conditional entropy $H(X|Y)$ represents Bob's *average uncertainty* about the message that Alice has sent. (Not another peculiar combination of words: "expected uncertainty". Let us stick to "average", that sounds better and means the same.) For a lossless channel we have $H(X|Y) = 0$, so there is no uncertainty for Bob. Following this interpretation of the conditional entropy it is straightforward to consider the expression

$$H(X\!:\!Y) := H(X) - H(X|Y), \tag{1.110}$$

which is called the *mutual information*. It represents Bob's *apriori ignorance* reduced by his *average uncertainty* about Alice's message. This difference yields Bob's *information gain* about X when he learns the value of Y. We have

$$H(X) \geq H(X\!:\!Y) \geq 0. \tag{1.111}$$

In case of a lossless channel the mutual information is maximized to $H(X\!:\!Y) = H(X)$, so the information contained in X is not reduced after having passed the channel. Simple calculations show that

$$H(X\!:\!Y) = H(Y) - H(Y|X), \tag{1.112}$$

and also

$$H(X\!:\!Y) = H(X) + H(Y) - H(X,Y), \tag{1.113}$$

where

$$H(X,Y) := -\sum_{x,y} p(x,y) \log_2 p(x,y) \tag{1.114}$$

is the *joint entropy* of X and Y. This entropy can be interpreted as the *overall ignorance* about what happens at all in the transmission. As one can see, the mutual information is symmetric,

$$H(X\!:\!Y) = H(Y\!:\!X), \tag{1.115}$$

which reflects the fact that one can learn as much about X by learning Y as one can learn about Y by learning X.

Lastly we should mention the *relative entropy* of two ensembles X and Y over the same alphabet,

$$H(X\|Y) := \sum_{x} p(x) \log_2 \frac{p(x)}{q(x)}. \tag{1.116}$$

As a functional of the two probability distributions p and q is is also known as the *Kullback-Leibler distance*,

$$D(p\|q) \equiv H(X\|Y). \tag{1.117}$$

However, $D(p\|q)$ is not a distance in the strict sense, because in general it is not symmetric and does not fulfill the triangle relation. The non-negativity of the relative entropy is known as the *Gibbs inequality*,

$$\begin{aligned} &D(p\|q) \geq 0 \\ \text{resp.} \quad &H(X\|Y) \geq 0. \end{aligned} \tag{1.118}$$

The Gibbs inequality implies that

$$\sum_x p(x) \log_2 q(x) \leq \sum_x p(x) \log_2 p(x), \tag{1.119}$$

which is a very useful relation.

1.7.3 Channel capacity

Say we have a noisy channel. In order to avoid decoding errors, we need to add some *redundancy* to the messages, so that the original message can still be decoded if there are some badly transmitted symbols. The aim of such *error correcting codes* is opposed to the aim of a compressing code, namely *expanding* the length of the codewords so that an error here and there does not affect the decoding fidelity. More precisely, an error correcting code is a block code $c : \mathcal{A} \to \mathcal{A}_c^n$ which maps each symbol from the source alphabet $\mathcal{A}$ to a block of n symbols from a code alphabet $\mathcal{A}_c$. The *rate* of the code is defined by

$$R := \frac{\log_2 |\mathcal{A}|}{n}, \tag{1.120}$$

which represents the number of data bits per code symbol. The codebook $\mathcal{C} = c(\mathcal{A}) \subset \mathcal{A}_c^n$ consists of $2^{nR} = |\mathcal{A}|$ codewords, one for each source symbol. An error correcting code needs a clever decoding function g which is in this case not simply given by c^{-1}. Instead, the decoding function $g : \mathcal{A}_c^n \to \mathcal{A}$ corrects errors by mapping messages in the neighbourhood of $c(x)$ back to x. This "neighbourhood" is defined by the *Hamming distance*, i.e. the number of different digits. Now given a noisy channel, the task is to find an error correcting code with an upper bound for the rate R such that all messages still can be send in an asymptotically faithful manner, i.e. for $n \to \infty$. Shannon's great achievement was to find an *explicit* and *non-vanishing* upper bound for the rate of the code. This bound is given by the *channel capacity*

$$C := \max_{p(x)} H(X\!:\!Y), \tag{1.121}$$

where the maximization is performed over all possible apriori probabilities $p(x)$ over the source alphabet $\mathcal{A}$.

Theorem 4 (Channel coding theorem) *An asymptotically faithful error correcing code c has a rate R which is bounded from above by*

$$R \leq C, \tag{1.122}$$

where C is the channel capacity. Conversely, to every $R \leq C$ there is an asymptotically faithful error correcting code with this rate.

In practice it is not so easy to actually *calculate* the channel capacity. However, the case of a noiseless channel is very simple: since $H(X:Y) = H(X)$ and since $H(X)$ can be maximized to $\log_2 |\mathcal{A}|$ by a uniform distribution, the capacity of a noiseless channel is $C = \log_2 |\mathcal{A}|$. Therefore we have $R \leq \log_2 |\mathcal{A}|$ and thus $n \geq 1$, which means that there is a faithful error correcting code with $n = 1$, i.e. without any redundancy.
Since it is not our task to consider noisy channels and error correcting codes, we will not further go into detail here. The reader may find detailed a discussion in [9, 12, 19, 20] and other related literature.

Chapter 2

Quantum Information

Assume that the system that Alice uses to store her message is a *quantum system*. A quantum system has much more possibilities than a classical system. It cannot only be in one out of several possible states, but also in any *superposition* of these states. If the system is measured, the superposition is in general destroyed. This peculiar behaviour of quantum systems has a deep impact on the processing of information, and so the notion of "quantum information" has been coined. To put it short: Classical information is the information that is stored by manipulating a classical system, and quantum information is the information that is stored by manipulating a quantum system. Quantum information is a generalized form of information: By choosing mutually orthogonal states both for encoding and decoding, and by addressing each qubit individually, the quantum information stored in the system behaves classical. If the states are non-orthogonal and/or the qubits are not all addressed individually, then the information behaves non-classical. Thus the distinguishing features between classical information and quantum information are *superposition* and *entanglement*.

2.1 States

2.1.1 Classical states

Classically, the *state* of a system is represented by a probability density ρ on a phase space $\mathcal{P}$. Let the phase space be endowed with the phase space measure μ,

$$\mu(\mathcal{A}) = \int_{\mathcal{A}} d\mu(x), \tag{2.1}$$

for all measurable subsets $\mathcal{A} \subset \mathcal{P}$. For example, the phase space of N particles is the space $\mathcal{P} = \mathbb{R}^{6N}$ of the phase coordinates $x = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N; \boldsymbol{p}_1 \ldots \boldsymbol{p}_N)$. The phase space measure is then given by $d\mu(x) = d^3x_1 \cdots d^3x_N d^3p_1 \cdots d^3p_N$. A probability density ρ is a function on the phase space with the following properties

1. ρ is real-valued,
$$\rho^* = \rho. \tag{2.2}$$

2. ρ is non-negative,
$$\rho \geq 0. \tag{2.3}$$

3. ρ is normalized to unity,
$$\int d\mu(x)\,\rho(x) = 1, \tag{2.4}$$
where we understand $\int d\mu(x) \equiv \int_{\mathcal{P}} d\mu(x)$.

The set of all probability densities ρ on $\mathcal{P}$ constitutes the *state space* $\mathcal{S}(\mathcal{P})$. Quite general, the state of a system is determined by its *properties*. Let us distinguish between *properties* and *property values*. For example, "Color" is a property, whereas "red, green, blue" are corresponding property values. Observable properties of the system are represented by *observables*, which are given by real-valued functions on the phase space. In our terminology the property values of an observable $A : \mathcal{P} \to \mathbb{R}$ are the function values $a \in \mathcal{A}$, where
$$\mathcal{A} = \{A(x) \mid x \in \mathcal{P}\} \tag{2.5}$$
is the range of A, thus we relate
$$\text{Function } A \quad \hat{=} \quad \text{Observable property} \tag{2.6}$$
$$\text{Function value } a \quad \hat{=} \quad \text{Property value.} \tag{2.7}$$

The *expectation value* of an observable A is defined as
$$\langle A \rangle := \int d\mu(x)\,\rho(x) A(x), \tag{2.8}$$
and the *uncertainty* of A as
$$\Delta A := \sqrt{\langle A^2 \rangle - \langle A \rangle^2}. \tag{2.9}$$

Time is introduced into the theory through the parametrization of states and observables by an external parameter $t \in \mathbb{R}$, thus $\rho \mapsto \rho(t)$ and $A \mapsto A(t)$. The expectation value of an observable A at time t is therefore given by
$$\langle A \rangle(t) = \int d\mu(x)\,\rho(x;t) A(x;t). \tag{2.10}$$

While the time-dependence of observables is *externally* determined, the time evolution of the system state is governed by the *Liouville equation*
$$\frac{\partial}{\partial t}\rho(t) = \{H(t), \rho(t)\}, \tag{2.11}$$
where $H(t)$ is the *Hamiltonian*, i.e. the total energy of the system, and where $\{\cdot,\cdot\}$ is the *Poisson bracket*, defined by
$$\{A, B\} := \sum_{i=1}^{d/2} \frac{\partial A}{\partial x_i}\frac{\partial B}{\partial p_i} - \frac{\partial B}{\partial x_i}\frac{\partial A}{\partial p_i}. \tag{2.12}$$
with $d = \dim \mathcal{P}$. The *initial state* is given by some $\rho_0 \in \mathcal{S}(\mathcal{P})$ so that $\rho(t)$ has to fulfill the initial condition
$$\rho(t_0) = \rho_0. \tag{2.13}$$

A system is said to be in a *microstate* if the probability density is a δ-function, i.e.

$$\rho(x;t) = \delta(x - x(t)), \tag{2.14}$$

for some *trajectory* $x(t) \in \mathcal{P}$, otherwise the system is in a *macrostate*. For a microstate the point $x(t)$ comprises the exact phase space coordinates of the system at time t. The expectation value of an observable A at time t then reads

$$\langle A\rangle(t) = A(x(t);t), \tag{2.15}$$

and the uncertainty is vanishing,

$$\Delta A(t) = 0. \tag{2.16}$$

If the system is at some time t in a microstate then its state is completely specified by the position and momentum of any particle which is part of the system. In this case any observable has at any time a unique value without any uncertainty, and the Liouville equation (2.11) guarantees that a microstate always evolves into a microstate. Therefore the property value corresponding to the observable A is at any time t precisely given by

$$a(t) = \langle A\rangle(t) = A(x(t);t). \tag{2.17}$$

The *entropy* of the system is defined as

$$S(\hat{\rho}) := -\int d\mu(x)\,\rho(x;t)\log_2\rho(x;t). \tag{2.18}$$

Because S is not a function on phase space, it does not represent an observable but rather a *feature of the system state*. For example, if only some macroscopic quantities like particle number, temperature and pressure are known, then the system is in a macrostate which is compatible with the value of these quantities and which at the same time maximizes the entropy.

The structure of the classical state concept allows the following *realistic interpretation*: Any system is *actually* in a microstate at any time, and only the observer's *incomplete knowledge* is responsible for the statistical description by macrostates. The values of all observables represent *objective* properties of the system, because for microstates there is no uncertainty throughout the course of time. The entropy can be regarded as a measure for the observer's *ignorance* about the system. Since the entropy of a microstate is negative infinity, this indicates that the observer needs an infinite amount of resources to gain complete knowledge about the system. Thus for any finite information-processing observer system (and thus for any human being) the system *appears* to be in a macrostate, although it *actually* is in a microstate.

2.1.2 Quantum states

The classical state concept can straightforwardly be transferred to the quantum case. The state of a quantum system is represented by a *density matrix* $\hat{\rho}$ on a Hilbert space $\mathcal{H}$. For example, the Hilbert space of N spin-1/2 particles is the space $\mathcal{H} = L^2(\mathbb{R}^{3N})\otimes\mathbb{C}^{2N}$. A density matrix is an operator $\hat{\rho}$ everywhere defined on $\mathcal{H}$ which fulfills the following conditions:

1. $\hat{\rho}$ is self-adjoint, i.e.
$$\hat{\rho}^\dagger = \hat{\rho}, \tag{2.19}$$
which is a short notation for $\langle\psi|\hat{\rho}|\psi\rangle^* = \langle\psi|\hat{\rho}|\psi\rangle$ for all $|\psi\rangle \in \mathcal{H}$.

2. $\hat{\rho}$ is non-negative, i.e.
$$\hat{\rho} \geq 0, \tag{2.20}$$
which is a short notation for $\langle\psi|\hat{\rho}|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}$.

3. $\hat{\rho}$ has unit trace, i.e.
$$\mathrm{Tr}\{\hat{\rho}\} = 1, \tag{2.21}$$
where the *trace* of an operator $\hat{A}$ is defined through
$$\mathrm{Tr}\{\hat{A}\} := \sum_{n=1}^{d} \langle e_n|\hat{A}|e_n\rangle, \tag{2.22}$$
with $\{|e_1\rangle, \ldots, |e_d\rangle\}$ being an arbitrary orthonormal basis of $\mathcal{H}$, and $d = \dim\mathcal{H}$.

For Hilbert spaces of finite dimension d the density matrix
$$\hat{\rho} = \frac{1}{d}\mathbb{1} \tag{2.23}$$
is called the *complete mixture*. For infinite-dimensional Hilbert spaces there is no complete mixture. In this case the *Gaussian states* form the class of states which are closest to a complete mixture, because they maximize the entropy for fixed first and second moments. A discussion of Gaussian states, however, would go beyond the scope of this thesis. The set of all density matrices on $\mathcal{H}$ constitutes the *state space* $\mathcal{S}(\mathcal{H})$. Note, however, that the state space is not a vector space, in contrast to $\mathcal{H}$. Any density matrix has a *spectral decomposition* of the form
$$\hat{\rho} = \sum_k \lambda_k |u_k\rangle\langle u_k|, \tag{2.24}$$
where the eigenvalues have the properties of a probability distribution, i.e. $\lambda_k \geq 0$, $\sum_k \lambda_k = 1$. An observable is represented by a self-adjoint operator on $\mathcal{H}$, and the eigenvalues of the operator represent the corresponding property values. If a is an eigenvalue of the self-adjoint operator $\hat{A}$, i.e. $\hat{A}|u_a\rangle = a|u_a\rangle$ for some eigenvector $|u_a\rangle \in \mathcal{H}$, then we relate
$$\text{Operator } \hat{A} \quad \hat{=} \quad \text{Observable property} \tag{2.25}$$
$$\text{Eigenvalue } a \quad \hat{=} \quad \text{Property value.} \tag{2.26}$$
The *expectation value* of an observable $\hat{A}$ is defined as
$$\langle\hat{A}\rangle := \mathrm{Tr}\{\hat{\rho}\hat{A}\}, \tag{2.27}$$
and the *uncertainty* of $\hat{A}$ is defined as
$$\Delta A := \sqrt{\langle\hat{A}^2\rangle - \langle\hat{A}\rangle^2}. \tag{2.28}$$

Time is introduced through the parametrization of states and observables by an external parameter $t \in \mathbb{R}$, thus $\hat{\rho} \mapsto \hat{\rho}(t)$ and $\hat{A} \mapsto \hat{A}(t)$. The expectation value of an observable $\hat{A}$ at time t is therefore given by

$$\langle \hat{A} \rangle(t) = \mathrm{Tr}\{\hat{\rho}(t)\hat{A}(t)\}. \tag{2.29}$$

Throughout this thesis we are working in the Schrödinger picture. In this picture, the time-dependence of observables is *externally given*, and the time evolution of the system state is governed by the *von-Neumann equation*

$$\frac{\partial}{\partial t}\hat{\rho}(t) = \frac{1}{i\hbar}[\hat{H}(t), \hat{\rho}(t)], \tag{2.30}$$

where $\hat{H}(t)$ is the Hamiltonian, and where $[\cdot, \cdot]$ is the *commutator*, defined through

$$[\hat{A}, \hat{B}] := \hat{A}\hat{B} - \hat{B}\hat{A}. \tag{2.31}$$

The *initial state* is given by some $\hat{\rho}_0 \in \mathcal{S}(\mathcal{H})$, so that $\hat{\rho}(t)$ has to fulfill the initial condition

$$\hat{\rho}(t_0) = \hat{\rho}_0. \tag{2.32}$$

The time-dependent system state can be expressed by

$$\hat{\rho}(t) = \hat{U}(t, t_0)\hat{\rho}_0\hat{U}^\dagger(t, t_0), \tag{2.33}$$

where the unitary *time evolution operator* $\hat{U}(t, t')$ obeys the *Schrödinger equation*

$$i\hbar\frac{\partial}{\partial t}\hat{U}(t, t') = \hat{H}(t)\hat{U}(t, t'), \tag{2.34}$$

together with the condition

$$\hat{U}(t, t) = \mathbb{1}. \tag{2.35}$$

Entanglement

If one considers a system which is composed out of several subsystems, then the Hilbert space of the total system is the *tensor product* of the individual Hilbert spaces,

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots . \tag{2.36}$$

Let us restrict here to *bipartite systems*, i.e. systems which are composed out of two subsystems,

$$\mathcal{H} = \mathcal{A} \otimes \mathcal{B}. \tag{2.37}$$

The *convex sum* of products of density matrices, i.e. the operator

$$\hat{\rho} = \sum_k \lambda_k \, (\hat{\sigma}_k \otimes \hat{\omega}_k), \tag{2.38}$$

where $\hat{\sigma}_k \in \mathcal{S}(\mathcal{A})$, $\hat{\omega}_k \in \mathcal{S}(\mathcal{B})$ and where $\lambda_k \geq 0$, $\sum_k \lambda_k = 1$, is called a *separable state*. Any non-separable state is called an *entangled state*. Separable states correspond to systems which are *classically correlated*, while entangled states contain *quantum* *correlations*. What is meant by this? If $\hat{\rho}$ is a separable state then the expectation value of some observable $\hat{A}$ on $\mathcal{S}(\mathcal{A}\otimes\mathcal{B})$ reads

$$\begin{aligned}
\langle\hat{A}\rangle &= \mathrm{Tr}\{\hat{\rho}\hat{A}\} && (2.39)\\
&= \mathrm{Tr}\Big\{\sum_k \lambda_k(\hat{\sigma}_k\otimes\hat{\omega}_k)\hat{A}\Big\} && (2.40)\\
&= \mathrm{Tr}_{\mathcal{A}}\{\hat{\rho}_{\mathcal{A}}\hat{A}_{\mathcal{A}}\}\mathrm{Tr}_{\mathcal{B}}\{\rho_{\mathcal{B}}\hat{A}_{\mathcal{B}}\} && (2.41)\\
&= \langle\hat{A}_{\mathcal{A}}\rangle\langle\hat{A}_{\mathcal{B}}\rangle, && (2.42)
\end{aligned}$$

where we have defined the reduced observables

$$\hat{A}_{\mathcal{A}} := \mathrm{Tr}_{\mathcal{B}}\{\hat{A}\}, \quad \hat{A}_{\mathcal{B}} := \mathrm{Tr}_{\mathcal{A}}\{\hat{A}\}, \tag{2.43}$$

and the density matrices

$$\hat{\rho}_{\mathcal{A}} := \sum_k \lambda_k\hat{\sigma}_k, \quad \hat{\rho}_{\mathcal{B}} := \sum_k \lambda_k\hat{\omega}_k. \tag{2.44}$$

Thus for separable states the observable $\hat{A}$ cannot be distinguished from the observable

$$\hat{A}' = \hat{A}_{\mathcal{A}}\otimes\hat{A}_{\mathcal{B}}. \tag{2.45}$$

Observables which are of the above product form are called *local observables*. As we can see by (2.42), for separable states the expectation value of any observable is just a product of local expectation values. This is a classical type of correlation, and it can be simulated with a classical ensemble. However, if the state $\hat{\rho}$ is entangled, then the expectation values can in general *not* be written in product form. Such a situation is *non-classical* in that it cannot be simulated by a classical ensemble. An entangled state contains information which is distributed over the subsystems, in other words the total system shows *non-local* properties. This motivates the introduction of the term *non-locality* in order to describe this typically quantum phenomenon.

Pure states

The system is said to be in a *pure state* if the density matrix is a projector, i.e.

$$\hat{\rho}^2 = \hat{\rho}, \tag{2.46}$$

otherwise the system is in a *mixed state*. Pure states are the extremal points in the state space $\mathcal{S}(\mathcal{H})$, because for any $\hat{\rho}\in\mathcal{S}(\mathcal{H})$ we have in general

$$\mathrm{Tr}\{\hat{\rho}^2\} \leq 1, \tag{2.47}$$

while for pure states we have $\mathrm{Tr}\{\hat{\rho}^2\} = 1$. A pure state has the form

$$\hat{\rho} = |\psi\rangle\langle\psi| \tag{2.48}$$

for some $|\psi\rangle \in \mathcal{H}$ with unit norm. If only pure states are considered, it suffices to identify the state of the system with a *unit ray* in $\mathcal{H}$, i.e.

$$|\psi\rangle \mathrel{\hat{=}} \{e^{i\phi}|\psi\rangle \mid \phi\in\mathbb{R}\}. \tag{2.49}$$

For pure states the expectation value of an observable $\hat{A}$ reads

$$\langle \hat{A} \rangle = \mathrm{Tr}\{|\psi\rangle\langle\psi|\hat{A}\} = \langle\psi|\hat{A}|\psi\rangle. \tag{2.50}$$

To any density matrix $\hat{\rho} \in \mathcal{S}(\mathcal{H})$ there is a pure state $|\Psi\rangle$ in a larger Hilbert space $\mathcal{H}\otimes\mathcal{K}$ such that

$$\hat{\rho} = \mathrm{Tr}_{\mathcal{K}}\{|\Psi\rangle\langle\Psi|\}, \tag{2.51}$$

where $\mathrm{Tr}_{\mathcal{K}}\{\cdot\}$ is the *partial trace* over the factor space $\mathcal{K}$. The state $|\Psi\rangle$ is called a *purification* of $\hat{\rho}$. Note, however, that the purification is not unique, i.e. there a many pure states in a larger Hilbert space whose partial trace yields the density matrix $\hat{\rho}$. A pure separable state is a *product state*,

$$|\Psi\rangle = |\psi_{\mathcal{H}}\rangle \otimes |\psi_{\mathcal{K}}\rangle, \tag{2.52}$$

while an entangled pure state cannot be written in product form, but rather in the general form

$$|\Psi\rangle = \sum_{k,l} \Psi(k,l)|u_k\rangle \otimes |e_l\rangle, \tag{2.53}$$

where $\{|u_k\rangle\}$ and $\{|e_l\rangle\}$ are orthonormal bases for $\mathcal{H}$ and $\mathcal{K}$, respectively.

Interpretation of quantum mechanics

Even for pure states the uncertainty does not vanish for all observables. The *Heisenberg uncertainty relation* [21] states that for any two observables $\hat{A}$ and $\hat{B}$ their uncertainties obey

$$\Delta A \Delta B \geq \frac{1}{2}|\langle[\hat{A},\hat{B}]\rangle|. \tag{2.54}$$

If $[\hat{A},\hat{B}] \neq 0$ then the two observables are called *incompatible*. In particular for position and momentum we find that

$$[\hat{x},\hat{p}] = i\hbar, \tag{2.55}$$

therefore the position and the momentum of a particle are not compatible. The uncertainty principle has a major implication on the interpretation of quantum mechanics. Because it is impossible to assign a precise value for two incompatible observables at one instant in time they cannot both correspond to elements of reality. This rigorous conclusion has firstly been drawn by *Einstein, Podolsky* and *Rosen* in 1935. In their landmark paper [22] the authors claim that quantum mechanics cannot be considered a complete theory, i.e. a theory where every element of the physical reality has a counterpart in the physical theory. Since then a neverending debate is is going on whether or not there is a realistic interpretation of quantum mechanics in the same way as there is one for classical mechanics. We will not enter this debate, but I favorize the following approach which offers in my view a satisfying solution to the controversies. In this approach, the state of a system is merely a *catalog of the observer's knowledge about the system*. A measure for the observer's *ignorance* about the system is given by the *von-Neumann entropy* of the system is defined as

$$S(\hat{\rho}) := -\mathrm{Tr}\{\hat{\rho}\log_2\hat{\rho}\}. \tag{2.56}$$

Just like in the classical theory, the entropy does not represent an observable but rather a *feature of the system state*, because S is not a self-adjoint operator on $\mathcal{H}$. The entropy is always non-negative,

$$S(\hat{\rho}) \geq 0, \tag{2.57}$$

and for pure states it becomes zero, which means that a pure state represents *complete knowledge* about the system. The mentioned "observer" does not need to be a human being, in fact it does not matter who or what the observer is. It is only important that the observer is a system which remains *outside* the description. The observer is the *subject* and the described system is the *object*. *Actual reality* is the result of interactions between subject and object. Any such interaction process represents an *observation* or *measurement* and the results of these measurements define discrete events happening in space and time. The physical theory then has to provide the probability that a given pair of measurement events follow each other.

2.2 The Qubit

The elementary unit of classical information is the *bit*. The bit is the smallest possible system that is able to store information: A system which can be in one of two states. Conventionally, these two states are labelled by `0` and `1`. Such a system acts as a *classical channel*. Alice, the sender, encodes her message ("yes/no" or "black/white" or "good/bad" or the like) by manipulating the state of the system and Bob, the receiver, reads out the state of the system and decodes it back to the original message. (Bob has to know the code that Alice uses, e.g. $c(\texttt{yes}) = \texttt{1}$ and $c(\texttt{no}) = \texttt{0}$.) If Alice wants to transmit more sophisticated messages then she has to use a channel system which can be in more than two states. If she has a couple of two-state systems at hand, then she can *combine* them to a larger system. As we have introduced by definition (1.12), the number of binary systems that Alice needs to encode her message is a measure for the information content of the message.

The quantum analogue to a classical bit is a two-level quantum system, and it is called the *quantum bit* or in short the *qubit*. Classical information is "quantized" by applying the map

$$\texttt{0} \mapsto |0\rangle, \quad \texttt{1} \mapsto |1\rangle, \tag{2.58}$$

where these two vectors have unit norm and are mututally orthogonal,

$$\langle 0|0\rangle = \langle 1|1\rangle = 1, \quad \langle 0|1\rangle = 0. \tag{2.59}$$

The set $\mathcal{B} = \{|0\rangle, |1\rangle\}$ is called the *computational basis*. The qubit state is a unit vector in the Hilbert space $\mathcal{H}$ spanned by the computational basis,

$$\mathcal{H} = \mathrm{Span}\{|0\rangle, |1\rangle\}, \tag{2.60}$$

and is of the form

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \tag{2.61}$$

where α, β are complex numbers fulfilling $|\alpha|^2 + |\beta|^2 = 1$.

A qubit can be implemented in many ways. Very common implementations are the *spin-$1/2$ particle* and the *polarized photon*. In case of a spin-1/2 particle the computational basis states are identified by the "spin-up" and "spin-down" states of the particle,

$$|0\rangle \equiv |\uparrow\rangle, \quad |1\rangle \equiv |\downarrow\rangle, \tag{2.62}$$

which represent the eigenstates of the spin component in a fixed direction. In case of a polarized photon the basis states are identified with the horizontal and vertical polarization states, respectively,

$$|0\rangle \equiv |H\rangle, \quad |1\rangle \equiv |V\rangle. \tag{2.63}$$

A convenient mathematical representation of qubits is the *tupel representation*

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{2.64}$$

such that the qubit statecom (2.61) can be written as

$$|\psi\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \tag{2.65}$$

However, the tupel representation is *basis dependent* in contrast to the abstract ket representation.

2.2.1 The Pauli matrices

There is yet another and very elegant representation of the pure qubit state as a point on the *Bloch sphere*. Generally, each point on a 3-dimensional sphere is uniquely defined by the two angles φ, ϑ, where $\varphi \in [0, 2\pi]$ and $\vartheta \in [0, \pi]$. Now it is a matter of fact that we can write down any normalized pure qubit state in the following fashion:

$$|\psi\rangle = \cos\frac{\vartheta}{2} e^{-i\frac{\varphi}{2}}|0\rangle + \sin\frac{\vartheta}{2} e^{i\frac{\varphi}{2}}|1\rangle, \tag{2.66}$$

where $\varphi \in [0, 2\pi]$ and $\vartheta \in [0, \pi]$. Thus each qubit state uniquely corresponds to a point on a unit sphere which is called the *Bloch sphere*, and vice versa (see Fig. 2.1). Now a point on the unit sphere is a unit vector in the 3-dimensional space. Any rotation of this vector affects the two parameters ϑ and φ, so the qubit space is in fact a *representation space for 3-dimensional rotations*. The rotation group $\mathcal{R}$ is the group of linear transformations R on the 3-dimensional space $\mathcal{V} = \mathbb{R}^3$ with $R^T R = \mathbb{1}$ and with $\det R = 1$, i.e. is the special orthogonal group in 3 dimensions, and one writes $\mathcal{R} = SO(3)$. A representation of the rotation group by orthogonal matrices in an n-dimensional vector space would be denoted by $SO(n)$. If the representation consists of *unitary* matrices on an n-dimensional complex vector space then one writes $SU(n)$. The rotation group is a *3-dimensional Lie group*, which means that it can be parametrized by three continuous parameters $\theta_1, \theta_2, \theta_3$. Considering these three parameters as components of a 3-vector $\boldsymbol{\theta}$, the parametrization has the form

$$R(\boldsymbol{\theta}) = e^{i\boldsymbol{\theta}\cdot\boldsymbol{J}}. \tag{2.67}$$

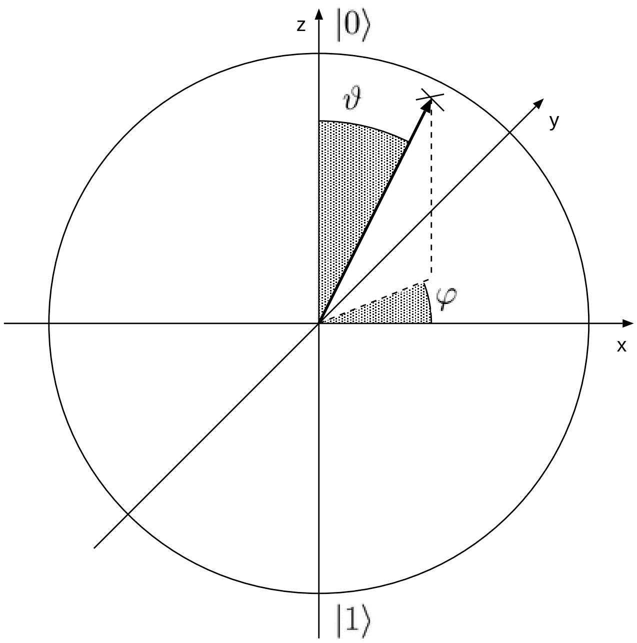


Figure 2.1: The Bloch sphere. Each pure state of a qubit can be represented by a point on the Bloch sphere.

An *infinitesimal rotation* about the angle $d\boldsymbol{\theta}$ is represented by the operator

$$R(d\boldsymbol{\theta}) = \mathbb{1} + i d\boldsymbol{\theta} \cdot \boldsymbol{J}. \tag{2.68}$$

The angular vector $\boldsymbol{\theta}$ can be considered as the rotation about a unit vector $\boldsymbol{n}$ by an angle θ with

$$\boldsymbol{\theta} = \theta \cdot \boldsymbol{n}. \tag{2.69}$$

The components of the vector operator $\boldsymbol{J}$ are called the *generators* of the group, and they are defined by

$$J_k := \frac{1}{i} \frac{dR}{d\theta_k}\Big|_{\boldsymbol{\theta}=0}. \tag{2.70}$$

The generators of *any* representation of the rotation group fulfill the commutator relation

$$[J_i, J_j] = i\epsilon_{ijk} J_k, \tag{2.71}$$

where $[A, B] := AB - BA$ and where we follow the Einstein convention and perform the sum over equal indices. As a consequence of (2.71), the generators form a *Lie algebra*, i.e. a vector space of operators(!) whose commutator is still in the vector space. In fact, one can construct a vector space $\mathcal{W}$ out of the basis $\mathcal{B} = \{J_1, J_2, J_3\}$ and then relation (2.71) guarantees that the commutator of two elements of $\mathcal{W}$ is still in $\mathcal{W}$. (This vector space $\mathcal{W}$ of operators has *nothing* to do with the representation space $\mathcal{V}$!) The eigenvalues m of any of the components of $\boldsymbol{J}$, e.g. J_3, fulfill $m = -j, \ldots, j$ where the number j is called the *spin* and is determined by the dimension n of the representation space through

$$2j + 1 = n. \tag{2.72}$$

For example, the spin-1 representation is the usual 3-dimensional representation $SO(3)$ in the vector space $\mathcal{V} = \mathbb{R}^3$. It can generally be shown that the spin j is either integer or half-integer.
The case that we are particularly interested in is the spin-$\frac{1}{2}$ representation $SU(2)$ of the rotation group, because its representation space is a 2-dimensional complex Hilbert space: the space of the *qubit*. The generators of the representation are given by the components of the vector operator

$$\hat{\boldsymbol{J}} = \frac{1}{\hbar}\hat{\boldsymbol{S}} = \frac{1}{2}\hat{\boldsymbol{\sigma}}, \tag{2.73}$$

where the operators

$$\hat{\sigma}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{\sigma}_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \hat{\sigma}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.74}$$

are called the *Pauli matrices*. The rotation matrix $R(\boldsymbol{\theta}) \in SO(3)$ is represented by the unitary operator $\hat{U}(\boldsymbol{\theta}) \in SU(2)$, where

$$\hat{U}(\boldsymbol{\theta}) = e^{\frac{i}{2}\boldsymbol{\theta}\cdot\hat{\boldsymbol{\sigma}}}. \tag{2.75}$$

However, there is a peculiar property of these representations: While the rotation about 2π around any axis $\boldsymbol{n}$ yields the identity operation, $R(2\pi\boldsymbol{n}) = \mathbb{1}$, this is not so for the spin-1/2 representation. Instead we find that $\hat{U}(2\pi\boldsymbol{n}) = -\mathbb{1}$. The $SU(2)$ is a representation of $SO(3)$ for *infinitesimal rotations*. If the angle $\theta = |\boldsymbol{\theta}|$ is too large then both representations do not coincide. In the $SU(2)$ it is the angle $\theta = 4\pi$ which leads to identity, $\hat{U}(4\pi\boldsymbol{n}) = \mathbb{1}$, so the domain of the parametrization of $SU(2)$ includes twice the domain of the parametrization of $SO(3)$. In plain words: You have to turn an electron *twice* around itself to reobtain its original state!
Now fix a normalized axis $\boldsymbol{n}$ and call it the *quantization axis*, then the operator

$$\hat{\sigma}_{\boldsymbol{n}} := \boldsymbol{n} \cdot \hat{\boldsymbol{\sigma}} \tag{2.76}$$

represents the spin component along this axis. The eigenvalues of any such operator are $+1$ or -1, so no matter what quantization axis we choose we will always find that the spin points either *up* or *down*, i.e. along or against the chosen axis. The corresponding eigenvectors are denoted by $|\uparrow_{\boldsymbol{n}}\rangle$ and $|\downarrow_{\boldsymbol{n}}\rangle$, so that

$$\hat{\sigma}_{\boldsymbol{n}}|\uparrow_{\boldsymbol{n}}\rangle = +|\uparrow_{\boldsymbol{n}}\rangle \tag{2.77}$$

$$\hat{\sigma}_{\boldsymbol{n}}|\downarrow_{\boldsymbol{n}}\rangle = -|\uparrow_{\boldsymbol{n}}\rangle. \tag{2.78}$$

(Note that the spin operator is given by $\hat{\boldsymbol{S}} = \frac{\hbar}{2}\hat{\boldsymbol{\sigma}}$, so that the eigenvalues of the physical spin component $\hat{S}_{\boldsymbol{n}} = \frac{\hbar}{2}\hat{\sigma}_{\boldsymbol{n}}$ are $\pm\frac{\hbar}{2}$.) For any two normalized vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ the corresponding spin components fulfill the *anticommutator relation*

$$\{\hat{\sigma}_{\boldsymbol{a}}, \hat{\sigma}_{\boldsymbol{b}}\} = 2\,\boldsymbol{a}\cdot\boldsymbol{b}, \tag{2.79}$$

where $\{\hat{A}, \hat{B}\} := \hat{A}\hat{B} + \hat{B}\hat{A}$ is the anticommutator of two operators $\hat{A}$ and $\hat{B}$. In particular, the Pauli operators $\hat{\sigma}_i$ obey

$$\{\hat{\sigma}_i, \hat{\sigma}_l\} = 2\delta_{ij}. \tag{2.80}$$

The expectation value of the spin component $\hat{\sigma}_{\boldsymbol{b}}$ in the state $|\uparrow_{\boldsymbol{a}}\rangle$ is given by

$$\langle\uparrow_{\boldsymbol{a}}|\hat{\sigma}_{\boldsymbol{b}}|\uparrow_{\boldsymbol{a}}\rangle = \boldsymbol{a}\cdot\boldsymbol{b}. \tag{2.81}$$

For a large number of electrons the above expectation value is proportional to the number of spin-polarized electrons passing a Stern-Gerlach apparatus directed along $\boldsymbol{b}$. This is also what one would expect *classically* when interpreting (2.81) as the (normalized) *intensity* of a beam of electrons passing the apparatus. Only when there are *very few* electrons, the true quantum nature of the electrons is revealed, because intensities appear as *probabilities* then. The conditional probability to find an electron with a spin $\uparrow_{\boldsymbol{b}}$ or $\downarrow_{\boldsymbol{b}}$ if it has been prepared with spin $\uparrow_{\boldsymbol{a}}$ is respectively given by

$$P(\uparrow_{\boldsymbol{b}} \mid \uparrow_{\boldsymbol{a}}) = |\langle\uparrow_{\boldsymbol{b}} \mid \uparrow_{\boldsymbol{a}}\rangle|^2 = \frac{1}{2}(1+\boldsymbol{a}\cdot\boldsymbol{b}) \tag{2.82}$$

$$P(\downarrow_{\boldsymbol{b}} \mid \uparrow_{\boldsymbol{a}}) = |\langle\downarrow_{\boldsymbol{b}} \mid \uparrow_{\boldsymbol{a}}\rangle|^2 = \frac{1}{2}(1-\boldsymbol{a}\cdot\boldsymbol{b}). \tag{2.83}$$

The Pauli matrices have some other fancy properties which allow for interesting operations. Since a qubit is a carrier of *quantum information*, these operations correspond to *calculations* and *gates*. Let us begin.

The Pauli matrices are *Hermitian*,

$$\hat{\sigma}_i^\dagger = \hat{\sigma}_i, \tag{2.84}$$

so they are *observables*. They are also *unitary*,

$$\hat{\sigma}_i^{-1} = \hat{\sigma}_i^\dagger, \tag{2.85}$$

so they are physically realizable *transformations*. Equations (2.84) and (2.85) imply that the Pauli matrices are *self-inverse*,

$$\hat{\sigma}_i^2 = \mathbb{1}. \tag{2.86}$$

They have vanishing trace,

$$\mathrm{Tr}\{\hat{\sigma}_i\} = 0, \tag{2.87}$$

and they obey

$$\mathrm{Tr}\{\hat{\sigma}_i\hat{\sigma}_j\} = 2\delta_{ij}. \tag{2.88}$$

From (2.71) and (2.73) we can infer the *commutator relation*

$$[\hat{\sigma}_i, \hat{\sigma}_j] = 2\epsilon_{ijk}\hat{\sigma}_k. \tag{2.89}$$

so the generators $\{\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3\}$ form the basis of a Lie algebra. Not only that, one can show that the set

$$\mathcal{B} = \{\mathbb{1}, \hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3\} \tag{2.90}$$

is a basis for the vector space containing *all* 2×2-matrices! This means, whenever we have an arbitrary 2×2-matrix $\hat{A}$ we can decompose it into a linear combination of Pauli matrices and unity operator,

$$\hat{A} = \sum_{\mu=0}^{3} \alpha_\mu \, \hat{\sigma}_\mu, \tag{2.91}$$

where $\hat{\sigma}_0 := \mathbb{1}$, and where the components are given by

$$\alpha_\mu = \frac{1}{2}\mathrm{Tr}\{\hat{\sigma}_\mu \hat{A}\}. \tag{2.92}$$

If we now consider a 2×2 density matrix $\hat{\rho}$ then it can be written as

$$\hat{\rho} = \frac{1}{2}(\mathbb{1} + \boldsymbol{\lambda} \cdot \hat{\boldsymbol{\sigma}}), \tag{2.93}$$

where $|\boldsymbol{\lambda}| \leq 1$. So there is a one-to-one correspondence between a given density matrix $\hat{\rho}$ and a point $\boldsymbol{\lambda}$ within a unit ball in three dimensions. On the surface of this ball, where $|\boldsymbol{\lambda}| = 1$, the density matrix is a one-dimensional projector and thus represents a pure state. The surface of a ball is a *sphere*, so we recover the *Bloch sphere* representation of pure states! Consequently, any qubit state $\hat{\rho}$ can be represented by a point in a three-dimensional unit ball which is called the *Bloch ball*. In the center of the ball, where $\boldsymbol{\lambda} = 0$, the qubit state is completely mixed, while on the surface, where $|\boldsymbol{\lambda}| = 1$, the qubit state is pure. Using (2.92) we can calculate the components of the *Bloch vector* $\boldsymbol{\lambda}$ for a given density matrix $\hat{\rho}$ by

$$\lambda_i = \mathrm{Tr}\{\hat{\sigma}_i \hat{\rho}\} = \langle \hat{\sigma}_i \rangle. \tag{2.94}$$

Now we have collected some important properties and features of qubit states and Pauli operators which are extensively used in quantum communication and computation.

2.3 Measurement

The most general operation that is allowed by the laws of quantum mechanics is a *completely positive map*. This is a map $\mathcal{E}$ which fulfills the following conditions:

1. $\mathcal{E}$ is linear, i.e.

$$\mathcal{E}(\sum_k \lambda_k \hat{\rho}_k) = \sum_k \lambda_k \mathcal{E}(\hat{\rho}_k), \tag{2.95}$$

   for any set of $\hat{\rho}_k \in \mathcal{S}(\mathcal{H})$ and $\lambda_k \geq 0$, $\sum_k \lambda_k = 1$.

2. $\mathcal{E}$ is positive, i.e.

$$\mathcal{E}(\hat{\rho}) \geq 0. \tag{2.96}$$

3. For any finite-dimensional Hilbert space $\mathcal{K}$ the map $\mathcal{E} \otimes \mathbb{1}_\mathcal{K}$ on $\mathcal{S}(\mathcal{H} \otimes \mathcal{K})$ is also positive.

We identify *quantum operations* with completely positive maps $\mathcal{E}$ that do not increase the trace,

$$\mathrm{Tr}\{\mathcal{E}(\hat{\rho})\} \leq 1. \tag{2.97}$$

Following a theorem by *Kraus* [23], any quantum operation $\mathcal{E}$ can be cast into the form

$$\mathcal{E}(\hat{\rho}) = \sum_{m \in \mathcal{M}} \hat{E}_m \hat{\rho} \hat{E}_m^\dagger, \tag{2.98}$$

where $\mathcal{M}$ is a discrete index set. The $\hat{E}_m$ are called *Kraus operators*. If $\mathcal{E}$ is also *trace-preserving*, i.e.

$$\mathrm{Tr}\{\mathcal{E}(\hat{\rho})\} = 1, \tag{2.99}$$

then the Kraus operators fulfill

$$\sum_{m \in \mathcal{M}} \hat{E}_m^\dagger \hat{E}_m = \mathbb{1}. \tag{2.100}$$

Trace-preserving quantum operations are all allowed operations that map quantum states to quantum states: unitary operations, non-selective measurements, addition of uncorrelated systems, the dismissal of parts of a compound system, and the replacement of the input state by some other state. In connection with trace-preserving quantum operations there is an important theorem by *Stinespring* which is not only of mathematical interest but also has a deep physical impact.

Theorem 5 (Stinespring Dilation Theorem) *Let $\mathcal{E}$ be a trace-preserving quantum operation on a Hilbert space $\mathcal{H}$. Then there is an ancilla space $\mathcal{K}$ of dimension $\dim \mathcal{K} \leq (\dim \mathcal{H})^2$ so that for any fixed $|\chi\rangle \in \mathcal{K}$ there is a unitary transformation $\hat{U}$ on $\mathcal{H} \otimes \mathcal{K}$ with*

$$\mathcal{E}(\hat{\rho}) = \mathrm{Tr}_{\mathcal{K}}\{\hat{U}(\hat{\rho} \otimes |\chi\rangle\langle\chi|)\hat{U}^\dagger\}. \tag{2.101}$$

This is an amazing theorem, because it shows that any allowed operation mapping quantum states to quantum states can be modeled by a unitary operation on a larger Hilbert space.

Another interesting fact about trace-preserving quantum operations is that they constitute a *positive operator-valued measure* or *POVM*. Let $\mathcal{E}$ be a trace-preserving quantum operation with Kraus operators $\hat{E}_m$ and define the operators

$$\hat{F}_m := \hat{E}_m^\dagger \hat{E}_m, \tag{2.102}$$

then the set

$$\mathcal{F} = \{\hat{F}_m\} \tag{2.103}$$

constitutes a POVM, because the $\hat{F}_m$ are Hermitian, non-negative and they fulfill

$$\sum_{m \in \mathcal{M}} \hat{F}_m = \mathbb{1}. \tag{2.104}$$

A POVM represents a *generalized measurement* where the outcome "m" occurs with the probability

$$P(m) = \mathrm{Tr}\{\hat{E}_m \hat{\rho} \hat{E}_m^\dagger\} = \mathrm{Tr}\{\hat{\rho} \hat{F}_m\}. \tag{2.105}$$

The trace-preserving property (2.100) ensures that the probabilities sum up to unity. A special case of a POVM is the *projection-valued measure* or in short *PVM*. Other synonymes for a PVM are the *von-Neumann measurement* and the *projective measurement*. In case of a PVM operation the Kraus operators are mutually orthogonal projectors, i.e.

$$\hat{E}_m^\dagger = \hat{E}_m = \hat{\Pi}_m, \quad \hat{\Pi}_m \hat{\Pi}_n = \delta_{mn} \hat{\Pi}_m, \tag{2.106}$$

so that the elements $\hat{F}_m$ are identical to the Kraus operators itself, i.e. the projectors

$$\hat{F}_m = \hat{E}_m = \hat{\Pi}_m. \tag{2.107}$$

A special feature of projective measurements is that they are *repeatable*,

$$\mathcal{E}^2(\hat{\rho}) = \sum_{m,n} \hat{\Pi}_m \hat{\Pi}_n \hat{\rho} \hat{\Pi}_n \hat{\Pi}_m \tag{2.108}$$

$$= \sum_{m \in \mathcal{M}} \hat{\Pi}_m \hat{\rho} \hat{\Pi}_m = \mathcal{E}(\hat{\rho}). \tag{2.109}$$

This indicates that many real-world measurements are *not* projective. For example, the detection of a photon on a silver screen destroys the photon, hence the measurement is not repeatable and thus cannot be represented by a PVM measurement. If the Kraus operators fulfill

$$\sum_m \hat{E}_m \hat{E}_m^\dagger = \mathbb{1}, \tag{2.110}$$

then $\mathcal{E}$ is called a *unital* operation. A quantum operation which is both trace-preserving and unital is called a *doubly stochastic map*. An important feature of doubly stochastic maps is that they increase the von-Neumann entropy,

$$S(\hat{\rho}) \leq S(\mathcal{E}(\hat{\rho})). \tag{2.111}$$

The doubly stochastic maps represent the class of *non-selective measurements*, which are measurements where the ensemble of individual output states is simply mixed together. It is clear that such a procedure cannot reduce the von-Neumann entropy, because the knowledge which is gained by the measurement is not used to manipulate the state in a selective manner. The post-measurement state of a non-selective measurement is given by

$$\hat{\rho}' = \mathcal{E}(\hat{\rho}) = \sum_{m \in \mathcal{M}} \hat{\rho}'_m, \tag{2.112}$$

with the individual output states

$$\hat{\rho}'_m = \hat{E}_m \rho \hat{E}_m^\dagger. \tag{2.113}$$

In a *selective measurement*, only a subset $\mathcal{N} \subset \mathcal{M}$ of results is selected and the other results are discarded, so the output state of a selective measurement has the form

$$\hat{\rho}' = \sum_{m \in \mathcal{N}} \hat{\rho}'_m = \sum_{m \in \mathcal{N}} \hat{E}_m \rho \hat{E}_m^\dagger, \tag{2.114}$$

where

$$\sum_{m\in\mathcal{N}} \hat{E}_m^\dagger \hat{E}_m \leq \mathbb{1}. \tag{2.115}$$

Generally, in a POVM measurement one does not care what happens with the state after the measurement. However, if one speaks of a "POVM operation" with elements $\{\hat{F}_m\}$ then one means that the post-measurement state is given by

$$\mathcal{E}(\hat{\rho}) = \sum_{m\in\mathcal{M}} \sqrt{\hat{F}_m}\hat{\rho}\sqrt{\hat{F}_m}, \tag{2.116}$$

so the Kraus operators are

$$\hat{E}_m = \sqrt{\hat{F}_m}. \tag{2.117}$$

Because the $\hat{F}_m$ are non-negative Hermitian operators, the above assignment is always well-defined. Such POVM operation is doubly-stochastic, as one can easily verify, so it does not decrease the von-Neumann entropy and represents a non-selective measurement.

2.4 Quantum channels

Alice plans to send messages $x \in \mathcal{X}$ with apriori probabilities $p(x)$. Instead sending the messages themselves, she *encodes* them into quantum states by applying the map

$$x \mapsto \hat{\rho}_x. \tag{2.118}$$

After encoding, Alice sends the states through a quantum channel to Bob. The quantum channel transforms each input state $\hat{\rho}_x$ via a quantum operation $\mathcal{E} : \mathcal{S}(\mathcal{H}) \to \mathcal{S}(\mathcal{H})$ into the output state

$$\hat{\rho}_x' = \mathcal{E}(\hat{\rho}_x). \tag{2.119}$$

The quantum channel is uniquely defined by this operation $\mathcal{E}$. For an external person who has no idea whatsoever about the encoding (2.118), the ensemble of quantum states is indistinguishable from the mixed state

$$\hat{\rho} = \sum_x p(x)\hat{\rho}_x. \tag{2.120}$$

The von-Neumann entropy of $\hat{\rho}$ is generally smaller or equal to the Shannon entropy of the corresponding source message ensemble $X = \{(x, p(x))|x \in \mathcal{X}\}$,

$$S(\hat{\rho}) \leq H(X), \tag{2.121}$$

where equality is reached only in case of mutually orthogonal quantum states, i.e. where

$$\mathrm{Tr}\{\hat{\rho}_x\hat{\rho}_{x'}\} = 0 \quad \text{for } x \neq x'. \tag{2.122}$$

As we will see later on, it is possible to compress an ensemble of pure quantum states down to $S(\hat{\rho})$ qubits per symbol, thus the von-Neumann entropy represents the quantum

analogon to the Shannon entropy, namely a measure for the quantum resources needed to encode the message.
Now Bob wants to find out which message $x \in \mathcal{X}$ Alice has sent. He performs a POVM measurement on $\hat{\rho}$ with elements $\{\hat{F}_y\}$ which produces some classical output $y \in \mathcal{Y}$. If Alice sends x then Bob decodes y with the conditional probability

$$q(y|x) = \mathrm{Tr}\{\hat{\rho}_x \hat{F}_y\}. \tag{2.123}$$

As one can see the quantum channel acts as a *classical noisy channel* with transition probabilities $p(x \to y) = q(y|x)$, so there is some intrinsic "quantum noise" which prevents Bob from decoding the correct input message. Only if

$$\mathrm{Tr}\{\hat{\rho}_x \hat{F}_y\} = \delta(x, y) \tag{2.124}$$

then the quantum channel acts as a noiseless classical channel. A necessary condition for this is that the input states must be mutually orthogonal. The marginal probability $q(y)$ for decoding y regardless of what Alice has sent, is given by

$$\begin{aligned} q(y) &= \sum_x q(y|x)p(x) && (2.125)\\ &= \sum_x \mathrm{Tr}\{\hat{\rho}_x \hat{F}_y\}p(x) && (2.126)\\ &= \mathrm{Tr}\{\hat{\rho}\hat{F}_y\}, && (2.127)\end{aligned}$$

so that the measurement output is represented an ensemble $Y = \{(y, q(y))|y \in \mathcal{Y}\}$. After decoding y, Bob's best guess is the message x with the highest *aposteriori* probability

$$\begin{aligned} p(x|y) &= \frac{q(y|x)p(x)}{q(y)} && (2.128)\\ &= \frac{\mathrm{Tr}\{\hat{\rho}_x \hat{F}_y\}p(x)}{\mathrm{Tr}\{\hat{\rho}\hat{F}_y\}}. && (2.129)\end{aligned}$$

Bob's *average information gain* about X when he learns the value of Y is given by the mutual information $H(X:Y) = H(X) - H(X|Y)$. Bob's task is to maximize this mutual information by choosing an appropriate POVM $\{\hat{F}_y\}$. The maximum over all POVMs is called the *accessible information*,

$$\mathrm{Acc}(\Sigma) := \max_{\{\hat{F}_y\}} H(X:Y). \tag{2.130}$$

Holevo's theorem [24, 25, 26, 27, 28] gives an upper bound to this value.

Theorem 6 (Holevo bound) *The accessible information of an ensemble* $\Sigma = \{(\hat{\rho}_x, p(x))\}$ *of quantum states is bounded from above by*

$$\mathrm{Acc}(\Sigma) \le \chi(\Sigma), \tag{2.131}$$

where

$$\chi(\Sigma) := S\Big(\sum_x p(x)\hat{\rho}_x\Big) - \sum_x p(x)S(\hat{\rho}_x) \tag{2.132}$$

is the Holevo information *of the ensemble* Σ*. The bound can be achieved asymptotically, i.e.*

$$\frac{1}{N}\mathrm{Acc}(\Sigma^{\otimes N}) \to \chi(\Sigma) \quad \textit{for } N \to \infty, \tag{2.133}$$

where $\Sigma^{\otimes N} := \{(\hat{\rho}_{x_1} \otimes \cdots \otimes \hat{\rho}_{x_N}, p(x_1)\cdots p(x_N)\}$ *is the ensemble of* N *quantum states sequentially prepared from* Σ*.*

In case of an ensemble of pure states, the Holevo information reduces to the von-Neumann entropy of the corresponding density matrix $\hat{\rho} = \sum_x p(x)\hat{\rho}_x$. Therefore the Holevo information can be regarded as a generalization of the von-Neumann entropy as a measure of the classical information that can be extracted from an ensemble of quantum states. Because

$$S\Big(\sum_x p(x)\hat{\rho}_x\Big) \geq \sum_x p(x)S(\hat{\rho}_x), \tag{2.134}$$

the Holevo information is always non-negative.

2.4.1 Channel capacity

During the transmission there are two operations which are performed on the quantum states: First, the channel performs the quantum operation $\mathcal{E}$ which in general introduces noise. Second, the receiver performs a measurement operation $\mathcal{M}$ which corresponds to a POVM measurement with some elements $\{\hat{F}_y\}$. In case of a noiseless quantum channel the operation $\mathcal{E}$ is the identity operation, but even then it is not guaranteed that all messages can be correctly decoded. In general we have

$$\mathrm{Acc}(\Sigma) \leq H(X), \tag{2.135}$$

where

$$H(X) = -\sum_x p(x)\log_2 p(x) \tag{2.136}$$

is the Shannon entropy of the input ensemble. In particular, if the input states cannot perfectly be distinguished from each other then the accessible information truly lies below the Shannon entropy of the input ensemble. In case of a noisy channel the situation even gets worse. The input ensemble $\Sigma = \{(\hat{\rho}_x, p(x))|x \in \mathcal{X}\}$ is affected by noise while passing the channel, so that the output ensemble is given by

$$\mathcal{E}(\Sigma) \equiv \{(\mathcal{E}(\hat{\rho}_x), p(x))|x \in \mathcal{X}\}. \tag{2.137}$$

Bob now tries to decode Alice's messages with the help of a suitable POVM $\{\hat{F}_y\}$ so that the mutual information $H(X\!:\!Y)$ is maximized. Because of Holevo's theorem (2.131) we know that this mutual information is bounded from above by the Holveo information (2.132),

$$H(X\!:\!Y) \leq \chi\big(\mathcal{E}(\Sigma)\big). \tag{2.138}$$

The capacity of a classical channel is given by the maximum mutual information over all input probabilities, $C = \max_{p_x} H(X\!:\!Y)$. For a quantum channel there is one more degree of freedom, namely the choice of the input states $\hat{\rho}_x$. In total, Alice has to maximize over all ensembles Σ, thus the capacity of the quantum channel is $C = \max_\Sigma H(X:Y)$. Because the maximal mutual information approaches the Holevo information asymptotically, and because the channel capacity is intended to be an asymptotic quantity, the *capacity of the quantum channel* reads

$$C = \max_{\Sigma} \chi(\mathcal{E}(\Sigma)). \tag{2.139}$$

The capacity of a quantum channel [26, 27] is almost impossible to calculate in most cases. Consequently, it is necessary to derive upper bounds restricted to a particular scenario, which are computable and tight enough to be useful for practical purposes.

2.5 Quantum messages

In section 1.1 we have defined the notion of a channel in a quite general fashion. In the quantum case, the source messages are particular states of a given quantum system. These states have to be encoded into a quantum analog of code messages, so that they can be transmitted to the receiver and reconstructed to give the original message. The encoding of quantum states into quantum messages is accomplished by the *quantum code*.

The classical code translates a source message into a code message, where the code message is a sequence of symbols taken from a certain alphabet $\mathcal{A}$. In the quantum case these symbols are represented by states of a quantum system. Let us in the following restrict the discussion to *pure* states. The set

$$\mathcal{B} = \{|\omega_1\rangle, \ldots, |\omega_k\rangle\} \tag{2.140}$$

is a quantum analogue to a classical alphabet $\mathcal{A}$ of k symbols, and it is called the *basis alphabet*. The symbols of the basis alphabet should be perfectly distinguishable, so the corresponding quantum message states have to be *mutually orthogonal*,

$$\langle\omega_i|\omega_j\rangle = \delta_{ij}. \tag{2.141}$$

In contrast to that, the messages that Alice composes from the basis alphabet by superposition do *not* have to be mutual orthogonal. The basis alphabet spans a Hilbert space

$$\mathcal{H} = \mathrm{Span}(\mathcal{B}), \tag{2.142}$$

which is called the *symbol space*. A symbol space with dimension $\dim \mathcal{H} = |\mathcal{B}| = k$ is a k-ary space and a code that uses this space as a code space is a k-ary quantum code. Quantum symbols are composed into *basis messages* by tensor multiplication,

$$|\omega^n\rangle := |\omega_1\rangle \otimes \cdots \otimes |\omega_n\rangle. \tag{2.143}$$

They form the set

$$\mathcal{B}^n := \{|\omega^n\rangle \mid |\omega_i\rangle \in \mathcal{B}\} \tag{2.144}$$

and span the *block space*

$$\mathcal{H}^{\otimes n} := \mathrm{Span}(\mathcal{B}^n), \tag{2.145}$$

giving

$$\mathcal{H}^{\otimes n} = \bigotimes_{i=1}^{n} \mathcal{H} = \mathcal{H} \otimes \cdots \otimes \mathcal{H}. \tag{2.146}$$

The space $\mathcal{H}^{\otimes n}$ is the quantum analogue to the set $\mathcal{A}^n$ of classical block messages and contains all superpositions of basis messages. The empty message, denoted by $|x^0\rangle \equiv |\square\rangle$, forms the set $\mathcal{B}^0 = \{|\square\rangle\}$ and spans the one-dimensional space $\mathcal{H}^{\otimes 0} := \mathrm{Span}(\mathcal{B}^0)$. Elements of $\mathcal{H}^{\otimes n}$ for some $n \in \mathbb{N}$ are called *block messages*. The set of all basis messages of finite length that can be composed from $\mathcal{B}$ is denoted by

$$\mathcal{B}^+ := \bigcup_{n=0}^{\infty} \mathcal{B}^n. \tag{2.147}$$

Now the *general message space* $\mathcal{H}^\oplus$ induced by a symbol space $\mathcal{H}$ can be defined by

$$\mathcal{H}^\oplus := \mathrm{Span}(\mathcal{B}^+), \tag{2.148}$$

giving

$$\mathcal{H}^\oplus = \bigoplus_{n=0}^{\infty} \mathcal{H}^{\otimes n} = \mathcal{H}^{\otimes 0} \oplus \mathcal{H} \oplus \mathcal{H}^{\otimes 2} \oplus \cdots . \tag{2.149}$$

The space $\mathcal{H}^\oplus$ is the quantum analogue to the set $\mathcal{A}^+$ of finite-length classical messages given by (1.1). A *quantum message* is any element $|x\rangle \in \mathcal{H}^\oplus$. Any quantum message which has the form

$$|x^n\rangle = |x_1\rangle \otimes \cdots \otimes |x_n\rangle, \tag{2.150}$$

is a *product message*. All other messages are *entangled messages*. Because superposition and entanglement have no classical interpretation, quantum information is truly different from classical information. $\mathcal{H}^\oplus$ is a separable Hilbert space with the countable basis $\mathcal{B}^+$. The space $\mathcal{H}^\oplus$ is the Fock space known from many-particle theory. The particles are symbols here which must be distinguishable, so there is no symmetrization or antisymmetrization. The general message space $\mathcal{H}^\oplus$ also contains superpositions of messages of distinct length. For example, for a binary symbol space $\mathcal{H}$ the vector

$$\frac{1}{\sqrt{2}}(|101\rangle + |11100\rangle) \tag{2.151}$$

is a valid quantum message in $\mathcal{H}^\oplus$. Any block space $\mathcal{H}^{\otimes n}$ is a subspace of $\mathcal{H}^\oplus$ and is orthogonal to any other block space $\mathcal{H}^{\otimes m}$ with $n \neq m$. Elements with components of distinct length are called *variable-length messages* (or *indeterminate-length messages*) to distinguish them from block messages. Any subspace $\mathcal{H}_C \subset \mathcal{H}^\oplus$ is called a *message space*.

2.5.1 Statistical messages

The generalization to statistical ensembles is straightforward. Consider an ensemble $X = \{p, \mathcal{X}\}$ of quantum messages $|x\rangle \in \mathcal{X} \subset \mathcal{H}^\oplus$ occurring with probability $p(x) > 0\ \forall |x\rangle \in \mathcal{X}$ such that $\sum_{x\in\mathcal{X}} p(x) = 1$. Then the density operator

$$\hat{\sigma} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|, \tag{2.152}$$

is the *statistical quantum message* corresponding to the ensemble X. The set of all density operators over the message space $\mathcal{H}^{\oplus}$ is denoted by $\mathcal{S}(\mathcal{H}^{\oplus})$. For a given density operator $\hat{\sigma} \in \mathcal{S}(\mathcal{H}^{\oplus})$ there is in general a non-countable set of ensembles which give the same density matrix $\hat{\sigma}$. This means that there is more information in the ensemble than in the corresponding density operator. As we will see, this additional apriori knowledge is in fact needed to make lossless compression possible.

2.5.2 Length operator

Define the *length operator* in $\mathcal{H}^{\oplus}$ measuring the length of a message as

$$\hat{L} := \sum_{n=0}^{\infty} n\,\hat{\Pi}_n, \tag{2.153}$$

where $\hat{\Pi}_n$ is the projector on the block space $\mathcal{H}^{\otimes n} \subset \mathcal{H}^{\oplus}$, given by

$$\hat{\Pi}_n = \sum_{\omega^n \in \mathcal{B}^n} |\omega^n\rangle\langle\omega^n|. \tag{2.154}$$

As $\hat{L}$ is a quantum observable, the length of a message $|x\rangle \in \mathcal{H}^{\oplus}$ is generally not sharply defined. Rather, the measurement of $\hat{L}$ generally disturbs the message by projecting it on a block space of the corresponding length. The *expected length* of a message $|x\rangle \in \mathcal{H}^{\oplus}$ is given by

$$L(x) := \langle x|\hat{L}|x\rangle. \tag{2.155}$$

However, in $\mathcal{H}^{\oplus}$ there are also messages whose expected length is infinite. Classical analoga are probability distributions with non-existing moments, e.g. the Lorentz distribution. Block messages are eigenvectors of $\hat{L}$, that is, $\hat{L}|x\rangle = n\,|x\rangle$ for all $|x\rangle \in \mathcal{H}^{\otimes n}$. The expected length of an ensemble X or of the corresponding statistical message $\hat{\sigma} \in \mathcal{S}(\mathcal{H}^{\oplus})$ is defined as

$$L(X) = L(\hat{\sigma}) := \mathrm{Tr}\{\hat{\sigma}\,\hat{L}\} = \sum_{x\in\mathcal{X}} p(x)\,L(x). \tag{2.156}$$

2.5.3 Base length

The expected length of a quantum message $|x\rangle$, given by (2.155), will in general not be the outcome of a length measurement. Every length measurement results in one of the length eigenvalues supported by $|x\rangle$ and generally disturbs the message. If there is a maximum value resulting from a length measurement of a state $|x\rangle$, namely the length of the longest component of $|x\rangle$, then let us call it the *base length* of $|x\rangle$, defined as

$$\underline{L}(x) := \max\{n \in \mathbb{N} \mid \langle x|\hat{\Pi}_n|x\rangle > 0\}. \tag{2.157}$$

For example, the quantum message

$$|x\rangle = \frac{1}{\sqrt{2}}(|\mathtt{abra}\rangle + |\mathtt{cadabra}\rangle) \tag{2.158}$$

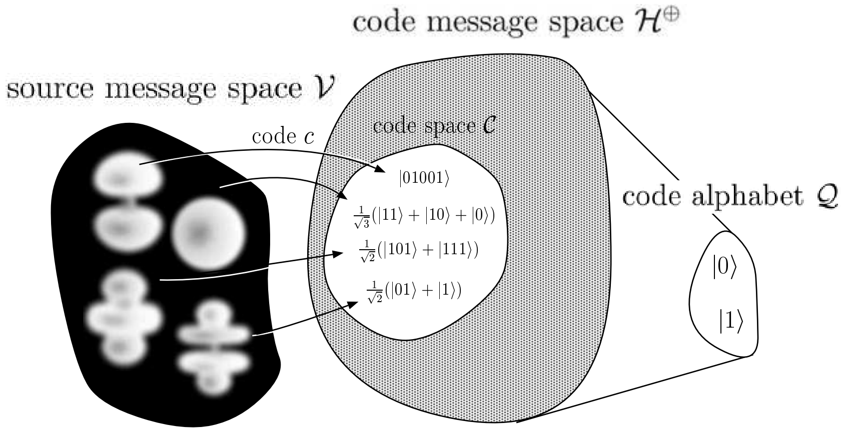


Figure 2.2: A quantum code is a linear isometric map from a source space of quantum objects into a code space of codewords composed from a quantum alphabet. Superpositions of source objects are encoded into superpositions of codewords. An ensemble of source objects is mapped to an ensemble of codewords. For a variable-length quantum code, the length of the codewords is allowed to vary. Superpositions of codewords of distinct length lead to codewords of indeterminate length. The *base length* of a codeword is defined as the length of the longest component.

has base length 7. Since the base length of a state is the size of its longest component, we have

$$\underline{L}(x) \geq L(x). \tag{2.159}$$

It is important to note that the base length is not an observable. It is only available if the message $|x\rangle$ is apriorily known.

2.5.4 Quantum code

Now we can precisely define a *k-ary quantum code* to be a linear map $c : \mathcal{V} \to \mathcal{H}^\oplus$, where $\mathcal{V}$ is a Hilbert space and $\mathcal{H}^\oplus$ is the general message space induced by a symbol space $\mathcal{H}$ of dimension k. The image of $\mathcal{V}$ under c is the *code space* $\mathcal{C} = c(\mathcal{V})$ (see Fig. 2.2). Being a quantum analogue to the codebook, $\mathcal{C}$ is the space of valid codewords. The code c is uniquely specified by the transformation rule

$$|\omega\rangle \stackrel{c}{\longmapsto} |\gamma\rangle, \tag{2.160}$$

where $|\omega\rangle$ are elements of a fixed orthonormal basis $\mathcal{B}_\mathcal{V}$ of $\mathcal{V}$ and $|\gamma\rangle = |c(\omega)\rangle$ are elements of an orthonormal basis $\mathcal{B}_\mathcal{C}$ of $\mathcal{C}$. A *lossless* quantum code is a *linear isometric* map $c : \mathcal{V} \to \mathcal{H}^\oplus$, i.e. $\langle\omega|\omega'\rangle = \langle c(\omega)|c(\omega')\rangle$, this implies that $|c(\omega)\rangle \neq |c(\omega')\rangle$ for all $|\omega\rangle \neq |\omega'\rangle$ in $\mathcal{V}$, so c is indeed a lossless code with an inverse c^{-1}. A quantum code c can be represented by the operator

$$\hat{C} := \sum_{\omega \in \mathcal{B}_\mathcal{V}} |c(\omega)\rangle\langle\omega|, \tag{2.161}$$

called the *encoder* of c. If c is lossless, there is an inverse operator

$$\hat{D} := \hat{C}^{-1} = \sum_{\omega \in \mathcal{B}_{\mathcal{V}}} |\omega\rangle\langle c(\omega)| = \sum_{\gamma \in \mathcal{B}_{\mathcal{C}}} |c^{-1}(\gamma)\rangle\langle\gamma|, \tag{2.162}$$

called the *decoder*. If c is an *error correcting code* (which is also lossless) then before decoding the message via $\hat{D}$ one performs an *error syndrom measurement* to see if an errror has modified the code basis states $|c(\omega)\rangle$. If so, one recovers the code states by applying an adequate unitary operation and then applies $\hat{D}$ to decode the original message. If c is a *lossy* code then the decoder is constructed in such a way that the decoded message is at least *similiar* to the original one, or give an error message. Error correcting and lossy codes are not our task here, so let us proceed with simple lossless quantum codes where $\hat{D} = \hat{C}^{-1}$.

In practice, the source space $\mathcal{V}$ and the code space $\mathcal{C}$ are often subspaces of one and the same physical space $\mathcal{R}$. Since $\hat{C}$ is an isometric operator between $\mathcal{V}$ and $\mathcal{C}$, there is a (non-unique) *unitary extension* $\hat{U}_C$ on $\mathcal{R}$ with

$$\hat{U}_C|x\rangle = \hat{C}|x\rangle, \quad \forall|x\rangle \in \mathcal{V} \subset \mathcal{R}, \tag{2.163}$$

$$\hat{U}_C^\dagger|y\rangle = \hat{C}^{-1}|y\rangle, \quad \forall|y\rangle \in \mathcal{C} \subset \mathcal{R}. \tag{2.164}$$

However, using $\hat{C}$ and distinguishing between $\mathcal{V}$ and $\mathcal{C}$ is more convenient and more general. Codes with $\mathcal{C} \subset \mathcal{H}^{\otimes n}$ for some $n \in \mathbb{N}$ are called *block codes*, otherwise *variable-length codes*.

2.6 Realizing variable-length messages

Variable-length messages could in principle directly be realized by a quantum system whose particle number is not conserved, for instance an electromagnetic field. Each photon may carry symbol information by its field mode, while the number of photons may represent the length of the message. The photons can be ordered either using their spacetime position (e.g. single photons running through a wire) or some internal state with many degrees of freedom (e.g. a photon with frequency ω_2 can be defined to "follow" a photon with frequency $\omega_1 < \omega_2$). The Hilbert space representing such a system of distinguishable particles with non-conserved particle number is the message space $\mathcal{H}^{\oplus}$. In case we have only a system at hand, where the number of particles is conserved, we can also realize variable-length messages by embedding them into block spaces.

It is a good idea to distinguish between the *message space*, which is a purely abstract space, from its physical realization. Let us call the physical realization of a message space $\mathcal{H}_{\mathrm{op}}$ the *operational space* $\tilde{\mathcal{H}}_{\mathrm{op}}$. Between $\mathcal{H}_{\mathrm{op}}$ and $\tilde{\mathcal{H}}_{\mathrm{op}}$, there is an isometric map, so $\dim \mathcal{H}_{\mathrm{op}} = \dim \tilde{\mathcal{H}}_{\mathrm{op}}$. This is expressed by $\mathcal{H}_{\mathrm{op}} \cong \tilde{\mathcal{H}}_{\mathrm{op}}$. The operational space $\tilde{\mathcal{H}}_{\mathrm{op}}$ is the space of physical states of a system representing valid codewords of $\mathcal{H}_{\mathrm{op}}$. Often the operational space is a subspace of the total space of all physical states of the system. Denoting the total *physical space* by $\mathcal{R}$ we have

$$\mathcal{H}_{\mathrm{op}} \cong \tilde{\mathcal{H}}_{\mathrm{op}} \subset \mathcal{R}. \tag{2.165}$$

One might object that *superselection rules* which forbid the superposition of distinct particle numbers also forbid the realization of variable-length messages. However, there are two crucial counterarguments. First, these superselection rules only apply to *massive* particles. In case of *massless* particles like photons the superposition of distinct particle numbers is rather the generic case than a bizarre exception (e.g. coherent states). Second, we actually do not deal with particles here but with *letters*. For example, a letter could be represented by a particular excited state of the Hydrogen atom and the empty letter is represented by the ground state. We then represent the variable-length message

$$\frac{1}{\sqrt{2}}(|0\rangle + |10\rangle) \tag{2.166}$$

by the physical state

$$\frac{1}{\sqrt{2}}(|0gg\rangle + |10g\rangle), \tag{2.167}$$

where $|g\rangle$ indicates the ground state. The entire message is now encoded into the joint state of three distinguishable atoms, so that the maximal length of the message is three. In the next section we will get more into detail with the possible realization of message spaces.

2.6.1 Bounded message spaces

The general message space $\mathcal{H}^{\oplus}$ is the "mother" of all message spaces induced by the symbol space $\mathcal{H}$. It contains just *every* quantum message that can be composed using symbols from $\mathcal{H}$ and the laws of quantum mechanics. However, it is an *abstract* space, i.e. independent from a particular physical implementation. It would be good to know if such a space can also physically be realized. It is clear that if one has a *finite system* one can only realize a *finite dimensional subspace* of the general message space, whose dimension is infinite. So let us start with the physical realization of the r-*bounded message space*

$$\mathcal{H}^{\oplus r} := \bigoplus_{n=0}^{r} \mathcal{H}^{\otimes n}, \tag{2.168}$$

containing all superpositions of messages of maximal length r.
Say we have a physical space $\mathcal{R} = \mathcal{D}^{\otimes s}$ representing a register consisting of s systems with $\dim \mathcal{D} = k$. Each subspace $\mathcal{D}$ represents one *quantum digit* in the register. In the case $k = 2$ the quantum digits are *quantum bits*, in short "qubits". The physical space $\mathcal{R}$ represents the space of all *physical states* of the register, while the message space $\mathcal{H}^{\oplus r}$ represents the space of *valid codewords* that can be held by the register and it is isomorphic to a subspace $\tilde{\mathcal{H}}^{\oplus r}$ of the physical space $\mathcal{R}$. Let $\dim \mathcal{H} = k$, then one must choose s such that

$$\dim(\mathcal{H}^{\oplus r}) \leq \dim(\mathcal{D}^{\otimes s}) \tag{2.169}$$

$$\Rightarrow \quad \sum_{n=0}^{r} k^n = \frac{k^{r+1} - 1}{k - 1} \leq k^s \tag{2.170}$$

$$\Rightarrow \quad s \geq r + 1. \tag{2.171}$$

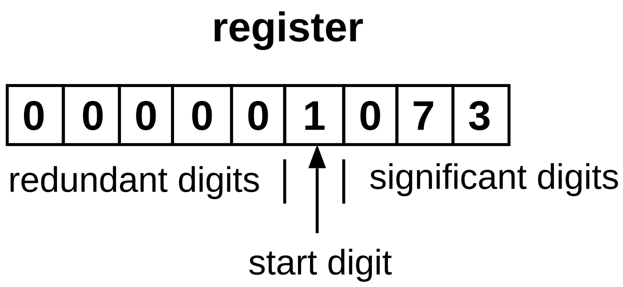


Figure 2.3: Realizing a general variable-length message.

Thus one needs a register of at least $(r+1)$ digits to realize the message space $\mathcal{H}^{\oplus r}$. Choose the smallest possible register space $\mathcal{R} = \mathcal{D}^{\otimes(r+1)}$. Since at most r digits are carrying information, one digit can be used to indicate either the beginning or the end of the message. Now we can conveniently use *k-ary representations of natural numbers* as codewords. Each natural number i has a unique k-ary representation $Z_k(i)$. For instance, $Z_2(3) = 11$ and $Z_{16}(243) = E3$. All k-ary representations have a *neutral prefix* “0” that can precede the representation without changing its value, e.g. $000011 \cong 11$. For a natural number $n > 0$, define $Z_k^n(i)$ as the *n-extended k-ary representation of i* by

$$Z_k^n(i) := \underbrace{0\cdots 0 Z_k(i)}_{n}, \quad 0 \le i \le k^r - 1. \tag{2.172}$$

For example, $Z_2^6(3) = 000011$ and $Z_{16}^6(243) = 0000E3$. Let us define that the message starts after the first appearance of “1”, e.g. $000102540 \cong 02540$. Now define orthonormal vectors

$$|e_i^n\rangle := |\underbrace{0\cdots 0}_{r-n} 1 Z_k^n(i)\rangle \in \mathcal{R} \tag{2.173}$$

where $n > 0$ and $0 \le i \le k^n - 1$. The n digits of $Z_k^n(i)$ are called *significant digits*. The empty message corresponds to the unit vector

$$|\Box\rangle := |e_0^0\rangle := |0\cdots 01\rangle. \tag{2.174}$$

Obviously, $|\Box\rangle$ has no significant digits. Next, define orthonormal basis sets

$$\tilde{\mathcal{B}}^n := \{|e_0^n\rangle, \ldots, |e_{k^n-1}^n\rangle\}, \quad 0 \le n \le r, \tag{2.175}$$

that span the operational block spaces

$$\tilde{\mathcal{H}}^{\otimes n} = \mathrm{Span}(\tilde{\mathcal{B}}^n). \tag{2.176}$$

Note that $\tilde{\mathcal{H}}^{\otimes n}$ is truly different from $\mathcal{H}^{\otimes n}$, because $\tilde{\mathcal{H}}^{\otimes n}$ has dimension k^{r+1}, while $\tilde{\mathcal{H}}^{\otimes n}$ has dimension k^n. Next, define an orthonormal basis

$$\tilde{\mathcal{B}}^+ := \bigcup_{n=0}^{r} \tilde{\mathcal{B}}^n, \tag{2.177}$$

and construct the operational space $\tilde{\mathcal{H}}^{\oplus r} \subset \mathcal{R}$ by

$$\tilde{\mathcal{H}}^{\oplus r} := \mathrm{Span}(\tilde{\mathcal{B}}^{+}). \tag{2.178}$$

Altogether, the physical space $\mathcal{R} = \mathcal{D}^{\otimes(r+1)}$ is the space of all physical states of the register, while the operational space $\tilde{\mathcal{H}}^{\oplus r} \subset \mathcal{R}$ is the space of those register states that represent valid codewords, and it is isomorphic to the abstract message space $\mathcal{H}^{\oplus r}$.
A message is represented by the vector

$$|x\rangle = \sum_{n=0}^{r} \sum_{i=0}^{k^n-1} x_{n,i}\, |e_i^n\rangle \tag{2.179}$$

with $\sum_{n=0}^{r} \sum_{i=0}^{k^n-1} |x_{n,i}|^2 = 1$. The length operator introduced in section 2.5.2 is here of the form

$$\hat{L} := \sum_{n=0}^{r} n\, \hat{\Pi}_n, \tag{2.180}$$

because there are at most r digits to constitute a message. Now we need to know how the projectors $\hat{\Pi}_n$ are constructed in the operational space $\tilde{\mathcal{H}}^{\oplus r}$. For a register state containing a message of sharply defined length, the length eigenvalue n is given by the *number of significant digits* in that register,

$$\hat{L}\, |e_i^n\rangle := n\, |e_i^n\rangle, \tag{2.181}$$

for $0 \le i \le k^n - 1$. Each projector is then defined by

$$\hat{\Pi}_n := \sum_{i=0}^{k^n-1} |e_i^n\rangle\langle e_i^n| \tag{2.182}$$

and projects onto the space $\mathcal{H}^{\otimes n} \subset \mathcal{R}$. Note that the *physical length* of each message is always given by the fixed size $(r+1)$ of the register. Only the *significant length* of a message, i.e. the number of digits that constitute a message contained in the register, is in general not sharply defined. Note further that the particular form of the length operator depends on the realization of the message space.
In the limit of large r we have $\lim\limits_{r\to\infty} \mathcal{H}^{\oplus r} = \mathcal{H}^{\oplus}$, but that space can no longer be embedded into a physical space $\mathcal{R} = \mathcal{D}^{\otimes\infty} := \lim\limits_{n\to\infty} \mathcal{D}^{\otimes n}$, since the latter is no separable Hilbert space anymore. However, we can think of r as *very large*, such that working in $\mathcal{H}^{\oplus}$ just means working with a quantum computer having enough memory.

2.6.2 Realizing more message spaces

A code is a map $c : \mathcal{V} \to \mathcal{H}^{\oplus}$ from source states in $\mathcal{V}$ to codewords in $\mathcal{H}^{\oplus}$. The space $\mathcal{C} = c(\mathcal{V})$ of all codewords is the *code space* and as a subspace of the general message space $\mathcal{H}^{\oplus}$ it is just a special message space. In order to implement a particular code c, it is in practice sufficient to realize only the corresponding code space $\mathcal{C}$ by a physical system. Let us realize some important code spaces now. However, we will not discuss the class of *error-correcting* code spaces here, since this would go beyond the scope of this thesis.

Block spaces

An important message space is the *block space* $\mathcal{H}^{\otimes n}$, that contains messages of fixed length n. Block spaces are *the* message spaces of standard quantum information theory. They can directly be realized by a register $\mathcal{R} = \mathcal{H}^{\otimes n}$ of n digits, e.g. n two-level systems representing one qubit each.

Prefix spaces

Another interesting message space is the space of *prefix codewords* of maximal length r. Such a space contains only superpositions of prefix codewords. A set of codewords is *prefix* (or *prefix-free*) if no codeword is the prefix of another codeword. For example, the set $P_3 = \{0, 10, 110, 111\}$ is a set of binary prefix codewords of maximal length 3. Prefix codewords have one significant advantage:

- Prefix codewords are *instantaneous*, that is, sequences of prefix codewords do not need a word separator. The separator can be added while reading the sequence from left to right. A sequence from P_3 can be separated like

$$110111010110 \mapsto 110, 111, 0, 10, 110. \tag{2.183}$$

However, there is also a drawback:

- Prefix codewords are in general not as short as possible.

This is a consequence of the fact that there are in general less prefix codewords than possible codewords. For example, if we want to encode 4 different objects, we can use the prefix set P_3 above with maximal length 3. If we renounce the prefix property we can use the set $\{0, 1, 01, 10\}$ with maximal length 2.
A *prefix space* $\mathcal{P}_r$ of maximal length r is given by the linear span of prefix codewords of maximal length r. For the set P_3, the corresponding prefix space is $\mathcal{P}_3 = \mathrm{Span}\{|0\rangle, |10\rangle, |110\rangle, |111\rangle\}$. The prefix space $\mathcal{P}_r \subset \mathcal{H}^{\oplus r}$ can physically be realized by a subspace $\tilde{\mathcal{P}}_r$ of the register space $\mathcal{R} = \mathcal{D}^{\otimes r}$ spanned by the prefix codewords which have been extended by zeroes at the end to fit them into the register. For example, $\tilde{\mathcal{P}}_3 = \mathrm{Span}\{|000\rangle, |100\rangle, |110\rangle, |111\rangle\} \subset \mathcal{D}^{\otimes 3}$ is a physical realization of the prefix space $\mathcal{P}_3$. The length operator measures the *significant length* of the codewords, given by the length of the corresponding prefix codewords.
Schumacher and Westmoreland [11] as well as *Braunstein et al.* [10] used prefix spaces for their implementation of variable-length quantum coding in form of a quantum analogue of the Huffman code. However, we will show later on that the significant advantage of prefix codewords in fact vanishes in the quantum case, whereas the disadvantage remains.

Neutral-prefix space

A specific code space will be of interest, namely the space of *neutral-prefix codewords*, which we define as follows. The k-ary representation of a natural number i is denoted by

Figure 2.4: Realizing variable-length messages by neutral-prefix codewords.

$Z_k(i)$ (see section 2.6.1). The empty message $\square$ is represented by $Z_k(0) = \square$. Define an orthonormal basis

$$\mathcal{B}_r := \{|Z_k(0)\rangle, \ldots, |Z_k(k^r - 1)\rangle\} \tag{2.184}$$

of variable-length messages of maximal length r. The length of each basis message $|Z_k(i)\rangle$ is given by

$$|Z_k(i)| = \lceil \log_k(i+1) \rceil, \tag{2.185}$$

where $\lceil x \rceil$ denotes the smallest integer $\geq x$. These basis messages span the r-*bounded neutral-prefix space*

$$\mathcal{N}_r := \mathrm{Span}(\mathcal{B}_r). \tag{2.186}$$

Note that $\mathcal{N}_r$ is not equal to the r-bounded message space $\mathcal{H}^{\oplus r}$ as one can see by comparing the dimension $\dim \mathcal{N}_r = k^r$ with $\dim \mathcal{H}^{\oplus r} = \frac{k^{r+1}-1}{k-1}$. $\mathcal{N}_r$ is smaller than $\mathcal{H}^{\oplus r}$, because not all messages of $\mathcal{H}^{\oplus r}$ are contained in $\mathcal{N}_r$. For example, the message $|01\rangle$ is in $\mathcal{H}^{\oplus r}$ but not in $\mathcal{N}_r$, hence we have

$$\mathcal{N}_r \subset \mathcal{H}^{\oplus r}. \tag{2.187}$$

Now we want to find a physical realization of $\mathcal{N}_r$. This turns out to be quite easy (see Fig. 2.4). As already noted in section 2.6.1, the k-ary representation $Z_k(i)$ of any natural number i can be extended by leading zeroes to the r-extended k-ary representation $Z_k^r(i) := 0 \cdots 0 Z_k(i)$. Take a register $\mathcal{R} = \mathcal{D}^{\otimes r}$ of r digits with $\mathcal{D} = \mathbb{C}^k$. Then the set

$$\mathcal{B}_{\mathcal{R}} := \{|Z_k^r(0)\rangle, \ldots, |Z_k^r(k^r - 1)\rangle\} \tag{2.188}$$

is an orthonormal basis for the register space $\mathcal{R}$. At the same time it can be regarded as an orthonormal basis for the operational space $\tilde{\mathcal{N}}_r$ representing the neutral-prefix space $\mathcal{N}_r$. While the *physical length* of each codeword is constantly r, the *significant length* is measured by the length operator

$$\hat{L} := \sum_{n=0}^{r} n\, \hat{\Pi}_n, \tag{2.189}$$

with mutually orthogonal projectors

$$\hat{\Pi}_n := \sum_{i:\ |Z_k(i)|=n} |Z_k^r(i)\rangle\langle Z_k^r(i)|. \tag{2.190}$$

Note that the so-defined length operator looks different from the one defined in section 2.6.1. While $\hat{L}$ is always of the same form (2.180), the projectors $\hat{\Pi}_n$ are different because the operational spaces are different.

The empty message can be defined by

$$|\square\rangle := |Z_k^r(0)\rangle = |0\cdots 0\rangle. \tag{2.191}$$

A code message space in $\tilde{\mathcal{N}}_r$ is given by

$$|x\rangle = \sum_{i=0}^{k^r-1} x_i \, |Z_k^r(i)\rangle. \tag{2.192}$$

We have realized the neutral-prefix space $\mathcal{N}_r$ by exhausting the entire register space $\mathcal{R}$, so the quantum resources are optimally used. In other words:

- All messages in $\mathcal{N}_r$ are as short as possible.

Remember that the physical realization of $\mathcal{H}^{\oplus r}$ requires one additional digit to represent the beginning or the end of a message. This digit does not contain any message information, it is sort of wasted. For quantum coding, the additional digit may really count, since it would have to be added each time a codeword is stored or transmitted! Also the prefix space considered in section 2.6.2 contains messages which are not as short as possible. One can encode a space $\mathcal{V}$ of dimension $\dim\mathcal{V} = 4$ by a prefix space spanned by $\{|000\rangle, |100\rangle, |110\rangle, |111\rangle\}$ with corresponding lengths $\{1, 2, 3, 3\}$, but then we need a register of 3 qubits. In contrast to that, $\mathcal{V}$ can be encoded by a neutral-prefix space spanned by the basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ with corresponding lengths $\{0, 1, 2, 2\}$, and we need a register of only 2 qubits. In the operational space $\tilde{\mathcal{N}}_r$, the basis messages reveal their length information by simply discarding leading zeroes. That way, not all variable-length messages can be realized, but we save 1 register digit, so $\mathcal{N}_r$ is a good candidate for variable-length quantum coding.

Part II

Quantum Data Compression

Chapter 3

Concepts of Quantum Data Compression

3.1 Information content of a quantum message

In section 1.4 we have defined compression as reducing the size of the message ensemble, i.e. reducing the effort of communication. The amount of compression is measured by the *effectivity* of the code (1.20), i.e. the ratio of uncompressed size to compressed size,

$$\eta_c = \frac{\mathcal{I}_0(M)}{\mathcal{I}_c(M)}. \tag{3.1}$$

The size $\mathcal{I}_c(M)$ (which we have also called the *code information content*) of the message ensemble M is the average over the size of the individual messages m, which is measured by (1.12),

$$\mathcal{I}_c(m) = \log_2 |\mathcal{A}_c| \cdot L_c(m), \tag{3.2}$$

where $\mathcal{A}_c$ is the code alphabet and $L_c(m)$ is the length of the encoded message m. The task is now to translate these concepts to the quantum case. The raw information (1.19) of an ensemble M is $\mathcal{I}_0(M) = \log_2 |\mathcal{M}|$ because we need $|\mathcal{M}|$ distinct symbols to encode each element of the message set $\mathcal{M}$ by a raw code. Interpreting $\mathcal{M}$ as an orthonormal basis for a Hilbert space $\mathcal{V}$, the raw information of $\mathcal{V}$ is also $\log_2 |\mathcal{M}|$, because we still need $|\mathcal{M}|$ distinguishable symbols to represent each element of the space $\mathcal{V}$. Since $|\mathcal{M}| = \dim \mathcal{V}$, we define the *quantum raw information content* of a space $\mathcal{V}$ as

$$\mathcal{I}_0(\mathcal{V}) := \log_2(\dim \mathcal{V}). \tag{3.3}$$

So the quantum raw information $\mathcal{I}_0$ corresponding to a space $\mathcal{V}$ equals the fixed number of qubits needed to represent all states in $\mathcal{V}$.
The amount of qubits that is occupied by a given quantum message $|x\rangle \in \mathcal{H}^{\oplus}$ is measured by the *information operator*

$$\hat{I} := \log_2 k \cdot \hat{L}, \tag{3.4}$$

where $\hat{L}$ is the length operator in $\mathcal{H}^{\oplus}$. For a given k-ary code $c : \mathcal{V} \to \mathcal{H}^{\oplus}$ represented by an encoder $\hat{C}$, the *code information operator* can be defined as

$$\hat{I}_c := \log_2 k \cdot \hat{L}_c, \tag{3.5}$$

where $\hat{L}_c := \hat{C}^{-1}\,\hat{L}\,\hat{C}$ is the length operator measuring the length of the codeword for a source vector in $\mathcal{V}$. If the code is based on a qubit alphabet, $\hat{I}_c$ measures the number of qubits forming the code message, hence the measuring unit of $\hat{I}_c$ is "1 qubit".
In short, the effective quantum information operator is defined in an arbitrary Hilbert space $\mathcal{V}$ and depends on a quantum code $c : \mathcal{V} \to \mathcal{H}^{\oplus}$, while the direct information operator is defined in a message space $\mathcal{H}^{\oplus}$ without referring to a quantum code. For a given code, the relation between both operators is

$$\hat{I}_c = C^{-1}\,\hat{I}\,C. \tag{3.6}$$

Now we want to compress a codeword by removing redundant quantum digits. The number of quantum digits carrying information is given by the base length of the codeword. All other digits are redundant and can be removed without loss of information. This motivates the definition of the *effective quantum information* of a state $|x\rangle \in \mathcal{V}$ respecting a code c by

$$\underline{\mathcal{I}}_c(x) := \log_2 k \cdot \underline{L}_c(x), \tag{3.7}$$

where $\underline{L}_c(x) = \underline{L}(c(x))$ is the base length of the codeword for $|x\rangle$. $\underline{\mathcal{I}}_c(x)$ represents the number of qubits needed to describe the state $|x\rangle$ by the code c. This value must be distinguished from the *expected* number of qubits $\mathcal{I}_c(x) = \langle x|\hat{I}_c|x\rangle$ that is found by performing a length measurement on the codeword for $|x\rangle$. In the classical case, the difference vanishes.
Now suppose we want to encode an ensemble $X = \{p, \mathcal{X}\}$ of states $|x\rangle \in \mathcal{X}$ that span the source space $\mathcal{V}$. Each individual message $|x\rangle$ can be compressed to $\underline{\mathcal{I}}_c(x)$ qubits, so the entire ensemble X will on average be compressed to the effective quantum information

$$\underline{\mathcal{I}}_c(X) := \log_2 k \sum_{x\in\mathcal{X}} p(x)\,\underline{L}_c(x). \tag{3.8}$$

In analogy to our classical definition (1.20), we define the effectivity of a quantum code by

$$\eta_c(X) := \frac{\mathcal{I}_0(X)}{\mathcal{I}_c(\mathcal{V})}. \tag{3.9}$$

A code c is compressive on the ensemble X if and only if

$$\eta_c(X) > 1 \quad \text{i.e.} \quad \underline{\mathcal{I}}_c(X) < \mathcal{I}_0(\mathcal{V}). \tag{3.10}$$

3.2 Schumacher compression

The compression scheme raised by Schumacher [6, 7] is the quantum analog to Shannon's source coding theorem. Let us now get into the details of Schumacher compression in order to understand the basic principles of quantum compression. These experiences will be helpful when we will later be looking out for a lossless compression scheme.

Alice composes a random message $\boldsymbol{x} \equiv x_1 \cdots x_N$ of length N from the message set $\mathcal{M} = \mathcal{A} \times \cdots \times \mathcal{A}$ by choosing N letters independently from the same letter ensemble. The resulting quantum message has the form

$$|\boldsymbol{x}\rangle \equiv |x_1\rangle \otimes \cdots \otimes |x_N\rangle, \tag{3.11}$$

where each letter $|x_n\rangle$ is an element of the quantum source alphabet

$$\mathcal{Q}_{\mathcal{A}} = \{|x\rangle | x \in \mathcal{A}\} \tag{3.12}$$

and the entire message is a vector from the Hilbert space

$$\mathcal{H}_M := \mathcal{H}_A^{\otimes N} \equiv \mathcal{H}_A \otimes \cdots \otimes \mathcal{H}_A. \tag{3.13}$$

Alice chooses each letter $|x\rangle \in \mathcal{Q}_A$ with apriori probability $p(x)$, so the ensemble of letters is represented by $X = \{\mathcal{Q}_A, p\}$, which corresponds to the letter matrix

$$\hat{\rho} = \sum_{x \in \mathcal{A}} p(x)\, |x\rangle\langle x|. \tag{3.14}$$

Each letter is chosen independently, so the message $|\boldsymbol{x}\rangle$ appears with probability $p(\boldsymbol{x}) = p(x_1) \cdots p(x_N)$, and the total message ensemble can be represented by the *message matrix*

$$\hat{\boldsymbol{\rho}} = \hat{\rho}^{\otimes N} = \sum_{\boldsymbol{x} \in \mathcal{M}} p(\boldsymbol{x})\, |\boldsymbol{x}\rangle\langle \boldsymbol{x}|. \tag{3.15}$$

Before sending her block message to Bob, Alice has to convert the message into a sequence of qubits, because the channel to Bob only accepts qubits. This conversion should be invertible, so Alice has to build an *encoder* $\hat{C}$ that unitarily maps the source space $\mathcal{H}_M$ to a code space $\mathcal{H}_C$ of qubits,

$$\hat{C} : \mathcal{H}_M \to \mathcal{H}_C. \tag{3.16}$$

In order for $\hat{C}$ to be unitary, the dimension of $\mathcal{H}_C$ must be equal to the dimension of the source space $\mathcal{H}_M$. The dimension of the alphabet space is at most $K = |\mathcal{Q}_A|$, but the quantum letters $|a_k\rangle \in \mathcal{Q}_A$ do not have to be *mututally orthogonal*, yet they do not even have to be *linearly independent*, so the dimension of $\mathcal{H}_A$ can in fact be *smaller* than the number of alphabet letters, which gives

$$\log(\dim \mathcal{H}_M) = N \log(\dim \mathcal{H}_A) \leq N \log K, \tag{3.17}$$

where logs are binary, here and in the following. A message $|\boldsymbol{x}\rangle$ is encoded into the binary message $|c(\boldsymbol{x})\rangle$ by applying the encoder $\hat{C}$,

$$|c(\boldsymbol{x})\rangle := \hat{C}|\boldsymbol{x}\rangle. \tag{3.18}$$

Note that while $|\boldsymbol{x}\rangle$ is by construction a product state, the code state $|c(\boldsymbol{x})\rangle$ can be highly entangled. Let the alphabet space $\mathcal{H}_A$ have dimension L, then in order to encode every message in $\mathcal{H}_M$ we need a qubit space of dimension

$$\dim \mathcal{H}_C = \dim \mathcal{H}_M = L^N, \tag{3.19}$$

where we assume that L and N are chosen such that the above number is a power of two. In other words, we need $N \log L$ qubits to encode each message in $\mathcal{H}_M$ with perfect fidelity. The decoding procedure is represented by the inverse operator $\hat{D} : \mathcal{H}_C \to \mathcal{H}_M$,

$$\hat{D} := \hat{C}^\dagger. \tag{3.20}$$

Since the use of a quantum channel is very expensive, we want to save qubits for the transmission. Let us look for an encoder $\hat{C}$ that is restricted to a proper subspace $\Lambda \subset \mathcal{H}_M$ with a dimension significantly smaller than L^N, such that we still achieve asymptotically faithful decoding. First, we perform a diagonalization of the letter matrix, resulting in

$$\hat{\rho} = \sum_{l=1}^{L} q_l \, |\lambda_l\rangle\langle\lambda_l|. \tag{3.21}$$

The number L of $\hat{\rho}$-eigenstates coincides with the dimension of the alphabet subspace $\mathcal{H}_A$. We have

$$\hat{\rho} \log \hat{\rho} = \sum_{l=1}^{L} q_l \log q_l \, |\lambda_l\rangle\langle\lambda_l|, \tag{3.22}$$

such that

$$\mathrm{Tr}\{\hat{\rho} \log \hat{\rho}\} = \sum_{l=1}^{L} q_l \log q_l = H(Y), \tag{3.23}$$

where Y denotes the ensemble of $\hat{\rho}$-eigenstates. The von-Neumann entropy of $\hat{\rho}$ equals the Shannon entropy of the ensemble of $\hat{\rho}$-eigenstates,

$$S(\hat{\rho}) = H(Y). \tag{3.24}$$

One can show that the von-Neumann entropy is bounded from above by the Shannon entropy of the letter ensemble X,

$$S(\hat{\rho}) \leq H(X), \tag{3.25}$$

where equality holds in the case of mutual orthogonal letter states.
Quantum mechanics tells us that the scenario where Alice sends the ensemble Y cannot by any experiment be distinguished from the actual scenario where Alice sends the ensemble X. However, sending the ensemble Y corresponds to a classical situation. Consider the sequence $|\boldsymbol{y}\rangle \equiv |y_1 \cdots y_N\rangle$ of basis states $|y_n\rangle \in \mathcal{B}_A$, which appear with probability $q(\boldsymbol{y}) = q(y_1) \cdots q(y_N)$. Just like in Shannon's noiseless coding theorem we introduce a typical subset T of messages $\boldsymbol{y}$ appearing with probability

$$2^{-N(S+\delta)} \leq q(\boldsymbol{y}) \leq 2^{-N(S-\delta)}, \tag{3.26}$$

where we have used the fact that $H(Y) = S(\hat{\rho}) \equiv S$. Then we define the *typical subspace* $\Lambda \subset \mathcal{H}_A$ as the space spanned by the typical messages,

$$\Lambda := \mathrm{Span}\{|\boldsymbol{y}\rangle \mid \boldsymbol{y} \in T\}. \tag{3.27}$$

Exploiting Shannon's theorem we know that for any fixed $\epsilon, \delta > 0$ there is a big enough N such that

$$P_\Lambda \geq 1 - \epsilon, \tag{3.28}$$

where P_Λ is the total probability of all members of T,

$$P_\Lambda = \sum_{\boldsymbol{y} \in T} q(\boldsymbol{y}). \tag{3.29}$$

Since the typical subspace Λ is spanned by the typical messages $|\boldsymbol{y}\rangle$ where $\boldsymbol{y} \in T$, the dimension of Λ is given by the size of T, so Shannon's theorem implies that

$$(1-\epsilon)2^{N(S-\delta)} \leq \dim\Lambda \leq 2^{N(S+\delta)}. \tag{3.30}$$

In the asymptotic limit $N \to \infty$, the dimension of the subspace approaches

$$\dim\Lambda \to 2^{NS}. \tag{3.31}$$

Because we have

$$S(\hat{\rho}) = \sum_{l=1}^{L} q_l \log q_l \leq \log L, \tag{3.32}$$

the dimension of Λ is smaller than or equal to the dimension of the space of all messages,

$$\dim\Lambda = 2^{NS(\hat{\rho})} \leq 2^{N\log L} = \dim\mathcal{H}_M. \tag{3.33}$$

In practice, except for the case of uniformly distributed letters, the typical subspace will have a *dramatically* smaller dimension (for large N). Hence we can save resources by encoding only the component of $|\boldsymbol{x}\rangle$ that lies in the typical subspace Λ. To this aim we need the projector onto the typical subspace, which is given by

$$\hat{\Pi}_\Lambda = \sum_{\boldsymbol{y}\in T} |\boldsymbol{y}\rangle\langle\boldsymbol{y}|. \tag{3.34}$$

Now we restrict the encoder to the typical subspace, $\hat{C} : \Lambda \to \mathcal{H}_C$, where it shall be a unitary operator. *Schumacher compression* goes as follows. First, Alice projects her source message $|\boldsymbol{x}\rangle$ onto the typical subspace Λ. With probability

$$P_\Lambda(\boldsymbol{x}) := \langle\boldsymbol{x}|\hat{\Pi}_\Lambda|\boldsymbol{x}\rangle = \mathrm{Tr}\{|\boldsymbol{x}\rangle\langle\boldsymbol{x}|\hat{\Pi}_\Lambda\}. \tag{3.35}$$

such projection will be successful and results in the state

$$|\phi(\boldsymbol{x})\rangle := \frac{1}{\sqrt{P_\Lambda(\boldsymbol{x})}}\hat{\Pi}_\Lambda|\boldsymbol{x}\rangle, \tag{3.36}$$

The average probability of a successful projection thus reads

$$\begin{aligned}
\langle P_\Lambda(\boldsymbol{X})\rangle &= \sum_{\boldsymbol{x}\in\mathcal{M}} p(\boldsymbol{x})P_\Lambda(\boldsymbol{x}) && (3.37)\\
&= \sum_{\boldsymbol{x}\in\mathcal{M}} p(\boldsymbol{x})\mathrm{Tr}\{|\boldsymbol{x}\rangle\langle\boldsymbol{x}|\hat{\Pi}_\Lambda\} && (3.38)\\
&= \mathrm{Tr}\Big\{\sum_{\boldsymbol{x}\in\mathcal{M}} p(\boldsymbol{x})|\boldsymbol{x}\rangle\langle\boldsymbol{x}|\hat{\Pi}_\Lambda\Big\} && (3.39)\\
&= \mathrm{Tr}\{\hat{\boldsymbol{\rho}}\,\hat{\Pi}_\Lambda\}, && (3.40)
\end{aligned}$$

that is,

$$\langle P_\Lambda(\boldsymbol{X})\rangle = \mathrm{Tr}\{\hat{\boldsymbol{\rho}}\hat{\Pi}_\Lambda\}. \tag{3.41}$$

Let us proceed,

$$\mathrm{Tr}\{\hat{\boldsymbol{\rho}}\hat{\Pi}_\Lambda\} = \sum_{\boldsymbol{y}\in T}\langle\boldsymbol{y}|\hat{\boldsymbol{\rho}}|\boldsymbol{y}\rangle \tag{3.42}$$

$$= \sum_{y\in T} q(\boldsymbol{y}) \equiv P_\Lambda \geq 1-\epsilon, \tag{3.43}$$

where in the last step we used Shannon's theorem. We arrive at

$$\langle P_\Lambda(\boldsymbol{X})\rangle = P_\Lambda \geq 1-\epsilon, \tag{3.44}$$

hence the projection will be 100% successful in the asymptotic limit of infinitely long messages. After projection, Alice can encode the resulting state $|\phi(\boldsymbol{x})\rangle$ by $\hat{C}$ and send it to Bob, who then applies the inverse operation $\hat{C}^\dagger$ to obtain the state $|\phi(\boldsymbol{x})\rangle$. If the overlap of $|\phi(\boldsymbol{x})\rangle$ with the original message is big enough, it was an approximately faithful transmission. If the projection was not successful, Alice prepares some garbage state $|\phi_0\rangle \in \Lambda$, encodes it by $\hat{C}$ and sends it to Bob. In this case, the overlap with the orininal message $|\boldsymbol{x}\rangle$ is hopefully very small. To put this more precisely, we describe the statistical ensemble of successful and unsuccessful projections by a density matrix. The probability that the projection is *not* successful reads

$$1 - P_\Lambda(\boldsymbol{x}) = \langle\boldsymbol{x}|(\mathbb{1}-\hat{\Pi}_\Lambda)|\boldsymbol{x}\rangle. \tag{3.45}$$

So after the projection procedure the message will be in the mixed state

$$\hat{\rho}_{\boldsymbol{x}} = P_\Lambda(\boldsymbol{x})|\phi(\boldsymbol{x})\rangle\langle\phi(\boldsymbol{x})| + (1-P_\Lambda(\boldsymbol{x}))|\phi_0\rangle\phi_0| \tag{3.46}$$

$$= \hat{\Pi}_\Lambda|\boldsymbol{x}\rangle\langle\boldsymbol{x}|\hat{\Pi}_\Lambda + (1-P_\Lambda(\boldsymbol{x}))|\phi_0\rangle\langle\phi_0|. \tag{3.47}$$

The subsequently performed encoding procedure by $\hat{C}$ maps the state $\hat{\rho}_{\boldsymbol{x}}$ to the qubit state $\hat{C}\hat{\rho}_{\boldsymbol{x}}\hat{C}^\dagger$, which is then send to Bob through the quantum channel. After receiving the code message, Bob applies the decoder $\hat{D} = \hat{C}^\dagger$ to it and since $\hat{C}$ is unitary, he recovers the state $\hat{\rho}_{\boldsymbol{x}}$. Originally, the message was given by the pure state $|\boldsymbol{x}\rangle\langle\boldsymbol{x}|$. The fidelity between original and decoded message is given by

$$F(\boldsymbol{x}) = \langle\boldsymbol{x}|\hat{\rho}_{\boldsymbol{x}}|\boldsymbol{x}\rangle \tag{3.48}$$

$$= \langle\boldsymbol{x}|\hat{\Pi}_\Lambda|\boldsymbol{x}\rangle\langle\boldsymbol{x}|\hat{\Pi}_\Lambda|\boldsymbol{x}\rangle + r_{\boldsymbol{x}} \tag{3.49}$$

$$= P_\Lambda^2(\boldsymbol{x}) + r_{\boldsymbol{x}} \tag{3.50}$$

$$\geq P_\Lambda^2 \geq 2\,P_\Lambda(\boldsymbol{x}) - 1, \tag{3.51}$$

where we used

$$r_{\boldsymbol{x}} := \langle\boldsymbol{x}|\big\{(1-P_\Lambda(\boldsymbol{x}))|\phi_0\rangle\langle\phi_0|\big\}|\boldsymbol{x}\rangle \geq 0 \tag{3.52}$$

together with the inequality $x^2 \geq 2x-1$, which holds for all real numbers x. So the average fidelity for the ensemble $\boldsymbol{x}$ of source messages reads

$$F = \sum_{\boldsymbol{x}\in\mathcal{M}} p(\boldsymbol{x})\,F(\boldsymbol{x}) \tag{3.53}$$

$$\geq \sum_{\boldsymbol{x}\in\mathcal{M}} p(\boldsymbol{x})\,\big(2\,P_\Lambda(\boldsymbol{x})-1\big) \tag{3.54}$$

$$= 2\,\mathrm{Tr}\{\hat{\boldsymbol{\rho}}\hat{\Pi}_\Lambda\} - 1 \geq 1-2\epsilon, \tag{3.55}$$

where we used (3.43). We arrive at the important conclusion: The average fidelity of the decoded states with the original messages tends to unity in the limit of infinitely long messages. States that do not survive the projection will be all encoded by the same junk state, which thus cannot be faithfully decoded to give the original message. Happily, the probability of such erroneous decoding vanishes in the limit of infinitely long messages. Since the dimension of the typical space approaches $d \to 2^{NS}$, we need $I_N = NS(\hat{\rho})$ qubits to encode each typical message, hence per source letter we need

$$I = S(\hat{\rho}) \tag{3.56}$$

qubits in the limit of infinitely long messages, which represents a significant compression in most practical cases. Now let us investigate if we can achieve a compression below $S(\hat{\rho})$ qubits. Just like in the classical case, we fix some $\epsilon' > 0$ and project the source message on a "subtypical subspace" $\Lambda' \subset \Lambda$ whose dimension is

$$\dim \Lambda' \leq (1-\epsilon)2^{N(S-\delta-\epsilon')} < 2^{N(H-\delta-\epsilon')}. \tag{3.57}$$

Let the space Λ' be spanned by the messages in a "subtypical set" $T' \subset T$,

$$\Lambda' := \mathrm{Span}\{|\boldsymbol{y}\rangle \mid \boldsymbol{y} \in T'\}, \tag{3.58}$$

so the dimension of Λ' equals the size of T',

$$\dim \Lambda' = |T'|. \tag{3.59}$$

The probability that a given message $|\boldsymbol{x}\rangle$ is successfully projected onto Λ' reads

$$\begin{aligned} P_{\Lambda'}(\boldsymbol{x}) &= \langle \boldsymbol{x}|\hat{\Pi}_{\Lambda'}|\boldsymbol{x}\rangle && (3.60)\\ &= \mathrm{Tr}\{|\boldsymbol{x}\rangle\langle \boldsymbol{x}|\hat{\Pi}_{\Lambda'}\}, && (3.61) \end{aligned}$$

and the projected state is then given by

$$|\phi'(\boldsymbol{x})\rangle := \frac{1}{\sqrt{P_{\Lambda'}(\boldsymbol{x})}}\hat{\Pi}_{\Lambda'}|\boldsymbol{x}\rangle. \tag{3.62}$$

The average probability that a message is successfully projected onto Λ' yields

$$\begin{aligned} P_{\Lambda'} &= \sum_{\boldsymbol{x}\in\mathcal{M}} p(\boldsymbol{x})P_{\Lambda'}(\boldsymbol{x}) && (3.63)\\ &= \sum_{\boldsymbol{x}\in\mathcal{M}} p(\boldsymbol{x})\,\mathrm{Tr}\{|\boldsymbol{x}\rangle\langle \boldsymbol{x}|\hat{\Pi}_{\Lambda'}\} && (3.64)\\ &= \mathrm{Tr}\{\hat{\boldsymbol{\rho}}\hat{\Pi}_{\Lambda'}\} = \sum_{\boldsymbol{y}\in T'}\langle \boldsymbol{y}|\hat{\boldsymbol{\rho}}|\boldsymbol{y}\rangle = \sum_{\boldsymbol{y}\in T'} q(\boldsymbol{y}) && (3.65)\\ &\leq q_{max}|T'| \leq 2^{-N(S-\delta)}2^{N(S-\delta-\epsilon')} && (3.66)\\ &= 2^{-N\epsilon'}, && (3.67) \end{aligned}$$

which vanishes for $N \to 0$. So already the projection will fail in the limit of long messages. This implies that the state $\hat{\rho}_{\boldsymbol{x}}$ after the projection will contain a vanishing component of the original message,

$$\begin{aligned} \hat{\rho}_{\boldsymbol{x}} &= P_{\Lambda'}(\boldsymbol{x})|\phi(\boldsymbol{x})\rangle\langle\phi(\boldsymbol{x})| + (1-P_{\Lambda'}(\boldsymbol{x}))|\phi_0\rangle\phi_0| && (3.68)\\ &= \hat{\Pi}_{\Lambda'}|\boldsymbol{x}\rangle\langle \boldsymbol{x}|\hat{\Pi}_{\Lambda'} + (1-P_{\Lambda'}(\boldsymbol{x}))|\phi_0\rangle\langle\phi_0|. && (3.69) \end{aligned}$$

The fidelity of $\hat{\rho}_{\boldsymbol{x}}$ with $|\boldsymbol{x}\rangle$ will also vanish,

$$\begin{aligned} F(\boldsymbol{x}) &= \langle \boldsymbol{x}|\hat{\rho}_{\boldsymbol{x}}|\boldsymbol{x}\rangle && (3.70)\\ &= P_{\Lambda'}^2(\boldsymbol{x}) + r_{\boldsymbol{x}} && (3.71)\\ &\leq P_{\Lambda'}(\boldsymbol{x}) + r_{\boldsymbol{x}} && (3.72)\\ &= \mathrm{Tr}\{|\boldsymbol{x}\rangle\langle\boldsymbol{x}|\hat{\Pi}_{\Lambda'}\} + r_{\boldsymbol{x}}, && (3.73) \end{aligned}$$

where we used $P_{\Lambda'}(\boldsymbol{x}) \leq 1$ and defined

$$r_{\boldsymbol{x}} := \langle\boldsymbol{x}|\big\{(1 - P_{\Lambda'}(\boldsymbol{x}))|\phi_0\rangle\langle\phi_0|\big\}|\boldsymbol{x}\rangle \geq 0 \tag{3.74}$$

The average fidelity becomes

$$\begin{aligned} F &= \sum_{\boldsymbol{x}\in\mathcal{M}} p(\boldsymbol{x})\, F(\boldsymbol{x}) && (3.75)\\ &\leq \mathrm{Tr}\{\hat{\boldsymbol{\rho}}\hat{\Pi}_{\Lambda'}\} + \sum_{\boldsymbol{x}\in\mathcal{M}} p(\boldsymbol{x})\, r_{\boldsymbol{x}} && (3.76)\\ &= P_{\Lambda'} + r && (3.77)\\ &\leq 2^{-N\epsilon'} + r, && (3.78) \end{aligned}$$

where

$$r := \sum_{\boldsymbol{x}\in\mathcal{M}} p(\boldsymbol{x})\, r_{\boldsymbol{x}}. \tag{3.79}$$

In the limit $N \to \infty$, the average fidelity will approach

$$F \to r, \tag{3.80}$$

which is just the average overlap of the source message ensemble with the garbage state $|\phi_0\rangle$. So even if the coding fails, there is still a chance to accidentally decode the correct message. However, such chance has nothing to do with faithful decoding, because the garbage state does not contain *any* information about the original message. Bob could simply *guess* the correct message with non-zero probability. We can get rid of r by choosing $|\phi_0\rangle$ orthogonal to all source messages.

Concluding, Schumacher derived the quantum analog of Shannon's source coding theorem: A symbol ensemble X of quantum states can be compressed to $S(\hat{\rho})$ qubits in the asymptotic limit of infinitely long messages, where $\hat{\rho}$ is the density matrix corresponding to the ensemble X. Compressing to fewer than $S(\hat{\rho})$ qubits results in the loss of all information in the asymptotic limit.

Chapter 4

Lossless Compression

As Schumacher has showed, *lossy* compression is possible in such a way that in the asymptotic limit of infinitely long messages the losses can be neglected. It is the aim of the following investigations to find statements about *lossless* quantum codes in analogy to the classical case.
The intention of using compressing codes is to minimize the effort of communication between two parties: Alice is preparing source messages $|x\rangle \in \mathcal{V}$ and encodes them into codewords $|c(x)\rangle \in \mathcal{H}^{\oplus r}$ by applying the encoder $\hat{C}$. She compresses the codewords by removing redundant quantum digits and sends the result to Bob, who receives them and decompresses them by appending quantum digits. After that he can decode the messages by applying the decoder $\hat{D}$ and read them or use them as an input for further computations. The communication has been lossless if the decoded message equals the source message. Note that it is not required for Bob to *read* the message he received! In fact, if Bob wants to use the message as an input for a quantum computer, he even *must not* do that, else he will potentially lose information. As we will soon see, it is important that Alice *apriorily knows* which source messages she prepares, otherwise no lossless compression would be possible.

4.1 How not to compress

Let us look for some statements about lossless codes. The first three of the following no-go theorems are also known in classical information theory and are easily transferred to the quantum case by general reasoning. However, we show them by applying the tools developed in this thesis. The last theorem is genuinely quantum with no classical analogue.

No lossless compression by block codes

A code is a *block code* if all codewords have the same length, else it is a *variable-length code*. Unfortunately, lossless block codes do not compress. Take an arbitrary ensemble $X = \{p, \mathcal{X}\}$ with $\mathcal{X} \subset \mathcal{V}$ and any lossless k-ary block code $c : \mathcal{V} \to \mathcal{H}^{\otimes n}$. Let $\mathcal{B}_{\mathcal{V}}$ and $\mathcal{B}_n$ be orthonormal basis sets of $\mathcal{V}$ and $\mathcal{H}^{\otimes n}$, respectively. In order to find for every basis vector $|\omega\rangle \in \mathcal{B}_{\mathcal{V}}$ a code basis vector $|c(\omega)\rangle \in \mathcal{B}_n$, the code must fulfill

$\dim\mathcal{V} \leq \dim\mathcal{H}^{\otimes n} = k^n$. For every $|x\rangle \in \mathcal{X}$, the corresponding codeword $|c(x)\rangle$ has sharp length $L(x) = n$, hence

$$\underline{\mathcal{I}}_c(X) = \log_2 k \sum_{x\in\mathcal{X}} p(x)\, L_c(x) = \log_2 k \cdot n = \log_2(k^n) \tag{4.1}$$

$$\geq \log_2(\dim\mathcal{V}) = \mathcal{I}_0(\mathcal{V}), \tag{4.2}$$

which violates condition (3.10). This implies that there is no lossless compressing block code. By choosing mutually orthogonal source states one can derive the analogue statement for the classical case.

For *long quantum messages* emitted by a memoryless source, block codes can achieve *almost lossless* compression by encoding only *typical subspaces*. The quantum code performing this type of lossy compression is the *Schumacher code* [8] which we have treated above. The only way to compress messages without loss of information is by use of a variable-length code. In order to achieve compression, more frequent messages must be encoded by shorter messages, less frequent messages by longer messages, so that the average length of the code messages is minimized. This is the general rule of lossless data compression.

No lossless compression by changing the alphabet

Trying to achieve compression by using a different alphabet does not work.

A code $c : \mathcal{H}_A^{\otimes n} \to \mathcal{H}_B^{\otimes m}$ that transforms messages over some symbol space $\mathcal{H}_A$ into messages over some symbol space $\mathcal{H}_B$ is lossless only if $\dim\mathcal{H}_A^{\otimes n} \leq \dim\mathcal{H}_B^{\otimes m}$, which implies that

$$\mathcal{I}_0(\mathcal{V}) = n\,\log_2(\dim\mathcal{H}_A) \tag{4.3}$$

$$\leq m\,\log_2(\dim\mathcal{H}_B) = \underline{\mathcal{I}}_c(x), \tag{4.4}$$

for every $|x\rangle \in \mathcal{H}_A$. So for every ensemble $X = \{p, \mathcal{X}\}$ of messages $|x\rangle \in \mathcal{H}_A^m$, we have $\underline{\mathcal{I}}_c(X) = \underline{\mathcal{I}}_c(x) \geq \mathcal{I}_0(\mathcal{V})$, which violates condition (3.10). By choosing mutually orthogonal source states, one can derive the analogue statement for the classical case. This thesis *looks* probably much shorter when written in chinese symbols. However, the effort of communication that is expressed by the effective quantum information I_c, would not be reduced.

No universal lossless compression

We have seen that it is not possible to compress messages without loss of information by using a block code or by using a different symbol space. Now we will see that no code can compress *all* messages without loss of information.

Say we have a space $\mathcal{H}^{\otimes n}$ of *block messages* of fixed length r and we want to compress all of them by use of a variable-length code $c : \mathcal{H}^{\otimes r} \to \mathcal{H}^{\oplus s}$ with $s < r$. The code can only be lossless if

$$\dim\mathcal{H}^{\otimes r} \leq \dim\mathcal{H}^{\oplus s}. \tag{4.5}$$

But since $\dim \mathcal{H}^{\otimes r} = k^r$ and $\dim \mathcal{H}^{\oplus s} = \frac{k^{s+1}-1}{k-1}$, we have

$$k^r \leq \frac{k^{s+1}-1}{k-1} \tag{4.6}$$

$$\Rightarrow \quad k^{r+1} \leq k^{s+1} + k - 1 \tag{4.7}$$

which is wrong for $r \geq s$ and $k > 1$, so we cannot compress all block messages of a given length. Now say we have a space $\mathcal{H}^{\oplus r}$ of *variable-length messages* with maximal length r. Assume that there is a universal lossless code c that reduces the length of all messages in $\mathcal{H}^{\oplus r}$. The code can only be lossless if $\dim \mathcal{H}^{\oplus r} \leq \dim \mathcal{H}^{\oplus s}$, which is obviously wrong for $r > s$, so we cannot compress all variable-length messages with a given maximal length. Concluding, there is no universal lossless compression that reduces the size of all messages. Some messages are unavoidably lengthened by a lossless code. By choosing mutually orthogonal source states, one can derive the analogue statement for the classical case.

No lossless compression of unknown messages

Now we come to a no-compression theorem that is typically quantum. In quantum mechanics there is a profound difference between a *known* and an *unknown* state. For example, a known state can be cloned (by simply preparing another copy of it), whereas an unknown state cannot be cloned.

Theorem 7 (No-compression Theorem) *It is impossible to compress an unknown quantum message without loss of information.*

Proof. Assume that there is a lossless quantum compression algorithm

$$c : \mathcal{H}^{\otimes r} \to \mathcal{H}^{\oplus s} \equiv \bigoplus_{n=0}^{s} \mathcal{H}^{\otimes n} \tag{4.8}$$

that compresses messages of fixed length r to variable-length messages of maximal length s. As shown above, a lossless code cannot compress *all* messages, so $s > r$. Now there is an oracle that hands Alice an arbitrary message $|x\rangle = \sum_{i=1}^{n} x_i |\omega_i\rangle$ where the $|\omega_i\rangle \in \mathcal{H}^{\otimes r}$ are mutually orthogonal states. The algorithm encodes the message $|x\rangle$ into $|c(x)\rangle = \sum_{i=1}^{n} x_i |c(\omega_i)\rangle$. In order for the code c to achieve lossless compression, it must be a variable-length code. Therefore, even if all the codeword components $|c(\omega_i)\rangle$ have determinate length, the total codeword $|c(x)\rangle$ has in general indeterminate length. If Alice wants to remove redundant digits without loss of information, she must know at least an upper bound for the base length of $|c(x)\rangle$, i.e. the length of its longest component. Since $|c(x)\rangle$ is an indeterminate-length message, Alice cannot measure the length without disturbing the state and thereby losing information, so she has to assume the maximal length s. Since $s > r$, no compression is achieved. The same argument applies to quantum compression algorithms $c : \mathcal{H}^{\oplus r} \to \mathcal{H}^{\oplus s}$ that compress variable-length messages of maximal length r to variable-length messages of maximal length s. $\square$

This theorem is not true for the classical case. A classical message is not disturbed by a length measurement, so it *can* in principle be compressed without loss of information. It would be nice to compress a quantum hard disk without loss of information just like a classical hard disk, but as we have seen this cannot be accomplished.

4.2 How to compress

Now that we have found a lot of impossible things to do with quantum messages, it is time to look for the possible things.

4.2.1 Why prefix quantum codes are not very useful

In classical information theory, prefix codes are favored for lossless coding. The reason is that they are *instantaneous*, which means that they carry their own length information (see section 2.6.2). Prefix codewords can be sent or stored without a separating signal between them. The decoder can add word separators ("commas") while reading the sequence from left to right. Whenever a message of letters yields a valid codeword, the decoder can add a comma and proceed. After all, a continuous stream of letters is separated into valid codewords.
Prefix codewords can be separated while *reading* the sequence, but in the quantum case this is potentially a very bad thing to do. Reading a stream of quantum letters means in general *disturbing* the message all the time. Therefore, the length information is generally not available. Furthermore, prefix codewords are in general *longer* than non-prefix codewords, because there are less prefix codewords of a given maximal length than possible codewords. Hence, by using prefix codewords qubits are wasted to encode length information which is unavailable anyway. We conclude that prefix quantum codes are practically not very useful.

4.2.2 A classical side-channel

One could try to encode length information in a different quantum channel, as proposed by *Braunstein et al.* [10] (unnecessarily they used prefix codewords anyhow). But that does not fix the problem. Whatever one does, reading out length information about different components of a variable-length codeword equals a length measurement and hence means disturbing the message. Though there should be *some* way to make sure where the codewords have to be separated, else the message cannot be decoded at all. Here is an idea: Use a *classical side-channel* to inform the receiver where the codewords have to be separated. This has two significant advantages:

- If the length information equals the base length of the codeword, the message is not disturbed and can be losslessly transmitted and decoded.
- Abandoning the prefix condition, shorter codewords can be chosen, such that the quantum channel is used with higher efficiency.

Let us give an example (see Fig. 4.1). Alice wants to send a message $|x_1\rangle$ which is encoded into the codeword $|c(x_1)\rangle = \frac{1}{\sqrt{3}}(|1001101\rangle + |1101\rangle + |10\rangle)$. The base length

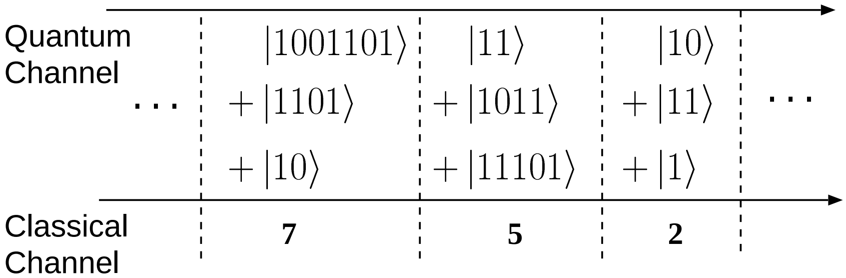


Figure 4.1: Storing length information in a classical side-channel.

of $|c(x_1)\rangle$ is 7, so she submits that information through the classical channel. Dependent on which realization of variable-length messages Alice and Bob have agreed to use, Alice sends enough qubits (at least 7) representing the codeword $|c(x_1)\rangle$ through the quantum channel. The next codeword is $|c(x_2)\rangle = \frac{1}{\sqrt{3}}(|11\rangle + |1011\rangle + |11101\rangle)$. The base length of $|c(x_2)\rangle$ is 5, so Alice sends the length information "5" through the classical channel and enough qubits (at least 5) representing the codeword $|c(x_2)\rangle$ through the quantum channel. She proceeds like that with all following messages. On Bob's side, there is a continuous stream of qubits coming through the quantum channel and a continuous stream of classical bits coming through the classical channel. Bob can read out the classical length information, separate the qubits into the specified blocks and apply the decoder to each codeword. After all, Bob obtains all source messages without loss of information.

4.2.3 Bounds for compression

Lower bound

How much compression can maximally be achieved by using the method sketched in section 4.2.2? Say Alice has an ensemble $X = \{p, \mathcal{X}\}$ of $m = |\mathcal{X}|$ messages $|x_i\rangle \in \mathcal{X}$, $i = 1, \dots, m$ that she wants to encode by k-ary codewords. The source space $\mathcal{V}$ is spanned by the elements of $\mathcal{X}$, thus $\mathcal{V} := \mathrm{Span}(\mathcal{X})$, and has dimension $d := \dim \mathcal{V}$. Alice fixes a basis set $\mathcal{B}_\mathcal{V}$ of d orthonormal vectors $|\omega_i\rangle$, $i = 1, \dots, d$. The ensemble X corresponds to the message matrix

$$\rho := \sum_{i=1}^{m} p(x_i)\, |x_i\rangle\langle x_i| = \sum_{i,j=1}^{d} \rho_{ij}\, |\omega_i\rangle\langle\omega_j|, \tag{4.9}$$

with $\rho_{ij} := \langle\omega_i|\rho|\omega_j\rangle$ and $\sum_{i=1}^{d} \rho_{ii} = 1$. The source messages are encoded by the isometric map $c : \mathcal{V} \to \mathcal{H}^{\oplus}$, defined by

$$|\omega_i\rangle \overset{c}{\longmapsto} |c(\omega_i)\rangle, \quad i = 1, \dots d. \tag{4.10}$$

The code space is k-ary, which means that $k = \dim \mathcal{H}$. Let each codeword $|c(\omega_i)\rangle$ have determinate length $L_c(\omega_i)$, such that the code length operator $\hat{L}_c$ on $\mathcal{V}$ is orthogonal in the basis $\mathcal{B}_\mathcal{V}$ and reads

$$\hat{L}_c = \sum_{i=1}^{d} L_c(\omega_i)\, |\omega_i\rangle\langle\omega_i|. \tag{4.11}$$

The codewords $|c(\omega_i)\rangle$ are not necessarily prefix, because Alice can encode the length information about each codeword in a classical side-channel. In order for the transmission to be lossless, she has to transmit the base length $\underline{L}_c(x_i)$ of each codeword corresponding to the source message $|x_i\rangle$. The base length is at least as long as the expected code length of the codeword, hence

$$\underline{L}_c(x_i) \geq \langle x_i|\hat{L}_c|x_i\rangle. \tag{4.12}$$

Now we are interested in the average base length, since this determines the compression rate. The average base length is bounded from below by

$$\underline{L}_c(X) = \sum_{i=1}^{m} p(x_i)\, \underline{L}_c(x_i) \tag{4.13}$$

$$\geq \sum_{i=1}^{m} p(x_i)\, \langle x_i|\hat{L}_c|x_i\rangle = \mathrm{Tr}\{\rho\, \hat{L}_c\} \tag{4.14}$$

$$= \sum_{i=1}^{m} \rho_{ii}\, L_c(\omega_i). \tag{4.15}$$

Now we perform the following trick. As already stated, non-prefix codewords can be chosen shorter than (or at most as long as) prefix codewords. Consider an arbitrary prefix code c', then

$$L_{c'}(\omega_i) = L_c(\omega_i) + l_{c'}(\omega_i) \geq L_c(\omega_i), \tag{4.16}$$

where $l_{c'}(\omega_i) \geq 0$ is the length difference between the prefix and the non-prefix codeword for $|\omega_i\rangle$. Prefix codes, just like all uniquely decodable symbol codes, have to fulfill the *Kraft inequality* [12, 19]

$$\sum_{i=1}^{d} k^{-L_{c'}(\omega_i)} \leq 1. \tag{4.17}$$

Since the code length operator $\hat{L}_{c'}$ is orthogonal in the basis $\mathcal{B}_\mathcal{V}$, we can express the above condition by the *quantum Kraft inequality*

$$\mathrm{Tr}_\mathcal{V}\{k^{-\hat{L}_{c'}}\} \leq 1, \tag{4.18}$$

where $\hat{L}_{c'} := \hat{L}_c + \hat{l}_{c'}$ and

$$\hat{l}_{c'} := \sum_{i=1}^{d} l_{c'}(\omega_i)\, |\omega_i\rangle\langle\omega_i|. \tag{4.19}$$

The quantum Kraft inequality was derived for the first time by *Schumacher and Westmoreland* [11]. Here, the quantum Kraft inequality requires that

$$Q := \sum_{i=1}^{d} k^{-L_c(\omega_i) - l_{c'}(\omega_i)} \leq 1. \tag{4.20}$$

Now define *implicit probabilities*

$$q(\omega_i) := \frac{1}{Q}\, k^{-L_c(\omega_i) - l_{c'}(\omega_i)}, \tag{4.21}$$

which can be rewritten as

$$L_c(\omega_i) = -\log_k q(\omega_i) - \log_k Q - l'(\omega_i). \tag{4.22}$$

Summing over the ρ_{ii} yields

$$\sum_{i=1}^{d} \rho_{ii}\, L_c(\omega_i) = -\sum_{i=1}^{d} \rho_{ii} \log_k q(\omega_i) - \log_k Q - l', \tag{4.23}$$

where

$$l' := \sum_{i=1}^{d} \rho_{ii}\, l_{c'}(\omega_i) = \mathrm{Tr}\{\rho\, \hat{l}_{c'}\} \tag{4.24}$$

is the average additional length. The inequality (4.15) can now be expressed by

$$\underline{L}_c(X) \geq -\sum_{i=1}^{d} \rho_{ii}\, \log_k q(\omega_i) - \log_k Q - l'. \tag{4.25}$$

Gibbs' inequality (1.119) implies that

$$\underline{L}_c(X) \geq -\sum_{i=1}^{d} \rho_{ii}\, \log_k \rho_{ii} - \log_k Q - l'. \tag{4.26}$$

The von-Neumann entropy of the message matrix ρ cannot decrease by a non-selective projective measurement in the basis $\mathcal{B}_{\mathcal{V}}$, hence

$$S(\hat{\rho}) \leq S(\rho'), \tag{4.27}$$

where

$$\rho' := \sum_{i=1}^{d} |\omega_i\rangle\langle\omega_i|\rho|\omega_i\rangle\langle\omega_i| = \sum_{i=1}^{d} \rho_{ii}|\omega_i\rangle\langle\omega_i|. \tag{4.28}$$

Since

$$S(\rho') = -\sum_{i=1}^{d} \rho_{ii}\, \log_2 \rho_{ii} = -\log_2 k\, \sum_{i=1}^{d} \rho_{ii}\, \log_k \rho_{ii}, \tag{4.29}$$

relation (4.27) states that

$$-\sum_{i=1}^{d} \rho_{ii}\, \log_k \rho_{ii} \geq \frac{1}{\log_2 k}\, S(\hat{\rho}). \tag{4.30}$$

Using (4.30) together with the Kraft inequality $Q \leq 1$, relation (4.26) transforms into

$$\log_2 k \cdot \left\{\underline{L}_c(X) + l'\right\} \geq S(\hat{\rho}) - \log_k Q \geq S(\hat{\rho}). \tag{4.31}$$

Recalling the definition of the effective quantum information (3.8) and defining the length information that can be drawn into the classical side-channel by

$$I' := \log_2 k \cdot l', \tag{4.32}$$

we finally arrive at the lower bound relation

$$\underline{\mathcal{I}}_c(X) + I' \geq S(\hat{\rho}). \tag{4.33}$$

If c is a uniquely decodable symbol code, e.g. a prefix code, we have $I' = 0$. Inequality (4.33) states that the ensemble X can be losslessly compressed not below $S(\hat{\rho})$ qubits. However, by drawing length information into a classical side-channel it is possible to reduce the average number of qubits passing through the quantum channel *below* the von-Neumann entropy. We will give an example later on where this really happens.

Upper bound

Let us look for an upper bound for the compression that can be achieved. In order to encode every source vector in $\mathcal{V}$ by a k-ary code, we need at most

$$\underline{L}_c(x) \leq \lceil \log_k(\dim \mathcal{V}) \rceil \leq \log_k(\dim \mathcal{V}) + 1 \tag{4.34}$$

digits. Using $\log_a x = \log_a b \cdot \log_b x$, we have

$$\underline{\mathcal{I}}_c(X) \leq \log_2(\dim \mathcal{V}) + \log_2 k. \tag{4.35}$$

This upper bound is neither very tight nor is it related to the von-Neumann entropy. However, our efforts to find a more interesting upper bound were not successful. It remains an open question to find such a bound and hence a quantum mechanical generalization to Shannon's theorem [9],

$$H(X) \leq \mathcal{I}_c(X) \leq H(X) + \log_2 k, \tag{4.36}$$

which looks more familiar for $k = 2$, such that $\log_2 k = 1$ and $\mathcal{I}_c(X) = L_c(X)$.

4.2.4 Quantum Morse codes

One way to avoid a classical side-channel is to leave a *pause* between the quantum codewords, which equals an additional orthogonal "separator state". Such a code is a quantum analogue to the *Morse code*, where the codewords are also separated by a pause, in order to avoid prefix codewords. Of course, the codewords *plus* the pause are prefix. Due to the close analogy one could speak of *quantum Morse codes*. Here, the information I' needed for the separator state is independent from the statistics, because the separator state must be sent after each letter codeword, no matter which one. In contrast to that, I' is in general dependent from the statistics. If one transmits the length of each codeword through a classical side-channel, one can use a Huffman code to find shorter codewords for more frequent length values. Such is done in the following compression scheme.

4.3 A lossless compression scheme

Let us construct an explicit coding scheme that realizes lossless quantum compression.

4.3.1 Basic idea

Alice and Bob have a quantum computer on both sides of the channel. They both allocate a register of r k-ary quantum digits, whose physical space is given by $\mathcal{R} = \mathcal{D}^{\otimes r}$ with $\mathcal{D} = \mathbb{C}^k$. They agree to use neutral-prefix codewords (see section 2.6.2) to implement variable-length coding, hence the message space is $\mathcal{N}_r$ of dimension k^r and is physically realized by the operational space $\tilde{\mathcal{N}}_r = \mathcal{R}$. Alice is preparing source messages $|x_i\rangle, i = 1, \ldots, m$ from a set $\mathcal{X}$. The space spanned by these messages is the source space $\mathcal{V} = \mathrm{Span}(\mathcal{X})$. Alice prepares each message $|x\rangle \in \mathcal{X}$ with probability $p(x)$, which gives the ensemble $X := \{p, \mathcal{X}\}$. She encodes the source messages into variable-length codewords $|c(x)\rangle \in \mathcal{N}_r$ of maximal length r. If the dimension of $\mathcal{V}$ is given by $d := \dim \mathcal{V}$, then the length of the register must fulfill

$$r \geq \lceil \log_k d \rceil. \tag{4.37}$$

If the set $\mathcal{X}$ is linearly dependent, Alice creates a set $\tilde{\mathcal{X}} = \mathcal{X}$, removes the most probable message from $\tilde{\mathcal{X}}$ and puts it into a list $\boldsymbol{M}$. Next, she removes again the most probable message from $\tilde{\mathcal{X}}$, appends it to the list $\boldsymbol{M}$ and checks if the list is now linearly dependent. If so, she removes the last element from $\boldsymbol{M}$ again. Then she proceeds with removing the next probable message from $\tilde{\mathcal{X}}$ and appending it to $\boldsymbol{M}$, checking for linearly dependence, and so on. In the end she obtains a list

$$\boldsymbol{M} = (|x_1\rangle, \ldots, |x_d\rangle) \tag{4.38}$$

of linearly independent source messages from $\mathcal{X}$, ordered by decreasing probability, such that $p(x_i) \geq p(x_j)$ for $i \leq j$. She performs a *Gram-Schmidt* orthononormalization on the list $\boldsymbol{M}$, giving a list $\boldsymbol{B}$ of orthornormal vectors $|\omega_i\rangle$, defined by

$$|\omega_1\rangle := |x_1\rangle, \tag{4.39}$$

$$|\omega_i\rangle := N_i \Big[\mathbb{1} - \sum_{j=1}^{i-1} |\omega_j\rangle\langle\omega_j|\Big] |x_i\rangle, \tag{4.40}$$

with $i = 2, \ldots, d$ and suitable normalization constants N_i. The elements of $\boldsymbol{B}$ form an orthonormal basis $\mathcal{B}_{\mathcal{V}}$ for the source space $\mathcal{V}$. Now she assigns codewords

$$|c(\omega_i)\rangle := |Z_k^r(i-1)\rangle, \quad i = 1, \ldots, d. \tag{4.41}$$

of increasing significant length

$$L_c(\omega_i) = \lceil \log_k(i) \rceil. \tag{4.42}$$

Note that the first codeword is the empty message $|\square\rangle = |Z_k^r(0)\rangle = |0 \cdots 0\rangle$, which does not have to be sent through the quantum channel at all. Instead, nothing is sent through the quantum channel and a signal representing "length 0" is sent through the classical channel. Alice implements the encoder

$$C := \sum_{i=1}^{d} |c(\omega_i)\rangle\langle\omega_i|, \tag{4.43}$$

by a gate array on $\mathcal{R}$. Then she calculates the base lengths of the codewords,

$$\underline{L}_c(x) = \max_{i=1,\ldots,d}\{L_c(\omega_i) \mid |\langle\omega_i|x\rangle|^2 > 0\}, \tag{4.44}$$

for every message $|x\rangle \in \mathcal{X}$ and writes them into a table. The classical information is compressed using Huffman coding of the set of distinct base length values $\mathcal{L} = \{L_c(\omega_1), \ldots, L_c(\omega_d)\}$. Alice constructs the Huffman codeword to each length $l \in \mathcal{L}$ appearing with probability

$$p_l = \sum_{x:\,\underline{L}_c(x)=l} p(x), \tag{4.45}$$

and writes them into a table. At last, Alice builds a gate array realizing the decoder $D = C^{-1}$ and gives it to Bob. For the classical channel she hands the table with the Huffman codewords for the distinct lengths to Bob. Now everything is prepared and the communication can begin.

4.3.2 Communication protocol

Alice prepares the message $|x\rangle \in \mathcal{X}$ and applies the encoder C to obtain $|c(x)\rangle$. She looks up the corresponding code base length $\underline{L}_c(x)$ in the table. If $\underline{L}_c(x) < r$, she truncates the message to $\underline{L}_c(x)$ digits by removing $r - \underline{L}_c(x)$ leading digits. She sends the $\underline{L}_c(x)$ digits through the quantum channel and the length information $\underline{L}_c(x)$ through the classical channel. Then she proceeds with the next message.

For any message $|x\rangle$ Alice sends, Bob receives the length information $\underline{L}_c(x)$ through the classical channel and $\underline{L}_c(x)$ digits through the quantum channel. He adds $r - \underline{L}_c(x)$ quantum digits in the state $|0\rangle$ at the beginning of the received codeword. He then applies the decoder D and obtains the original message $|x\rangle$ with perfect fidelity. Note that Alice can send *any* message from the source message space $\mathcal{V}$, the protocol will ensure a lossless communication of the message. For such arbitrary messages, however, compression will in general not be achieved, since the protocol is only adapted to the particular ensemble X. Also, Bob can as well store all received quantum digits on his quantum hard disk and the received length information on his classical hard disk, and go to bed. The next day, he can scan the classical hard disk for length information and separate and decode the corresponding codewords on the quantum hard disk. The protocol works as well for online communication as for data storage.

4.3.3 An explicit example

Alice and Bob want to communicate vectors of a 4-dimensional Hilbert space $\mathcal{V} = \mathrm{Span}\{|0\rangle, |1\rangle, |2\rangle, |3\rangle\}$, where we use the row notation in the following. Alice decides to use the (linearly dependent) source message set

$$\mathcal{X} = \{|a\rangle, |b\rangle, |c\rangle, |d\rangle, |e\rangle, |f\rangle, |g\rangle, |h\rangle, |i\rangle, |j\rangle\}, \tag{4.46}$$

whose elements are given by

$$
\begin{aligned}
|a\rangle &= \frac{1}{2}(1,1,1,1), \quad |b\rangle = \frac{1}{\sqrt{5}}(1,2,1,1) \\
|c\rangle &= \frac{1}{\sqrt{6}}(1,3,1,1), \quad |d\rangle = \frac{1}{\sqrt{7}}(1,4,1,1) \\
|e\rangle &= \frac{1}{\sqrt{2}}(1,0,1,0), \quad |f\rangle = \frac{1}{\sqrt{3}}(2,0,1,0) \\
|g\rangle &= \frac{1}{2}(3,0,1,0), \quad |h\rangle = \frac{1}{\sqrt{2}}(0,1,0,1) \\
|i\rangle &= \frac{1}{\sqrt{3}}(0,2,0,1), \quad |j\rangle = \frac{1}{2}(0,3,0,1)
\end{aligned} \tag{4.47}
$$

and which are used with the probabilities

$$
\begin{aligned}
p(a) &= 0.6, \quad p(b) = p(c) = p(d) = 0.1, \\
p(e) &= \ldots = p(j) = \frac{0.3}{3}.
\end{aligned} \tag{4.48}
$$

The Shannon entropy of the ensemble $X = \{p, \mathcal{X}\}$ is

$$H(X) = 2.02945, \tag{4.49}$$

and the classical raw information (1.19) reads

$$\mathcal{I}_0(\mathcal{X}) = \log_2 |\mathcal{X}| = 3.32193, \tag{4.50}$$

which gives an optimal classical compression rate of $R = H/\mathcal{I}_0 = 0.610924$. If Bob knows Alice's list of possible messages, then this rate could in the optimal case be achieved by pure classical communication. However, Bob does not know the list and classical communication is not the task here. The message matrix $\rho = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|$, given by

$$\rho = \begin{pmatrix} 0.214549 & 0.224624 & 0.197882 & 0.177882 \\ 0.224624 & 0.40302 & 0.224624 & 0.244624 \\ 0.197882 & 0.224624 & 0.191216 & 0.177882 \\ 0.177882 & 0.244624 & 0.177882 & 0.191216 \end{pmatrix} \tag{4.51}$$

has von-Neumann entropy

$$S(\hat{\rho}) = 0.571241. \tag{4.52}$$

The orthogonalization procedure yields the basis $\mathcal{B}_\mathcal{V} = \{|\omega_i\rangle\}$ with

$$
\begin{aligned}
|\omega_1\rangle &= (0.5, 0.5, 0.5, 0.5) \\
|\omega_2\rangle &= (-0.288675, 0.866025, -0.288675, -0.288675) \\
|\omega_3\rangle &= (0.408248, 0, 0.408248, -0.816497) \\
|\omega_4\rangle &= (0.707107, 0, -0.707107, 0).
\end{aligned} \tag{4.53}
$$

Let the quantum channel be binary, i.e. let $k = 2$. The codewords are constructed along $|c(\omega_i)\rangle = |Z_2(i-1)\rangle$, yielding the variable-length states

$$
\begin{aligned}
|c(\omega_1)\rangle &= |\Box\rangle, \quad |c(\omega_2)\rangle = |1\rangle \\
|c(\omega_3)\rangle &= |10\rangle, \quad |c(\omega_4)\rangle = |11\rangle,
\end{aligned} \tag{4.54}
$$

that span the code space $\mathcal{C}$. In a neutral-prefix code they are realized by the 2-qubit states

$$\begin{aligned} |\tilde{c}(\omega_1)\rangle &= |00\rangle, \quad |\tilde{c}(\omega_2)\rangle = |01\rangle \\ |\tilde{c}(\omega_3)\rangle &= |10\rangle, \quad |\tilde{c}(\omega_4)\rangle = |11\rangle \end{aligned} \tag{4.55}$$

that span the operational code space $\tilde{\mathcal{C}}$, which is a subspace of the physical space $\mathcal{R} = \mathbb{C}^2 \otimes \mathbb{C}^2$. Alice realizes the encoder $C : \mathcal{V} \to \tilde{\mathcal{C}}$, $C = \sum_i |\tilde{c}(\omega_i)\rangle\langle\omega_i|$, given by

$$C = \begin{pmatrix} 0.5 & 0.5 & 0.5 & 0.5 \\ -0.288675 & 0.866025 & -0.288675 & -0.288675 \\ 0.408248 & 0 & 0.408248 & -0.816497 \\ 0.707107 & 0 & -0.707107 & 0 \end{pmatrix} \tag{4.56}$$

and the decoder $D = C^{-1}$, given by

$$D = \begin{pmatrix} 0.5 & 0.408248 & -0.288675 & 0.707107 \\ 0.5 & 0 & 0.866025 & 0 \\ 0.5 & 0.408248 & -0.288675 & -0.707107 \\ 0.5 & -0.816497 & -0.288675 & 0 \end{pmatrix} \tag{4.57}$$

by gate arrays and gives the decoder to Bob. The encoded alphabet is obtained by $|c(x)\rangle = C|x\rangle$. Alice writes the base lengths of the codewords

$$\begin{aligned} &\underline{L}_c(a) = 0,\ \underline{L}_c(b) = \underline{L}_c(c) = \underline{L}_c(d) = 1, \\ &\underline{L}_c(e) = \ldots = \underline{L}_c(j) = 2 \end{aligned} \tag{4.58}$$

in a table and calculates the corresponding probabilities

$$p_0 = 0.6, \quad p_1 = 0.3, \quad p_2 = 0.1 \tag{4.59}$$

She constructs Huffman codewords for each length

$$c_0 = 1, \quad c_1 = 01, \quad c_2 = 00, \tag{4.60}$$

such that the average bit length is

$$L' = \sum_{l=0}^{2} p_l \cdot l = 1.4, \tag{4.61}$$

which is the optimal value next to the Shannon entropy of the length ensemble

$$I' = -\sum_{l=0}^{2} p_l \log_2 p_l = 1.29546\ . \tag{4.62}$$

Alice hands the table with the Huffman codewords to Bob and tells him that he must listen to the classical channel, decode the arriving Huffman codewords into numbers, receive packages of qubits, whose size corresponds to the decoded numbers, and add to each package enough leading qubits in the state $|0\rangle$ to end up with 2 qubits. Then he must apply the decoder D to each extended package and he will get Alice's original messages.

Say, Alice wants to send the message $|a\rangle$. She prepares $|a\rangle$ and applies the encoder C to obtain the codeword $|00\rangle$. She looks up the corresponding base length $\underline{L}_c(a) = 0$ and truncates the codeword to $\underline{L}_c(a) = 0$ qubits. In this case there are no qubits left at all, so she sends nothing through the quantum channel and the Huffman codeword for "length 0" through the classical channel. Bob receives the classical length information "0" and knows that nothing comes through the quantum channel and that in this case he has to prepare 2 qubits in the state $|00\rangle$. He applies the decoder D and obtains Alice's original message $|a\rangle$. In order to send message $|b\rangle$, Alice truncates the codeword to $\underline{L}_c(b) = 1$ qubit and obtains $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. She sends the qubit through the quantum channel together with the classical signal "length 1". Bob receives the length message and knows that he has to take the next qubit from the quantum channel and that he has to add 1 leading qubit in the state $|0\rangle$. He applies D and obtains Alice's original message $|b\rangle$. The whole procedure works instantaneous and without loss of information. Timo Felbinger and me have implemented the above example by a Mathematica™ program and numerical simulations show that the procedure works fine and the specified compression of quantum data is achieved. (You can find the program and the package at [29]).

Let us look for the compression that has been achieved. The quantum effective quantum information, i.e. the average number of qubits being sent through the quantum channel,

$$\underline{\mathcal{I}}_c = \sum_{x\in\mathcal{X}} p(x)\underline{L}_c(x) = 0.5, \tag{4.63}$$

falls below(!) the von-Neumann entropy:

$$\underline{\mathcal{I}}_c < S = 0.571241. \tag{4.64}$$

Such a behaviour has already been suspected in section 4.2.3. The quantum raw information, i.e. the size of the non-compressed messages, is given by

$$\underline{\mathcal{I}}_c < \mathcal{I}_0 = \log_2(\dim \mathcal{V}) = 2, \tag{4.65}$$

hence the compression rate on the quantum channel reads

$$\mathrm{R}_c = \frac{\underline{\mathcal{I}}_c}{\mathcal{I}_0} = 0.25. \tag{4.66}$$

In other words, the number of qubits passing through the quantum channel is reduced by 75 %. Sending 100 messages without compression requires 200 qubits. Using the compression scheme, Alice typically sends 50 qubits. The sum of both quantum and classical information,

$$I_{\text{tot}} = \underline{\mathcal{I}}_c + I' = 1.79546, \tag{4.67}$$

is smaller than the Shannon entropy (4.49) of the original ensemble X,

$$I_{\text{tot}} < H = 2.02945, \tag{4.68}$$

but greater than the von-Neumann entropy (4.52),

$$I_{\text{tot}} > S = 0.571241. \tag{4.69}$$

The classical part of the compression depends on the algorithm. Only in the ideal case the information can be compressed down to the Shannon entropy of the length ensemble, given by I'. Using the Huffman scheme, the average length $L' = 1.4$ represents the information that is effectively sent through the classical channel, such that the total *effective* information is given by

$$I_{\text{eff}} = \underline{\mathcal{I}}_c + L' = 1.9. \tag{4.70}$$

The the total compression rate of both channels reads

$$R_{\text{tot}} = \frac{\underline{\mathcal{I}}_c + I'}{\mathcal{I}_0} = 0.897731 < 1, \tag{4.71}$$

where it is assumed that the information on the classical channel can be compressed down to its Shannon entropy I'. Using the Huffman scheme (as we have done in our example), the information on the classical channel can only be compressed to $L' > I'$, such that the *effective* total compression rate is given by

$$R_{\text{eff}} = \frac{\underline{\mathcal{I}}_c + L'}{\mathcal{I}_0} = 0.95 < 1. \tag{4.72}$$

Thus in any case there is an overall compression. For higher dimensional source spaces (hence more letters), the compression is expected to get better (provided the letter distribution is not too uniform). However, the numerical effort for higher dimensional letter spaces increases very fast and we want to keep the example as simple as possible.

Part III

Cryptography

Chapter 5

Classical Cryptography

Cryptography is the art of secret communication. A message must be brought from A to B so that no one else can read it. This task is raised since ages, whenever people have reason to distrust each other, mostly in a situation of war, of conspiracy, of business, but also in a private context, in case of criminal action, forbidden love, secret friendship, or simply in case that personal letters and notes ought to stay truly personal.
The generic situation envolves three parties: Alice, the sender, Bob, the authorized receiver, and Eve, the evil enemy. What can Alice do to send Bob a message without letting Eve read it? Whatever mechanism Alice uses, it will amount to *hiding* the message from Eve while sending it to Bob. This process of message hiding is called *encryption* and the result is called a *cryptogram*. The instructions which uniquely define the encryption process represent the *key*, and we will further specify it as the *encoding key*. After encryption Alice sends the resulting cryptogram to Bob. But now she faces a problem: How should Bob *recover* the message? Bob is effectively in the same situation as Eve, because both are potential receivers of the message, hence there must be some *distinctive feature* that enables Bob to read the message and not Eve. This distinctive feature is the posession of the instructions that uniqely define the decryption process. These instructions represent another *key* and we will call it the *decoding key*. In order to provide a successful transfer, the encoding and decoding key must fit together, so that the decryption process reverts the encryption process. The entire transfer very much resembles the situation of putting the message in a box (encryption), sending the box to Eve (transmission), and getting the message out of the box (decryption), where the box can only be opened by use of a key. It is evidently necessary that the decoding key is only available to the authorized receiver, Bob. Whenever Eve also manages to get this key, she can slip into the role of Bob and recover the message.
There are several methods of hiding the message from the enemy. These methods are called *cryptographic systems* or in short *cryptosystems*. It was again Shannon who made a major contribution to the systematic investigation of secret communication [30]. He distinguished three kinds of cryptographic systems: 1) *Concealment systems*, where the mere existence of the message is concealed, e.g. by use of invisible ink or by writing the message on a shaved head or into a microdot. This type of cryptography is also called *steganography*, and it remains an important cryptographic method, especially in times of digital communication where one can hide a message in the pixels of a harmless picture published on the internet. 2) *Privacy systems*, where special equipment is required to

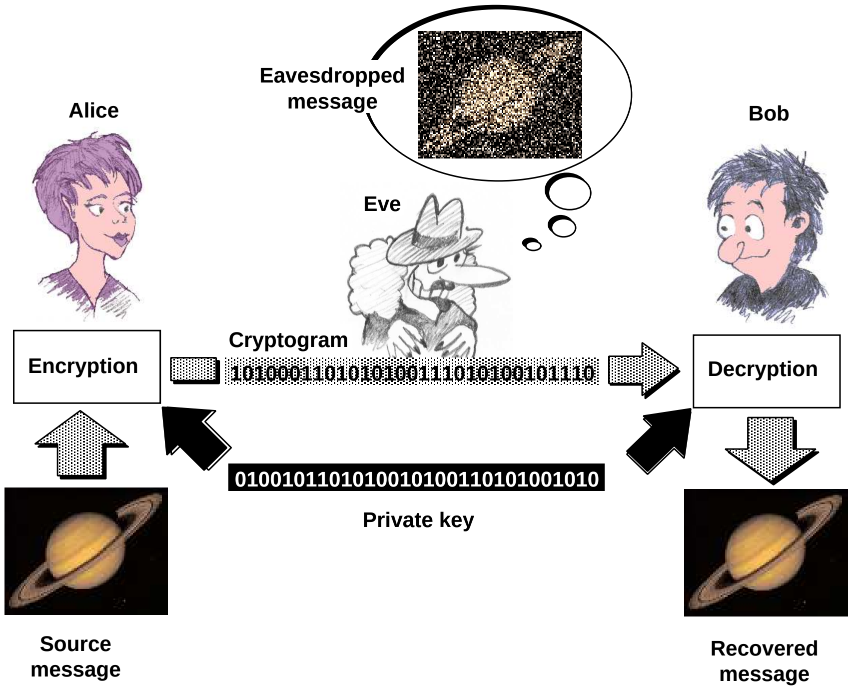


Figure 5.1: General scheme for a private-key cryptosystem. Alice encodes her message into a cryptogram by use of a private key. She sends the cryptogram to Bob over a channel which is attacked by Eve. Bob decodes the message by use of the same key. Without the key Eve can gain very few information. If the cryptosystem is perfectly secure then Eve gain gain no information at all. However, the problem is how to establish a shared secret key. . . ?

recover the message, e.g. speech inversion or hidden tracks on a vinyl record. 3) *Secrecy systems*, where the meaning of the message is concealed by a special code.

We see that all three types of cryptosystems follow the same strategy of encryption, transmission, and decryption. In the case of concealment systems, the encoding key is the set of instructions for concealing the message (e.g. "shave the head of the messenger and write the message on the skin"), and the decoding key is the set of instructions to revert this procedure ("shave the messenger's head and read the message"). In the case of privacy systems, the encryption is a technical procedure of converting the message into a signal which can only be reverted into the original message by use of special technical equipment. The instructions how to build the equipment for encryption and decryption represent the encoding and decoding key, respectively. In the case of secrecy systems the encryption is a mathematical function on the message message, the decryption is the inverse of this function and the encoding and decoding key is the parameter needed to identify the enrypting and decrypting function, respectively.

It is the latter of these three types of cryptosystems, the *secrecy systems*, that is actually addressed in communication theory, because it is of a purely *mathematical* nature and is therefore preferred by mathematicians. The other two types involve *physics*. However, the advent of *quantum cryptography* has brought physics back to the focus of modern cryptography. It is special equipment which is needed to execute a quantum cryptographic protocol: One needs a pulsed laser, a high-quality glass fibre and a sensible detector. While the classical cryptographic protocols are secure for *mathematical* reasons, the quantum cryptographic protocols are secure for *physical* reasons. They make use of the fact that it is impossible to gain information from a quantum system without disturbing it. This disturbance can be detected: The observer becomes observable. In a cryptographic context this typically quantum feature becomes useful, since the observer is now the enemy who wants to eavesdrop the secret as it is passing from A to B. By extracting information he disturbs the state of the channel, therefore he modifies the signal and becomes detectable. As soon as the eavesdropping attack is detected, the communication is stopped and since the signal only carries information about a random key later to be used for encryption, the actual message is saved from the enemy's ruthless eyes.

5.1 Private-key cryptosystems

Consider our definition of secret communication in the previous section. If the encoding key is the same as the decoding key then we have a *private-key cryptosystem*. A private-key cryptosystem is given by a family $\mathcal{F}_\mathcal{K} = \{f_k \mid k \in \mathcal{K}\}$ of encrypting functions f_k parametrized by keys $k \in \mathcal{K}$. Sender and receiver agree upon some private key $k \in \mathcal{K}$ and a message $m \in \mathcal{M}$ is encrypted by

$$c = f_k(m). \tag{5.1}$$

The message is decrypted by use of a function g_k, such that

$$g_k(f_k(m)) = m \tag{5.2}$$

for all $m \in \mathcal{M}$. Obviously, g_k is the inverse of f_k on the set

$$\mathcal{C}_k := f_k(\mathcal{M}) \tag{5.3}$$

of all cryptograms that can be obtained by use of the key k. The set of all possible cryptograms is given by

$$\mathcal{C} = \bigcup_{k\in\mathcal{K}} \mathcal{C}_k. \tag{5.4}$$

5.1.1 Perfect security

A cryptosystem is called *perfectly secure* or *unconditionally secure* or *information-theoretically secure* if the knowledge available to the eavesdropper does not reveal any information about the original message. Let us precise this for the case of a private-key cryptosystem. Let $p(m)$ be the apriori probability of the message $m \in \mathcal{M}$ and let λ_k be the apriori probability of the key $k \in \mathcal{K}$. Then the total probability of all keys transforming m into c represents the conditional probability that the message m is encoded into the cryptogram c,

$$p(c|m) = \sum_{k:f_k(m)=c} \lambda_k. \tag{5.5}$$

Perfect security means that without knowing the key it is impossible to deduce the original message, which is equivalent to the requirement that

$$\forall\, m \in \mathcal{M}, c \in \mathcal{C}: \quad p(c|m) = q(c), \tag{5.6}$$

where

$$q(c) = \sum_m p(c|m)p(m) \tag{5.7}$$

is the probability of obtaining the cryptogram c. Stated otherwise, a necessary and sufficient condition for perfect security is that

$$\sum_{k:f_k(m)=c} \lambda_k \quad \text{should be independent from } m. \tag{5.8}$$

We can find still other defining conditions for perfect security. The aposteriori probability of the message m for a given cryptogram c reads

$$q(m|c) = \frac{p(c|m)p(m)}{\sum_m p(c|m)p(m)}, \tag{5.9}$$

from where we infer that perfect security requires that the aposteriori probability of a message m should coincide with its apriori probability,

$$q(m|c) \stackrel{!}{=} p(m) \tag{5.10}$$

for all $m \in \mathcal{M}$ and $c \in \mathcal{C}$. The joint probability of choosing m and obtaining c is given by

$$p(m, c) = p(c|m)p(m) = q(m|c)q(c), \tag{5.11}$$

so (5.8) or (5.10) are equivalent to

$$p(m, c) \stackrel{!}{=} p(m)q(c), \tag{5.12}$$

thus perfect security means that the messages and the cryptograms are *statistically independent* from each other. Consequently, by learning the cryptogram one learns nothing about the original message, so the *mutual information* of the message ensemble $M = \{(m, p(m)), m \in \mathcal{M}\}$ and the cryptogram ensemble $C = \{(c, q(c)), c \in \mathcal{C}\}$ is zero,

$$I(M\!:\!C) \stackrel{!}{=} 0. \tag{5.13}$$

Conditions (5.6, 5.8, 5.10, 5.12, 5.13) are all equivalent for perfect security.
For a fixed key k the encrypting function f_k gives a one-to-one correspondence between all messages $m \in \mathcal{M}$ and their encryptions $c \in \mathcal{C}_k$, so $|\mathcal{M}| = |\mathcal{C}_k|$ for all $k \in \mathcal{K}$. Since $\mathcal{C} = \bigcup_k \mathcal{C}_k$ we have

$$|\mathcal{M}| \leq |\mathcal{C}|. \tag{5.14}$$

Each cryptogram $c \in \mathcal{C}$ fulfills $q(c) \neq 0$, because by definition of $\mathcal{C}$ there is at least one key that produces this cryptogram from a message. Perfect security requires that $p(c|m) = q(c) \neq 0$ for any $m \in \mathcal{M}$. Hence for any pair (m, c) there must be at least one key $k \in \mathcal{K}$ transforming m into c. Now fix m, then there are $|\mathcal{C}|$ different pairs (m, c)

and for each such pair there must be at least one different key, thus we have $|\mathcal{C}| \leq |\mathcal{K}|$. Using (5.14) we find that there are at least as many keys as there are messages,

$$|\mathcal{M}| \leq |\mathcal{K}|. \tag{5.15}$$

This is an important conclusion found by Shannon [30] which has the consequence that in case that the messages and keys are messages over the same alphabet one finds:

> *The key needed for a perfectly secure transmission of a message is at least as long as the message to be transmitted.*

Perfect security can already be obtained with $|\mathcal{M}| = |\mathcal{K}|$ as the following example shows. Let there be n distinct messages m_i and cryptograms c_j and keys k, then encrypt each message with the key k by

$$f_k(m_i) = c_j, \quad j = i + k \,(\mathrm{mod}\, n). \tag{5.16}$$

We see that $p(c|m) = \frac{1}{n} = q(c)$, thus we have perfect security.

5.1.2 The Vernam cipher

There is a perfectly secure private-key cryptosystem, the *Vernam cipher* or *one-time pad*. Alice and Bob agree upon a K-ary alphabet $\mathcal{A} = \{0, 1, 2, \ldots, K-1\}$ and a private key $k \in \mathcal{A}^N$ of length N. Alice encodes her message $m \in \mathcal{A}^N$ by adding the key k modulo K:

$$f_k(x_1 \cdots x_N) = f_{k_1}(x_1) \cdots f_{k_N}(x_N), \tag{5.17}$$

where

$$f_{k_n}(x_n) = x_n + k_n \quad (\mathrm{mod}\, K). \tag{5.18}$$

Bob decrypts the message by subtracting the key letter by letter

$$g_{k_n}(c_n) = c_n - k_n \quad (\mathrm{mod}\, K). \tag{5.19}$$

This method is for each single letter equivalent to Shannon's encryption method mentioned in the previous section. The conditional probability is $p(c_n|x_n) = \frac{1}{K} = q(c_n)$, thus the scheme is perfectly secure. The Vernam cipher is also called *one-time pad* because the key can only be used one time. This is typical for private-key cryptosystems. If the key would be in use multiple times then the method would no longer be perfectly secure, because statistical correlations would appear that make it possible to estimate the source message up to a certain fidelity.

5.1.3 The key distribution problem

The problem with private-key cryptosystems is: How should Alice and Bob agree upon the private key? This is the so-called *key distribution problem*. If Alice and Bob are at separate locations they must find some procedure to transmit the private key from one location to another in a secure manner. This again would require another private key which would in turn have to be securely transmitted and so on. If the key needed for each secure transmission would be shorter than the message to be transmitted, the entire procedure would come to an end and so there would effectively be a finite procedure providing unconditionally secure communication from one party to another. This cannot be the case because Eve can at each step eavesdrop the cryptogram and therefore receive the same information as Bob. In the end Alice would have securely communicated a message to Bob *and* to Eve, which is a contradiciton. This also shows that the length of the key in a perfectly secure cryptosystem cannot be shorter than the length of the message.

As the key distribution problem makes clear, the problem of secure communication is effectively shifted from communicating the original message to communicating the key. If there would be no difference between a message and a key, we could forget about secure communication. There is, however, an important difference: The message is *willingly chosen* while the key may be *random*. In other words: The message contains *meaningful* information, and the key does not. The distinction between "meaningful" and "not meaningful" is difficult to formalize. Let us agree upon the following: A message is meaningful if the sender composes the message by his own will, and it is not meaningful if he composes it by use of a random generator. For an external person both messages appear random, because the will of the sender is absolutely private, and therefore both messages are unpredictable. We understand "random" in the sense of "unpredictable", thus both the message and the key appear random to an external person but not to the sender. This distinctive feature between sender and external person can be exploited in practice: Sender and receiver privately agree upon the key *in advance* (at a secret meeting) and then use this key later when communicating the actual message.

Quantum cryptography offers another solution to the key distribution problem: Sender and receiver generate a shared secret key by use of a quantum communication protocol which is perfectly secure against eavesdropping in the limit of infinitely long keys. Such a protocol is the BB84 which we will discuss below.

5.2 Public-key cryptosystems

Consider once more our definition of secure communication in section 5. If the encoding key is *different* from the decoding key, then we have a *public-key cryptosystem*. A public-key cryptosystem is given by a family $\mathcal{F}_{\mathcal{K}}$ of encrypting functions and a family $\mathcal{G}_{\mathcal{S}}$ of decrypting functions, such that for each encoding key $k \in \mathcal{K}$ there is at least one decoding key $s \in \mathcal{S}$ so that the message is encrypted by $f_k \in \mathcal{F}_{\mathcal{K}}$ and is decrypted by $g_s \in \mathcal{G}_{\mathcal{S}}$,

$$g_s(f_k(m)) = m \tag{5.20}$$

for all $m \in \mathcal{M}$. The *secret key* s is known to the authorized receiver, and the *public key* k is known to everybody, including the sender. The communication protocol goes as follows: Bob generates a pair (k, s) of matching keys, he transmits the public key k by use of a *public channel* to Alice, Alice uses this key to encrypt her message, she transmits the encrypted message to Bob who then decrypts it by use of the secret key s. The "public channel" is a channel that can only be modified by an authorized person, and it is an important element of the protocol. If Eve (who is definitely *not* authorized) would have control over the public channel, then she could replace Bob's public key by a key of her own choice. Alice would use this key to encrypt her message which then can seemlessly be decrypted by Eve. Such an attack is called a *man-in-the-middle attack*, because Eve slips into the role of an authorized person, i.e. either Alice or Bob.

5.2.1 Computational security

The drawback of public-key cryptosystems is that they are not perfectly secure, because the public key can in principle be used to decrypt the message. Perfect security would imply that Eve's knowledge, i.e. the knowledge about the cryptosystem, the cryptogram, and the public key, does not reveal any information about the original message. However, g_s is obviously the inverse of f_k on the set $\mathcal{C}_k = f_k(\mathcal{M})$. Since Eve knows the public key k she also knows the function f_k, so she can determine the set $\mathcal{C}_k$, and since the inverse of f_k on $\mathcal{C}_k$ is unique, it is in principle possible for Eve to determine this inverse, so she can decrypt any message. Eve can in principle even determine all possible secret keys s for any given k, because all of them must give the inverse function g_s of f_k on the range $\mathcal{C}_k$, so by brute force she could try them all out and write the matching keys down in a table. Whenever Bob announces the public key, Eve chooses a matching secret key from her table and decrypts the message. This is not at all what we would understand under "secure communication", so what is the public-key cryptosystem actually good for?
The answer is *computational security*. A cryptosystem is computationally secure if the best currently known methods cannot break the cryptosystem by consuming a tractable amount of time and resources. "Tractable" can be precised to "depending on the length of the input in an at most polynomial way". Any computational task that exceeds this polynomial dependence of time and/or resources on the length of the input is called a *hard* computational task. One also speaks of *non-polynomial problems* or *np*–problems.
Public-key cryptosystems are computationally secure, because the encrypting function f_k is a *trapdoor function*, i.e. a function whose inverse is very difficult to compute. Provided that the computers are not fast enough and the public channel is really public, there is nothing Eve can do. The secret key always remains secret to Alice, the inverse encrypting function cannot be computed by Eve, the public key cannot be altered by Eve, the encrypted message and the public key reveal nothing about the original message.
The problem with computationally secure public-key cryptosystems is: *They might be insecure.* The most famous public-key cryptosystem, the RSA protocol [2], is based upon sophisticated number theoretical algebra, but nobody has found a proof that it is really a *hard* computational task to reverse the encryption. More precisely: It is not yet decided if prime number factorization is truly an np–problem. The situation is even worse: There already *is* an efficient algorithm for factorization, and it has been found by *Peter Shor* in 1994 [1]. Only, this algorithm makes use of a *quantum computer* which has not yet been built and which cannot be efficiently simulated on a classical computer. If one day a quantum computer can be built then any message ever encoded by a public key cryptosystem can be deciphered.
However, as long as nobody has found an efficient classical algorithm to break the cryptosystem on contemporary computer networks, the cryptosystem is assumed to be computationally secure.

Chapter 6

Quantum Cryptography

6.1 Quantum key distribution

The key distribution problem which plagues any private-key cryptosystem can be solved in a way which has not been foreseen by the information theorists, because it is a *physical* solution. Assume that there is a public *quantum channel* between Alice and Bob, then there are two physical theorems which make it possible to establish a shared secret key between Alice and Bob: 1) The *no-cloning theorem* and 2) the *uncertainty principle*. Roughly speaking, the no-cloning theorem states that *it is impossible to clone an unknown quantum state* and the uncertainty principle states that *any measurement disturbs the system*. Let us address these two important theorems.
The no-cloning theorem has independently been discovered by *Dieks* [31] and by *Wootters* and *Zurek* [32] in 1982, and it can be formulated as follows.

Theorem 8 (No-cloning theorem) *Assume that there is a machine that acts on a Hilbert space $\mathcal{H} \otimes \mathcal{H}$ and realizes for a fixed initial state $|\chi\rangle \in \mathcal{H}$ and any input state $|\psi\rangle \in \mathcal{H}$ the operation*

$$|\psi\rangle|\chi\rangle \mapsto |\psi\rangle|\psi\rangle, \tag{6.1}$$

then this machine contradicts the laws of quantum mechanics.

Proof. By the superposition principle any state $|\psi\rangle \in \mathcal{H}$ can be decomposed into two distinct states $|\psi_1\rangle, |\psi_2\rangle \in \mathcal{H}$ such that

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|\psi_1\rangle + |\psi_2\rangle). \tag{6.2}$$

Any allowed operation on a quantum system must be *linear*, so along (6.1) the cloning machine would act on $|\psi\rangle|\chi\rangle$ as

$$\frac{1}{\sqrt{2}}(|\psi_1\rangle + |\psi_2\rangle)|\chi\rangle = \frac{1}{\sqrt{2}}(|\psi_1\rangle|\chi\rangle + |\psi_2\rangle|\chi\rangle) \tag{6.3}$$

$$\mapsto \frac{1}{\sqrt{2}}(|\psi_1\rangle|\psi_1\rangle + |\psi_2\rangle|\psi_2\rangle). \tag{6.4}$$

This output is by construction different from $|\psi\rangle|\psi\rangle$ in contradiciton to (6.1), which proves the theorem. □

The crucial point is that the state $|\psi\rangle$ is assumed to be *unknown*. If $|\psi\rangle$ would be *known* then the machine could easily clone the state by preparing another copy of it. Such a machine would only work for this particular known state $|\psi\rangle$ and not for *any* state. There is yet another possibility: If $\mathcal{M} = \{|e_1\rangle, \ldots, |e_n\rangle\}$ is a set of mutually orthogonal states in $\mathcal{H}$ then one can construct a *quantum copy machine* which performs

$$|e_i\rangle|\chi\rangle \mapsto |e_i\rangle|e_i\rangle, \tag{6.5}$$

for any $|e_i\rangle \in \mathcal{M}$. Any state $|\psi\rangle \in \mathcal{H}$ can be decomposed into

$$|\psi\rangle = \sum_i \psi_i \,|e_i\rangle + |\phi\rangle, \tag{6.6}$$

where $\psi_i = \langle e_i|\psi\rangle$ and $\langle e_i|\phi\rangle = 0$ for all $|e_i\rangle \in \mathcal{M}$. Along (6.5) the quantum copy machine would act as

$$\begin{aligned} |\psi\rangle|\chi\rangle &= \sum_i \psi_i \,|e_i\rangle|\chi\rangle + |\phi\rangle|\chi\rangle & (6.7)\\ &\mapsto \sum_i \psi_i \,|e_i\rangle|e_i\rangle + |\phi\rangle|\chi\rangle, & (6.8)\end{aligned}$$

without any contradiciton to the assertions.
Heisenbergs *uncertainty principle* [21] (see page 31) has an information-theoretical implication: If information is gained about one of these two observables then this information gain *disturbs* the system in such a way that the subsequent information gain about the other observable is biased. In [33] one can find a theorem which expresses this information-theoretic implication of Heisenbergs uncertainty principle in a more explicit way:

Theorem 9 (Information gain implies disturbance) *In any attempt to distinguish between two non-orthogonal states, information gain is only possible at the expense of introducing disturbance to the signal.*

Proof. Due to the Stinespring dilation theorem (see page 38), any allowed quantum operation can be realized by a unitary operation on a larger Hilbert space. Let $|\psi\rangle, |\phi\rangle$ be two non-orthogonal quantum states taken from a Hilbert space $\mathcal{H}_A$. Then there is an ancilla space $\mathcal{H}_E$ such that any measurement aiming at the distinction between $|\psi\rangle$ and $|\phi\rangle$ is realized by a unitary operation $\hat{U}$ acting on the space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_E$, where the ancilla state represents the pointer state of the measuring device. Assuming that the operation does not disturb the original state, the operation most generally reads

$$\begin{aligned} \hat{U}|\psi\rangle|\chi\rangle &= |\psi\rangle|\chi_1\rangle & (6.9)\\ \hat{U}|\phi\rangle|\chi\rangle &= |\phi\rangle|\chi_2\rangle. & (6.10)\end{aligned}$$

In order to distinguish the two states $|\psi\rangle$ and $|\phi\rangle$ the pointer states $|\chi_1\rangle$ and $|\chi_2\rangle$ should be different. Taking the inner product of these two equation one obtains

$$\begin{aligned} \langle\chi_1|\chi_2\rangle\langle\psi|\phi\rangle &= \langle\chi|\chi\rangle\langle\psi|\phi\rangle & (6.11)\\ \langle\chi_1|\chi_2\rangle &= \langle\chi|\chi\rangle = 1, & (6.12)\end{aligned}$$

therefore $|\chi_1\rangle = |\chi_2\rangle$, so the above operation cannot be used to distinguish $|\psi\rangle$ and $|\phi\rangle$.
□
The proof only works for non-orthogonal states. If $\langle\psi|\phi\rangle = 0$ then the line (6.11) does not imply the next line. In fact, two orthogonal states can be distinguished without disturbing the system. The above theorem can be linked to the uncertainty principle in the following way. The distinction between $|\psi\rangle$ and $|\phi\rangle$ corresponds to the measurement of the two observables $\hat{A} = |\psi\rangle\langle\psi|$ and $\hat{B} = |\phi\rangle\langle\phi|$. If the two states are non-orthogonal then $\hat{A}$ and $\hat{B}$ do not commute and hence Heisenbergs uncertainty principle applies.

6.2 The BB84 protocol

The most famous quantum key distribution protocol is the BB84 which has been introduced by *Bennet* and *Brassard* in 1984 [3]. With this protocol it is possible to establish a shared random sequence between two parties. The protocol is secure against eavesdropping in that an eavesdropper can get no information about the sequence without disturbing it, thus his presence can be detected in which case the communication is stopped. The BB84 is a *non-deterministic* protocol, which means that it is only possible to distribute a *random* sequence. The BB84 cannot be used to send a *message* from Alice to Bob, where we understand a message as a sequence of symbols that can be *willingly* chosen. Even if Alice chooses the symbols by will, there is no way for her to determine which of these symbols are *correctly decoded* by Bob. Both parties can only agree upon a random subsequence of symbols which have successfully been communicated. Let us see how it works.

1. Alice randomly chooses a basis $\mathcal{B}_i$ out of two orthogonal bases $\mathcal{B}_0 = \{|0\rangle, |1\rangle\}$ and $\mathcal{B}_1 = \{|+\rangle, |-\rangle\}$, where

$$|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle). \tag{6.13}$$

then she randomly chooses a logical bit $j \in \{0, 1\}$ and encodes it by choosing the corresponding member of the previously chosen basis,

$$\begin{aligned} 0 &\mapsto |0\rangle, |+\rangle \\ 1 &\mapsto |1\rangle, |-\rangle. \end{aligned} \tag{6.14}$$

This procedure is equivalent to encoding two bits ij into the state $|\psi_{ij}\rangle$, where

$$\begin{aligned} |\psi_{00}\rangle &= |0\rangle \\ |\psi_{01}\rangle &= |1\rangle \\ |\psi_{10}\rangle &= |+\rangle \\ |\psi_{11}\rangle &= |-\rangle. \end{aligned} \tag{6.15}$$

It is the second bit j which can later be used as a key bit. After encoding, Alice sends the qubit to Bob.

2. Bob receives the qubit, randomly chooses $i' \in \{0, 1\}$ and measures the qubit in the basis $\mathcal{B}_{i'}$. The result $|\psi_{i'j'}\rangle$ is decoded using the inverse of the scheme (6.14), such that Bob's key bit reads j'.

3. Alice and Bob repeat the steps 1 and 2 $4N$ times.

4. Alice and Bob publicly compare their basis bit sequences $\boldsymbol{i}$ and $\boldsymbol{i}'$. They discard those key bits j_n and j'_n, respectively, where the measurement basis differs from the preparation basis, i.e. where $i_n \neq i'_n$. The remaining subsequence $\boldsymbol{s}$ of key bits typically is of the length $4N/2 = 2N$. If no eavesdropper is in the line and the channel is perfect, this subsequence coincides on Alice's and Bob's side.

5. Alice randomly selects a control sequence $\boldsymbol{c}$ of $4N/4 = N$ bits from $\boldsymbol{s}$ and publicly announces the position and the value of each bit of the control sequence.

6. Bob compares Alice's control sequence with the corresponding subsequence on his side. If both sequences do not coincide then the communication is stopped. Otherwise the remaining subsequence $\boldsymbol{k} = \boldsymbol{s} - \boldsymbol{c}$ is taken as the key. The typical length of $\boldsymbol{k}$ is $4N/4 = N$. If one increases the length $4N$ of the original sequence by some sufficiently large δ then one can make the probability very close to 1 that at least N key bits are left.

7. From the N key bits Alice and Bob distill a private key by using a classical privacy amplification protocol.

In the above form the protocol is only secure in the case of a *noiseless* quantum channel. If there is noise on the channel then Alice and Bob have to check the control sequence for a certain *threshold* of errors at the end of the communication. If this threshold is exceeded then there must have been an eavesdropper in the line and the key is discarded. The classical privacy amplification used in the last step to distill the final key is not a necessary element of the BB84. It just improves the security for the case of imperfect channels, where a certain amount of information may have leaked out to Eve who has hidden her presence in the channel noise.

6.3 The ping-pong protocol

The BB84 protocol is the most successful quantum cryptosystem. Its experimental realization has already reached a level of high practicality [5, 4, 34]. Other protocols [13, 14, 35, 36] are either basically equivalent to BB84, have significant limitations or cannot be realized with standard components. (For a very readable review on quantum cryptography see [37].) However, there are also some limitiations of the BB84 protocol. First, the BB84 is *non-deterministic*, i.e. Alice can encode a classical bit into a quantum state which is then sent to Bob, but she cannot determine the bit value that Bob eventually *decodes*. Inspite of that, such non-deterministic communication can be used to establish a *shared secret key* between Alice and Bob, consisting of a sequence of random bits. This secret key can then be used to encrypt a message which is sent through a classical public channel. Second, the BB84 is not *instantaneous*, i.e. Bob must wait until the transmission stops, then he establishes a public connection with Alice, they exchange some more information, after that he is able to decode the key, and yet they have to transmit the encrypted message over a classical channel. Finally, the BB84 is not optimally *effective*, because on average every second transmitted qubit must be discarded due to a mismatch in preparation basis and measurement basis. The development of other cryptographic codes which overcome these limitations and which are experimentally feasible with relatively small effort, is an important issue and poses a great challenge.

Recently, a novel quantum communication protocol has been presented [16] that allows *secure direct communication*, where the message is deterministically sent through the quantum channel, but can only be decoded after a final transmission of classical information.

In [P2] a novel secure communication protocol is proposed which is based on an entangled pair of qubits and which allows asymptotically secure key distribution and quasi-secure direct communication. Since the information is transferred in a deterministic manner, no qubits have to be discarded. The security against arbitrary eavesdropping attacks is shown for the case of a perfect quantum channel. In case of eavesdropping attacks with full information gain, the detection rate is 50% per control transmission. The experimental realization of the protocol is feasible with relatively small effort, which also makes commercial applications conceivable. The ping-pong protocol can be used for the transmission of either a secret key or a plaintext message. In the latter case, the protocol is *quasi-secure*, i.e. an eavesdropper is able to gain a small amount of message information before being detected. In case of a *key transmission* the protocol is asymptotically secure. In contrast to other quantum cryptographic schemes, the presented scheme is *instantaneous*, i.e. the information can be decoded during the transmission and no final transmission of additional information is needed. The basic idea of the protocol, encoding information by local operations on an EPR pair, has already been raised by Bennett and Wiesner [15]. In our protocol, we follow this idea, but abandon the *dense* coding feature in favour of a *secure* transmission.

6.3.1 Basic idea

When two photons are maximally *entangled* in their polarization degree of freedom, then each single photon is not polarized at all. Denote the horizontal and vertical polarization state by $|0\rangle$ and $|1\rangle$, respectively, then the *Bell states* $|\psi^\pm\rangle = \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle)$ are maximally entangled states in the two-particle Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. A measurement of the polarization of one photon, say A, leads to a completely random result. This is reflected by the fact that the corresponding *reduced density matrices*, $\rho_A^\pm := \mathrm{Tr}_B\{|\psi^\pm\rangle\langle\psi^\pm|\}$ are both equal to the complete mixture, $\rho_A^\pm = \frac{1}{2}\mathbb{1}_A$. Hence, no experiment performed on only one photon can distinguish these states from each other. However, since the states $|\psi^\pm\rangle$ are mutually orthogonal, a measurement on *both* photons can perfectly distinguish the states from each other. In other words: One bit of information can be encoded in the states $|\psi^\pm\rangle$, which is completely unavailable to anyone who has only access to one of the photons. As one can easily verify, the unitary operator $\hat{\sigma}_z^A \equiv (\hat{\sigma}_z \otimes \mathbb{1}) = (|0\rangle\langle 0| - |1\rangle\langle 1|) \otimes \mathbb{1}$ flips between the two states $|\psi^\pm\rangle$,

$$\hat{\sigma}_z^A|\psi^\pm\rangle = |\psi^\mp\rangle. \tag{6.16}$$

Altough $\hat{\sigma}_z^A$ acts *locally*, i.e. on one photon only, it has a *non-local* effect. Someone who has access to one single photon only, can *encode* one bit of information, but he cannot

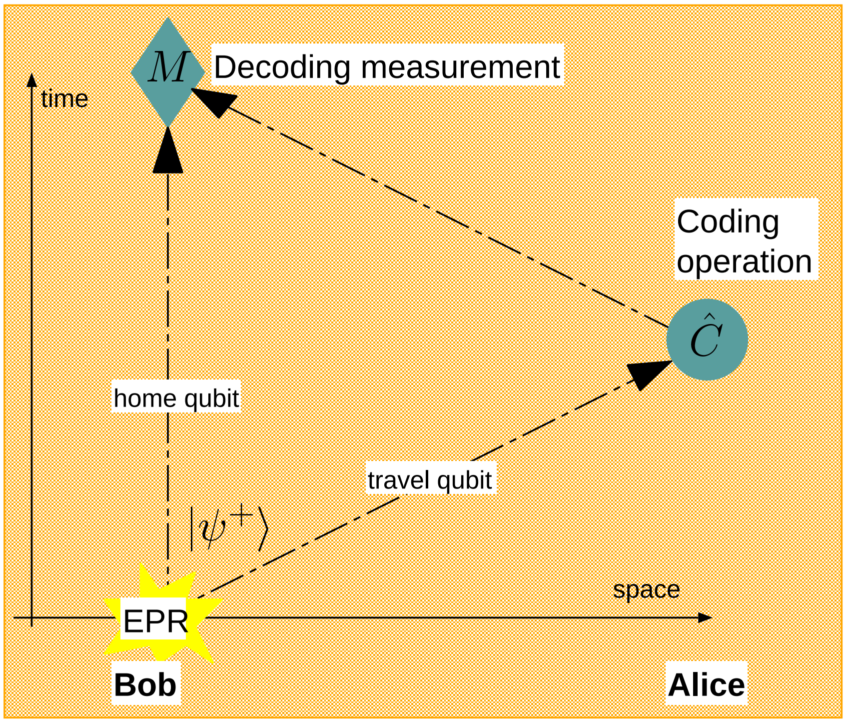


Figure 6.1: Message mode. Dashed lines are qubit transfers.

decode it, since he has no access to the other photon. This is a situation perfectly suited for a cryptographic scenario.

6.3.2 Scheme

Bob prepares two photons in the state $|\psi^+\rangle$. He keeps one photon, the "*home qubit*", and sends the other one, the "*travel qubit*", to Alice ("ping!"). Alice decides either to perform the operation $\hat{\sigma}_z$ on the travel qubit or to do nothing, i.e. to perform the operation $\mathbb{1}$. Then she sends the travel qubit back to Bob ("pong!"). Bob, who has now both qubits again, performs a Bell measurement resulting in either $|\psi^+\rangle$ or $|\psi^-\rangle$, depending on what Alice did. Thus, he has received one bit of information from Alice. One qubit travels forth and back ("ping-pong!") and one bit of information flows from Alice to Bob. Let us introduce two communication modes, "*message mode*" and "*control mode*" (see Figs. 6.1,6.2). By default, Alice and Bob are in message mode and communicate the way described above. With probability c, Alice switches to control mode and instead of performing her operation on the travel qubit, she performs a measurement in the basis $\mathcal{B}_z = \{|0\rangle, |1\rangle\}$. Using the public channel, she sends the result to Bob, who then also switches to control mode and performs a measurement in the same basis $\mathcal{B}_z$. Bob compares his own result with Alice's result. If both results coincide, Bob knows that Eve is in the line and stops the communication. [t] Let us give an explicit algorithm for the protocol.

p.0) Protocol is initialized. $n = 0$. The message to be transmitted is a sequence

$x^N = (x_1, \ldots, x_N)$, where $x_n \in \{0, 1\}$.

p.1) $n = n + 1$. Alice and Bob are set to message mode. Bob prepares two qubits in the Bell state $|\psi^+\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$.

p.2) He stores one qubit, the *home qubit*, and sends the other one, the *travel qubit*, to Alice through the quantum channel.

p.3) Alice receives the travel qubit. With probability c she switches to control mode and proceeds with c.1, else she proceeds with m.1.

c.1) Alice measures the travel qubit in the basis $\mathcal{B}_z$ and obtains the result $i \in \{0, 1\}$ with equal probability.

c.2) She sends i through the public channel to Bob.

c.3) Bob receives i from the public channel, switches to control mode and measures the home qubit in the basis $\mathcal{B}_z$ resulting in the value j.

c.4) $(i = j)$: Eve is detected. Abort transmission. $(i \neq j)$: Set $n = n - 1$ and Goto p.1.

m.1) Define $\hat{C}_0 := \mathbb{1}$ and $\hat{C}_1 := \hat{\sigma}_z$. For $x_n \in \{0, 1\}$, Alice performs the coding operation $\hat{C}_{x_n}$ on the travel qubit and sends it back to Bob.

m.2) Bob receives the travel qubit and performs a Bell measurement on both qubits resulting in the final state $|\psi'\rangle \in \{|\psi^+\rangle, |\psi^-\rangle\}$. He decodes the message as

$$|\psi'\rangle = \begin{cases} |\psi^+\rangle \Rightarrow x_n = 0 \\ |\psi^-\rangle \Rightarrow x_n = 1 \end{cases} . \tag{6.17}$$

m.3) $(n < N)$: Goto p.1. $(n = N)$: Goto p.4.

p.4) Message x^N is transmitted from Alice to Bob. Communication successfully terminated.

6.3.3 Security proof

Eve is an evil quantum physicist able to build all devices that are allowed by the laws of quantum mechanics. Her aim is to find out which operation Alice performs. Eve has no access to Bob's home qubit, so all her operations are restricted to the travel qubit, whose state is (to Eve) indistinguishable from the complete mixture $\rho_A = \mathrm{Tr}_B\{|\psi^+\rangle\langle\psi^+|\} = \frac{1}{2}\mathbb{1}_A$. The most general quantum operation is a *completely positive map* $\mathcal{E} : \mathcal{S}(\mathcal{H}_A) \to \mathcal{S}(\mathcal{H}_A)$ on the state space $\mathcal{S}(\mathcal{H}_A)$. Due to the *Stinespring dilation theorem* [38], any completely positive map can be realized by a unitary operation on a larger Hilbert space. For $\mathcal{H}_A$ and $\mathcal{E}$ given, there is an *ancilla space* $\mathcal{H}_E$ of dimension $\dim \mathcal{H}_E \leq (\dim \mathcal{H}_A)^2$, an ancilla state $|\chi\rangle \in \mathcal{H}_E$, and a unitary operation $\hat{E}$ on $\mathcal{H}_A \otimes \mathcal{H}_E$, such that for all states $\rho_A \in \mathcal{S}(\mathcal{H}_A)$, we have

$$\mathcal{E}(\rho_A) = \mathrm{Tr}_E\{\hat{E}(\rho_A \otimes |\chi\rangle\langle\chi|)\hat{E}^\dagger\}. \tag{6.18}$$

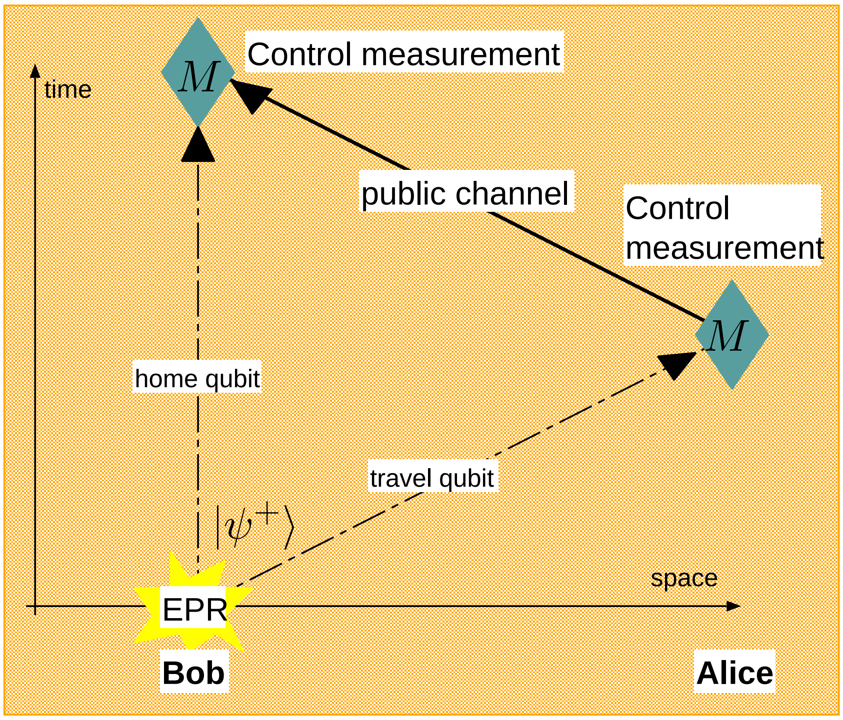


Figure 6.2: Control mode. Solid lines are classical transfers.

In order to gain information about Alice's operation, Eve should first perform the unitary *attack operation* $\hat{E}$ on the composed system, then let Alice perform her coding operation $\hat{C}$ on the travel qubit, and finally perform a measurement on the composed system (see Fig. 6.3). Since a probable control measurement by Alice takes place *before* Eve's final measurement, the latter has no influence on the detection probability for Eve's attack. All that can be detected is the attack operation $\hat{E}$. Let us analyze the detection probability d, given an attack operation $\hat{E}$. Since for Eve the state of the travel qubit is indistinguishable from the complete mixture, we can replace the state of the travel qubit by the apriori mixture $\rho_A = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$, which corresponds to the situation where Bob sends the travel qubit in either of the states $|0\rangle$ or $|1\rangle$, with equal probability $p = 1/2$. Let us at first consider the case where Bob sends $|0\rangle$. Alice adds an ancilla in the state $|\chi\rangle$ and performs the unitary operation $\hat{E}$ on both systems, resulting in

$$|\psi'\rangle = \hat{E}|0,\chi\rangle = \alpha|0,\chi_0\rangle + \beta|1,\chi_1\rangle, \tag{6.19}$$

where $|\chi_0\rangle, |\chi_1\rangle$ are pure ancilla states uniquely determined by $\hat{E}$, and $|\alpha|^2 + |\beta|^2 = 1$. In a subsequent control measurement, Alice measures the travel qubit in the basis $\mathcal{B}_z = \{|0\rangle, |1\rangle\}$ and sends the result to Bob. Without Eve, the result will always read "0", hence the detection probability for Eve's attack in a control run reads

$$d = |\beta|^2 = 1 - |\alpha|^2. \tag{6.20}$$

Now let us analize how much information Eve can maximally gain when there is no

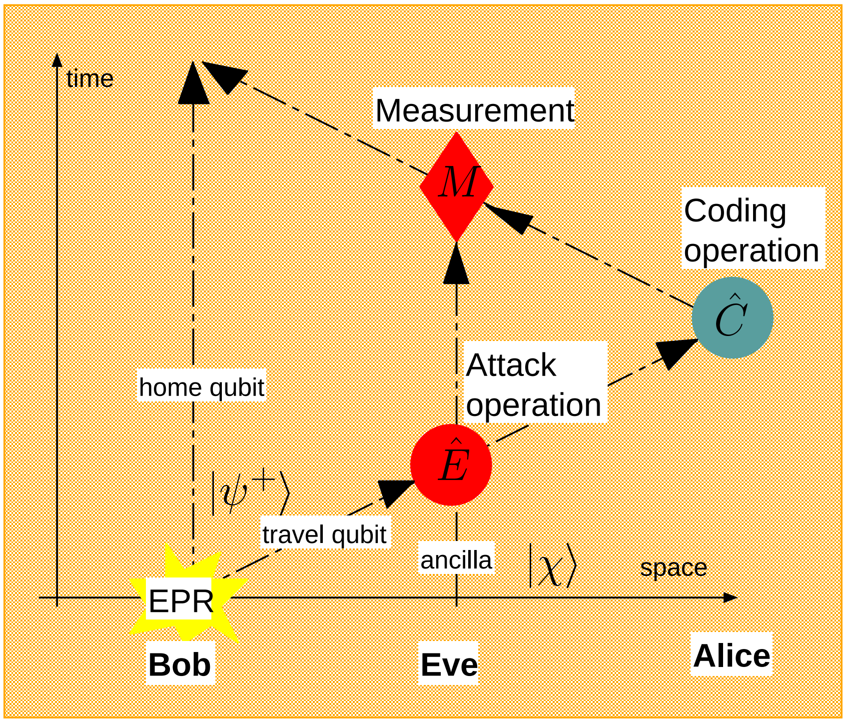


Figure 6.3: A general eavesdropping attack.

control run. After Eve's attack operation, the state of the system reads

$$\rho' = |\psi'\rangle\langle\psi'| = |\alpha|^2|0,\chi_0\rangle\langle 0,\chi_0| + |\beta|^2|1,\chi_1\rangle\langle 1,\chi_1| \tag{6.21}$$

$$+\alpha\beta^*|0,\chi_0\rangle\langle 1,\chi_1| + \alpha^*\beta|1,\chi_1\rangle\langle 0,\chi_0|, \tag{6.22}$$

which can be rewritten in the orthogonal basis $\{|0,\chi_0\rangle, |1,\chi_1\rangle\}$ as

$$\rho' = \begin{pmatrix} |\alpha|^2 & \alpha\beta^* \\ \alpha^*\beta & |\beta|^2 \end{pmatrix}. \tag{6.23}$$

Alice encodes her bit by applying the operation $\hat{C}_0 = \mathbb{1}$ or $\hat{C}_1 = \hat{\sigma}_z$ to the travel qubit, with probability p_0 and p_1, respectively. The state of the travel qubit after Eve's attack operation and after Alice's encoding operation reads

$$\rho'' = \begin{pmatrix} |\alpha|^2 & \alpha\beta^*(p_0 - p_1) \\ \alpha^*\beta(p_0 - p_1) & |\beta|^2 \end{pmatrix}. \tag{6.24}$$

The maximal amount I_0 of classical information that can be extracted from this state is given by the *von-Neumannn entropy*, $I_0 = S(\rho'') \equiv -\mathrm{Tr}\{\rho'' \log_2 \rho''\}$. In order to calculate the von-Neumann entropy we need the eigenvalues λ of ρ'', which are the roots of the characteristic polynomial $\det(\rho'' - \lambda\mathbb{1})$, yielding the two eigenvalues

$$\lambda_{1,2} = \frac{1}{2}\left(1 \pm \sqrt{1 - 4|\alpha\beta|^2[1 - (p_0 - p_1)^2]}\right), \tag{6.25}$$

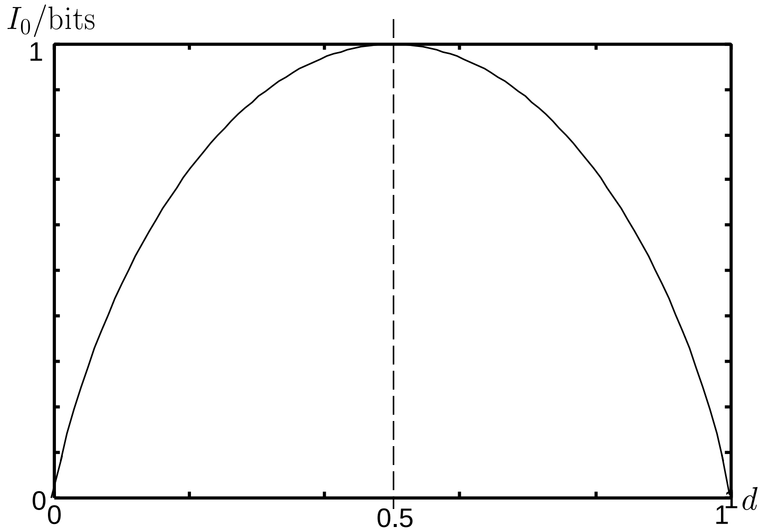


Figure 6.4: Eve's maximal eavesdropping information gain I_0 as a function of the detection probability d. The function is equal to the Shannon entropy of a binary source. For $d = 0$ there is no information gain, so Eve can only gain information at the cost of being detectable.

so we have

$$I_0 = -\lambda_1 \log_2 \lambda_1 - \lambda_2 \log_2 \lambda_2. \tag{6.26}$$

The maximal information gain I_0 can be expressed as a function of the detection probability d. Using (6.20), we have $|\alpha\beta|^2 = (1 - |\beta|^2)|\beta|^2 = (d - d^2)$, and therefore

$$\lambda_{1,2} = \frac{1}{2} \pm \frac{1}{2}\sqrt{1 - (4d - 4d^2)[1 - (p_0 - p_1)^2]}. \tag{6.27}$$

Now assume that Bob sends $|1\rangle$ rather than $|0\rangle$. The above calculations can be done in full analogy, resulting in the same crucial relations (6.26,6.27). Eve's task is, of course, to minimize d. Though if she chooses an eavesdropping action $\hat{E}$ that provides $d = 0$, then $\lambda_1 = 1$, $\lambda_2 = 0$, which implies $I_0 = 0$, therefore Eve can gain no information at all. Thus we have shown that *any effective eavesdropping attack can be detected*.
In the case $p_0 = p_1 = 1/2$, where Alice encodes exactly 1 bit, expression (6.27) simplifies to $\lambda_{1,2} = \frac{1}{2} \pm |\frac{1}{2} - d|$, or $\lambda_1 = d$, $\lambda_2 = 1 - d$. Interestingly, the maximal information gain is equal to the Shannon entropy of a binary source,

$$I_0(d) = -d \log_2 d - (1 - d) \log_2(1 - d). \tag{6.28}$$

The function $I_0(d)$ which is plotted in Fig. 6.4 has a maximum at $d = 1/2$, and can be inversed on the interval $[0, 1/2]$, giving a monotonous function $0 \leq d(I_0) \leq 1/2$, $I_0 \in [0, 1]$. By choosing a desired information gain $I_0 > 0$ per attack, Eve has to face a detection probability $d(I_0) > 0$. If she wants to gain the *full* information ($I_0 = 1$), the detection probability is $d(I_0 = 1) = 1/2$.

6.3.4 Direct communication versus key distribution

In contrast to quantum key distribution protocols like BB84 [3], the ping-pong protocol provides a *deterministic* transmission of bits, hence it is possible to communicate the message *directly* from Alice to Bob. Assuming that Eve wants to gain full information in each attack, the ping-pong protocol provides a detection probability of $d = 1/2$, which is significantly higher than the detection probability of the BB84 protocol, where we have $d = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ for the same situation. Furthermore, the BB84 protocol has a probability of $1/2$ that a transmitted bit has to be discarded due to the wrong choice of basis on both sides.

Taking into account the probability c of a control run, the *effective transmission rate*, i.e. the number of message bits per protocol run, reads $r = 1 - c$, which is equal to the probability for a message transfer. Say, Eve wants to eavesdrop one message transfer without being detected. The probability for this event reads

$$\begin{aligned} s(c,d) &= (1-c) + c(1-d)(1-c) + \\ &\quad c^2(1-d)^2(1-c) + \ldots \qquad (6.29) \\ &= \frac{1-c}{1-c(1-d)}, \qquad (6.30) \end{aligned}$$

where the terms in the (geometric) series correspond to Eve having to survive $0, 1, 2, \ldots$ control runs before she gets to eavesdrop on a message run, finally yielding the desired information of $I_0(d)$ bits. After n successful attacks Eve gains $nI_0(d)$ bits of information and survives with probability s^n, thus the probability to successfully eavesdrop $I = nI_0(d)$ bits reads $s(I,c,d) = s(c,d)^{I/I_0(d)}$, so

$$s(I,c,d) = \left(\frac{1-c}{1-c(1-d)}\right)^{I/I_0(d)}, \qquad (6.31)$$

where $I_0(d)$ is given by (6.28). For $c > 0$, $d > 0$, this value decreases exponentially but is nonzero. In the limit $I \to \infty$ (a message or key of infinite length) we have $s \to 0$, so the protocol is *asymptotically secure*, just like the BB84 protocol. Let us give an example. A convenient choice of the control parameter is $c = 0.5$, where on average every second bit is a control bit. Say, Eve wants to gain full information in each attack, thus $I_0 = 1$ and $d = 1/2$. The probability that Eve successfully eavesdrops 1 character (8 bits) is already as low as $s \approx 0.039$. In Fig. 6.5 we have plotted the eavesdropping success probability as a function of the information gain I, for $c = 0.5$ and for different detection probabilities d that Eve can choose. (Note that for $d < 1/2$ Eve only gets *part* of the message right and does not even know *which* part.) If desired, the security can arbitrarily be improved by increasing the control parameter c at the cost of decreasing the transmission rate. Let us call such communication "*quasi-secure*". If we want a *perfectly* secure communication (which is, strictly speaking, also not really perfect), we must abandon the direct transfer in favour of a *key* transfer. In this case, Alice does not transmit the message directly to Bob but rather takes a random sequence of N bits from a secret random number generator. After a succesful transmission, the random sequence is used as a *shared secret key* between Alice and Bob. Eve has virtually no advantage in eavesdropping only a few bits, because one can choose classical privacy

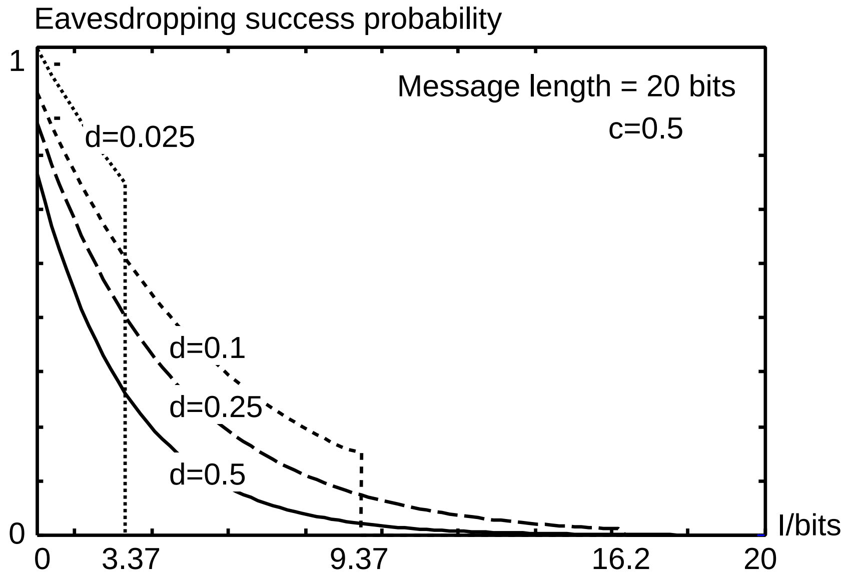


Figure 6.5: Eavesdropping success probability as a function of the maximal eavesdrop information, plotted for different detection probabilities d. The graph ends as soon as the message length of 20 bits is reached. for example, if Eve chooses a very low detection probability of $d = 0.025$ then she can eavesdrop at most 3.37 bits without being detected and still her success probability is significantly below 1.

amplification protocols that make it *very* hard to decode parts of the message with only *some* of the key bits given. The one-time-pad scheme, by the way, is not quite a good choice, because here each eavesdropped key bit directly yields one decoded message bit. Anyway, as soon as Eve is detected, the transfer stops and she has learned nothing but a sequence of nonsense random bits.

6.4 Attacking the ping-pong protocol

In its present form, the ping-pong protocol is designed for a noiseless quantum channel. Security is provided by making use of the fact that information can only be extracted from the channel at the cost of disturbing the quantum state and thereby generating errors in the message decoded by the authorized receiver. As soon as the quantum channel gets noisy, which is the generic situation in the real world, a door is opened for eavesdroppers to probably attack the protocol without being detected. To find out how far this door is opened is an important issue to be studied in future research. There are two recent publications [18, 17] where two different strategies to attack the protocol are proposed. In the following we will address these two attack scenarios.

6.4.1 Denial-of-Service attack

Qing-yu Cai proposes a *Denial-of-Service attack*, in short *DoS attack*, on the ping-pong protocol [18]. Such type of attack tries to disturb the communication between the au-

thorized parties without aiming at information gain, therefore it is not an "eavesdropping attack". And in fact, as the author of the attack concedes, there are many effective methods of classical *message authentification* which provide sufficient protection against this kind of attack.
Eve captures the travel qubit on its way from Bob to Alice and performs a measurement in the z-basis $\mathcal{B}_z = \{|0\rangle, |1\rangle\}$. This measurement destroys the entanglement between travel qubit and home qubit, therefore it is no longer possible for Alice to encode her message. The state after Eve's measurement reads either $|01\rangle$ or $|10\rangle$, each with probability $1/2$. Alice's encoding operation

$$\hat{C}_j = \hat{\sigma}_z^j \tag{6.32}$$

on the attacked travel qubit does not modify the state. After sending the travel qubit back to Bob, he will perform a decoding measurement in the Bell basis $\{|\psi^+\rangle, |\psi^-\rangle\}$ on travel qubit and home qubit. With probability $1/2$ he will either get the result $|\psi^+\rangle$ or $|\psi^-\rangle$ and thus he will decode a completely random sequence which has no correlation with Alice's message. In a control run there is no chance to detect the attack because Alice performs the control measurement in the z-basis. The author proposes an additional mechanism to detect the attack with probability $1/2$ per control run. The proposed modification is in fact equivalent to the comparison of control bits: Being in control mode, Alice decides with probability $1 - c_0$ to return the travel qubit untouched, thus sending a logical "0". If Bob afterwards decodes a logical "1" (represented by the state $|\psi^-\rangle$) and both compare their results, then Eve is detected. However, apart from control bit comparison there are numerous other classical (and probably more efficient) methods of message authentification, e.g. checksum analysis, so in fact there is no need for a modification of the protocol on the level of *quantum* operations.
It should also be noted that there is no way for Eve to modify the message *directedly*, i.e. she cannot send *disinformation*, because the message information is stored in the entanglement correlations to which Eve has no access. If she takes the role of Alice and sends the directedly modified travel qubit back to Bob, then she has to provide another qubit for Alice which cannot be entangled with Bob's home qubit, and thus her action will be detected with probability $1/2$ per control run.

6.4.2 Eavesdropping attack on an imperfect quantum channel

Wojcik has proposed an attack scheme on the ping-pong protocol which makes use of the fact that the decoding errors generated by an eavesdropping attack can be hidden in *transmission losses* of the channel [17].
There is a widely accepted criterion for the *practicality* and *security* of communication schemes which has been formulated by *Brassard et al.* [39]:

> *"In order to be practical and secure, a quantum key distribution scheme must be based on existing – or nearly existing – technology, but its security must be guaranteed against an eavesdropper with unlimited computing power whose technology is limited only by the laws of quantum mechanics."*

As has been shown in [39], the imperfections of a channel put limitations on the security of *any* quantum cryptographic protocol. The ping-pong protocol is practicable, because it requires technology which is available nowadays. However, its security has only been

shown for the case of a *perfect* quantum channel. In the real world there is no perfect channel, therefore it is an important task to study the security of the protocol also for the case of an imperfect channel.
Now we come to Wojciks attack scheme. The lossy quantum channel is described by a single-photon transmission efficiency η, where an ideal channel means that $\eta = 1$. Eve uses two auxiliary spatial modes x, y together with a single photon in the state $|0\rangle$. She attacks the channel twice, for the first time during the transmission from Bob to Alice (B-A attack) and for the second time during the transmission from Alice to Bob (A-B attack). Eve prepares her ancilla systems x and y in the state $|\mathrm{vac}\rangle_x$ and $|0\rangle_y$, respectively, where $|\mathrm{vac}\rangle$ denotes the vacuum state. The initial state of the total system is thus given by

$$|\mathrm{initial}\rangle = |\Psi^+\rangle_{ht}|\mathrm{vac}\rangle_x|0\rangle_y. \tag{6.33}$$

where the indices h and t refer to the home and travel qubit, respectively, while x and y refer to the auxiliary modes which are under Eve's control. After Bob sends the travel qubit to Alice, Eve intercepts it and applies a joint operation

$$\hat{Q}_{txy} = \mathrm{S\hat{W}AP}_{tx}\mathrm{C\hat{B}PS}_{txy}\hat{\mathrm{H}}_y. \tag{6.34}$$

The SWAP_{tx} corresponds to swapping the states of travel qubit and x-qubit,

$$\mathrm{S\hat{W}AP}|\psi\rangle|\phi\rangle = |\phi\rangle|\psi\rangle, \tag{6.35}$$

the CBPS_{txy} corresponds to the action of a controlled polarized beam splitter,

$$\left.\begin{array}{l} |0\rangle|\mathrm{vac}\rangle|0\rangle \\ |0\rangle|\mathrm{vac}\rangle|1\rangle \\ |1\rangle|\mathrm{vac}\rangle|0\rangle \\ |1\rangle|\mathrm{vac}\rangle|1\rangle \end{array}\right\} \stackrel{\mathrm{CBPS}}{\longmapsto} \left\{\begin{array}{l} |0\rangle|0\rangle|\mathrm{vac}\rangle \\ |0\rangle|\mathrm{vac}\rangle|0\rangle \\ |1\rangle|\mathrm{vac}\rangle|0\rangle \\ |1\rangle|1\rangle|\mathrm{vac}\rangle \end{array}\right., \tag{6.36}$$

and H_y corresponds to the *Hadamard* gate,

$$\left.\begin{array}{l} |0\rangle \\ |1\rangle \end{array}\right\} \stackrel{\mathrm{H}}{\longmapsto} \left\{\begin{array}{l} \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle \\ \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle \end{array}\right.. \tag{6.37}$$

The action of $\hat{Q}_{txy}$ transforms the initial state into the state

$$\begin{aligned} |B-A\rangle = & \frac{1}{2}|0\rangle_h(|\mathrm{vac}\rangle_t|1\rangle_x|0\rangle_y + |1\rangle_t|1\rangle_x|\mathrm{vac}\rangle_y) \\ & +\frac{1}{2}|1\rangle_h(|\mathrm{vac}\rangle_t|0\rangle_x|1\rangle_y + |0\rangle_t|0\rangle_x|\mathrm{vac}\rangle_y). \end{aligned} \tag{6.38}$$

We see that the state of the home qubit is now entangled with both the travel qubit and Eve's x qubit. The cost of this "entanglement splitting" is that the travel qubit has now a vacuum component, i.e. Bob will with probability 1/2 receive no photon. This photon loss can in principle be detected, but the ping-pong protocol contains in its present form no mechanism to check such losses. We see that Eve's attack reproduces the correct correlations between travel and home qubit: Whenever Alice and Bob perform a control measurement, Alice either detects no travel photon, or their respective photon states are anticorrelated. Therefore, the protocol is not aborted inspite of Eve's activities. Alice now encodes the bit j by the operation $\hat{Z}^j$, where we conveniently denote

$$\hat{X} := \hat{\sigma}_x, \quad \hat{Y} := \hat{\sigma}_y, \quad \hat{Z} := \hat{\sigma}_z. \tag{6.39}$$

After Alice performs her encoding operation on the travel qubit, the system is in the state

$$\begin{aligned} \hat{Z}^j|B-A\rangle &= \frac{1}{2}|0\rangle_h(|\mathrm{vac}\rangle_t|1\rangle_x|0\rangle_y + (-1)^j|1\rangle_t|1\rangle_x|\mathrm{vac}\rangle_y) \\ &+\frac{1}{2}|1\rangle_h(|\mathrm{vac}\rangle_t|0\rangle_x|1\rangle_y + (-1)^j|0\rangle_t|0\rangle_x|\mathrm{vac}\rangle_y). \end{aligned} \tag{6.40}$$

As soon as Alice sends the travel qubit back to Bob, Eve intercepts it and applies the second attack operation $\hat{Q}_{txy}^{-1}$, so that the system is now in the state

$$|A-B\rangle = \frac{1}{\sqrt{2}}(|0\rangle_h|1\rangle_t|j\rangle_y + |1\rangle_h|0\rangle_t|0\rangle_y)|\mathrm{vac}\rangle_x). \tag{6.41}$$

We see that the information is now partially encoded in Eve's y qubit. Eve finishes her attack by measuring the y qubit resulting in the bit value k. The encoding of Alice's bit into Eve's qubit is only partial because if Alice encodes $j = 1$ there is still a probability of $1/2$ that Eve erroneously decodes $k = 0$. The rest of the information is still stored in the entanglement between home and travel qubit. Bob performs his Bell measurement and decodes the bit value l. If $j = 0$ then Bob correctly decodes $l = 0$ because home and travel qubit factorize off into the Bell state $|\Psi^+\rangle$. If $j = 1$ then the total system is entangled in such a way that the partial trace over Eve's x and y qubits leaves home and travel qubit in the mixed state

$$\hat{\rho}_{ht} = \frac{1}{2}(|01\rangle\langle 01| + |10\rangle\langle 10|). \tag{6.42}$$

Therefore with probability $1/2$ Bob erroneously decodes the bit value $l = 0$. Eve's actions transforms a perfect channel into an imperfect one. Denoting the bit values of Alice, Eve and Bob by i, j and k, respectively, one calculates the following joint probabilities $p_{jkl} = p(j,k,l)$:

$$p_{000} = \frac{1}{2} \tag{6.43}$$

$$p_{100} = p_{101} = p_{110} = p_{111} = \frac{1}{8}, \tag{6.44}$$

and all other probabilities equal to zero. This gives the following mutual informations:

$$I_{AE} = I_{AB} = \frac{3}{4}\log_2\frac{4}{3} \simeq 0.311 \tag{6.45}$$

$$I_{BE} = 1 + \frac{5}{8}\log_2 5 - \frac{3}{2}\log_2 3 \simeq 0.074. \tag{6.46}$$

We see that Alice and Bob share the same information as Alice and Eve. Eve can further reduce the mutual information between Alice and Bob by applying with probability 1/2 the additional operation

$$\hat{S}_{ty} = \hat{X}\hat{Z}_t \mathrm{C\hat{N}OT}_{ty}\hat{X}_t \tag{6.47}$$

right after the operation $\hat{Q}^{-1}_{txy}$, so that the the final state after this "symmetrization" procedure reads

$$|A-B\rangle^{(S)} = \frac{1}{2}(|\Psi^+\rangle_{ht}|j\rangle_y + |\Psi^-\rangle_{ht}|j\rangle_y - |\Psi^+_{ht}|1\rangle_y + |\Psi^+\rangle_{ht}|1\rangle_y)|\mathrm{vac}\rangle_x). \tag{6.48}$$

The additional operation disturbs the communication between Alice and Bob by reducing their mutual information

$$I_{AB} = \frac{3}{4}\log_2 3 - 1 \simeq 0.189. \tag{6.49}$$

In other words, the attack scheme works partially as a Denial-of-Service attack. Surprisingly, the *qubit error rate (QBER)* is not affected by the additional operation, in both cases it reads

$$\mathrm{QBER} = \sum_k (p_{0k1} + p_{1k0}) = \frac{1}{4}. \tag{6.50}$$

As already stated, the eavesdropping attack is not detected, because the control mechanism does not check for photon losses. Moreover, in the case of an imperfect channel the photon losses can be hidden in the natural channel losses. The channel losses induced by Eve read 50%. If the transmission efficiency of the channel is $\eta \leq 1/2$ then Eve can replace the channel by a better one with the efficiency 2η, so that she compensates the eavesdropping losses. However, now the channel is "too good", because in message mode the channel is used twice, thus the efficiency now should read η^2 and not $4\eta^2$ as with the better channel. In order to compensate this effect, Eve has to filter out 75% of the photons reaching Bob in the message mode. In the end, Eve has completely hidden her eavesdropping action in losses that appear to be natural. If $\eta > 1/2$ then Eve should not attack all the time but instead only a fraction of $\mu = 2(1-\eta)$, which reduces the mutual information between Alice and Eve. Though, as a closer investigation shows, it is possible to eavesdrop a bigger amount of information than what is exchanged between Alice and Bob, i.e. $I_{AE} \geq I_{AB}$, up to efficiencies of $\eta \leq 0.6$.
Wojcik now proposes two modifications of the control mechanism to restore the security of the protocol. First, the QBER could be tested by sacrificing a part of the message. Eve's attacks produces a QBER of 1/4 which is very high as compared to the typical QBER of a few percent encountered in long-distance quantum cryptography [4, 5, 34], so there is a good chance to detect Eve's attack. But there is a second, more effective detection strategy without such a sacrifice. Say, Alice switches to control mode and measures the home qubit in the z-basis. Now she *delays* her announcement of the mode status and waits exactly that amount of time which would be needed for the travel qubit to go back to Bob. After Alice's control measurement, the travel qubit is in a vacuum state which is subsequently filled with a photon by Eve's B-A attack operation $\hat{Q}^{-1}_{txy}$. Thus with a certain probability Bob will detect a travel photon although Alice has not sent it back to Bob. This *double detection* can be used as an evidence of Eve's action.

Alice should announce the mode status after this small period of time and if Bob notices that he received a travel photon although Alice has switched to control mode, he aborts the communication. Otherwise Bob performs his ususal control measurement in the z-basis and they both compare their measurement results.
Concluding, Wojcik has proposed an attack scheme which is undetectable as long as the channel is imperfect and the communication protocol is not altered. He has given an explicit mechanism to protect the protocol against his attack so that the modified ping-pong protocol fulfills the condition of both practicality and security.

Altogether, this is rather good news for the ping-pong protocol, as it shows that the protocol can be protected against a sophisticated eavesdropping attack even for the case of an imperfect channel. However, there remains the need for a rigorous and general proof that the ping-pong protocol is secure against arbitrary attacks on an imperfect quantum channel. For the BB84 this has eventually been shown, but it was a lot of work and it took a lot of time [40, 41, 42, 43, 44]. In view of Wojcik's work there is good hope that also for imperfect channels the ping-pong protocol will once be proven to be as secure as the BB84.

6.4.3 Realizing the ping-pong protocol

The ping-pong protocol can be implemented by use of standard optical components, therefore its realization is relatively easy. The Bell state $|\psi^+\rangle$ can be created by *parametric down-conversion*. Bob's Bell measurement must only distinguish between the states $|\psi^\pm\rangle$, which can also be accomplished. The storage of one photon is necessary only for a duration corresponding to twice the distance between Alice and Bob. The encoding procedure corresponds to a controlled $\hat{\sigma}_z$-operation, which can be realized by triggered optical elements. The correlation test involves a simple measurement of the linear polarization in a fixed basis. Altogether, the experimental realization of the ping-pong protocol should be feasible using nowaday's technology.
Since October 2002 there is a collaboration with an experimentalist group at the university of Potsdam under the guidance of Dr. Martin Ostermeyer. Nonlinear crytals, beamsplitters, polarizers, detectors, and glass fibers are used to generate, transmit and receive the signals and perform the necessary operations in message mode and control mode. In Fig. 6.6 the experimental setup is schematically represented, and Fig. 6.7 shows a photographic picture of the setup. Since the protocol only requires the distinction between the two Bell states $|\psi^\pm\rangle$, the Bell analyzers do not face the usual difficulty in distinguishing between all four Bell states. Moreover, the number of detectors can be reduced to only two at both sides of the channel by a clever design of the setup.

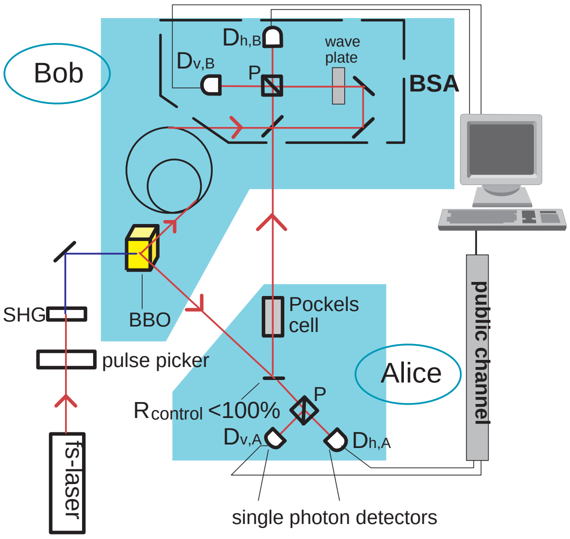


Figure 6.6: Schematic setup to realize ping-pong coding: P denotes polarizing beam splitters, SHG denotes the frequency doubling crystal, BSA the bell state analyzer, BBO the crystal to be used for parametric down conversion. The detectors are denoted as DV, DH, etc.

Figure 6.7: The experimental setup of the ping-pong protocol.

Chapter 7

Summary and Outlook

The guiding line through this thesis was the encoding and transmission of information through a quantum channel. Coding is a necessary element for the transmission of information: The source message has to be translated into a sequence of symbols which is compatible with the physical properties of the channel. In the quantum case the channel is a compound quantum system whose state can be manipulated by the sender and read out by the receiver. The resources which are needed to encode a particular message into a channel state yields a measure for the information content of the message. Reducing these resources on average for a given ensemble of source messages is the task of compression. Finding a clever way to make the information sensitive to eavesdropping is the task of quantum cryptography. We have studied several ways to compress messages with or without loss of information. Lossless compression requires variable-length coding which is a new task to quantum information theory. Communicating a message directly while providing security against eavesdropping is also a new task. My thesis has addressed these tasks and has therefore hopefully enriched the scientific discussion in the field of quantum information theory. In the following let us briefly summarize the single chapters of the thesis.

Chapter 1: Classical Information
The basic concepts of classical information theory were reviewed. The notions of code, message, channel and information were introduced and discussed, also some important elements of probability theory were presented and related to the theory of communication. The idea of compression was outlined and illustrated by a discussion of Shannon's source coding theorem. It was explained why lossless compression is only possible by use of variable-length codes.

Chapter 2: Quantum Information
The axiomatic foundations of classical and quantum mechanics were presented in a compact version so that the analogies become clear. The focus was then put on quantum theory, where the notions of pure and mixed state, of entanglement and measurement have been clarified. It was shown how the theory of quantum channels is connected to the theory of quantum measurement. The qubit as the elementary unit of quantum information was introduced and some theoretical background was given which is very helpful for the mathematical treatment of

qubit systems. Quantum messages were defined in close analogy to the classical case, which includes the extension of the standard concept of block messages to that of variable-length messages. The length operator has been defined and its relation to the information content of a message has been discussed. The concept of a quantum code has been introduced and illustrated by explicit examples. In particular some ideas how to realize variable-length message spaces have been proposed and discussed.

Chapter 3: Concepts of Quantum Data Compression
The idea of quantum data compression was explained and illustrated by the Schumacher compression scheme, which is a lossy block compression scheme.

Chapter 4: Lossless Compression
In analogy to the classical case the concept of variable-length coding was applied to obtain a compression scheme which reduces the size of quantum messages without any loss of information. Some theorems were given concerning the properties of certain compression strategies. In particular it has been shown that it is impossible to compress an unknown quantum message without loss of information. It was also pointed out that quantum prefix codes are not very useful in the context of lossless quantum data compression. The main idea of a successful compression scheme was outlined, which makes use of a classical side-channel to store length information about the encoded messages. By means of this side-channel it becomes possible to reduce the quantum resources below the von-Neumann entropy of the source message ensemble. It was shown that if the resources of the classical side-channel are additionally considered, then the total size of the compressed message is bounded from below by the von-Neumann entropy of the source message ensemble. This statement represents an anologue to Shannon's noiseless coding theorem for lossless codes. An explicit compression protocol has been given and evaluated for a sample message ensemble.

Chapter 5: Classical Cryptography
The concepts of classical cryptography were reviewed and some basic methods, in particular the private-key and public-key cryptosystems, were discussed. The notion of perfect security was explained and illustrated by the Vernam cipher.

Chapter 6: Quantum Cryptography
The basic idea of quantum cryptography was explained and opposed against the basic idea of classical cryptography. While classical cryptography is based on the mathematical properties of suitable encoding functions, quantum cryptography is based on the physical properties of suitable quantum channels. In order to illustrate this concept, the BB84 protocol was discussed, which is a non-deterministic protocol providing the distribution of a secret random key between two parties. In contrast to that, a deterministic protocol was presented which can be used to communicate a message directly in a secure manner. The security of this so-called "ping-pong protocol" against arbitrary eavesdropping attacks has been shown for the case of a perfect quantum channel. It remains an open task to prove the security also for the case of an imperfect quantum channel. Two recently published

attack schemes on the ping-pong protocol were discussed and it was pointed out that none of them represents a serious threat to the security of the protocol. Lastly, the experimental realization of the ping-pong protocol was briefly discussed, also mentioning the collaboration with an experimental group in Potsdam.

It could be interesting to combine lossless quantum compression with quantum cryptography. By combining the methods of quantum cryptography with the methods of lossless compression, the efficiency of secure data transfer could possibly be increased. One also should investigate how the framework of variable-length messages applies to quantum computation, since the data stored in the register of a quantum computer could also be regarded as a variable-length quantum message. Furthermore, the fact that the Fock space is used for the representation of variable-length messages suggests that it might be advantageous to apply the techniques of second quantization here. This idea has recently been followed by *Rallan* and *Vedral* in [45] and it offers a lot of interesting possibilities.
Concerning the ping-pong protocol the next important thing is certainly to find a rigorous security proof for the case of imperfect quantum channels. The work of Wojcik [17] feeds the hope that also for imperfect channels the security of the ping-pong protocol can be shown in general. It is also interesting to follow the experimental progress in realizing the protocol. Based on these experiences it might be possible to envisage a future commercial application of the ping-pong protocol, just as it has already become true for the BB84.

Bibliography

[1] P.W. Shor. Algorithms for quantum computation: Discrete log and factoring. *Proc. 35th IEEE Symposium on Foundations of Computer Science*, pages 124–134, 1994.

[2] R.L. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120–126, 1978.

[3] C.H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. *Proc. IEEE Int. Conf. on Computers, Systems, and Signal Processing, Bangalore*, pages 175–179, 1984.

[4] D. Stucki, N. Gisin, O. Guinnard, G. Ribordy, and H. Zbinden. Quantum key distribution over 67 km with a plug&play system. *New J. Phys.*, 4:41, 2002.

[5] R.J. Hughes, J.E. Nordholt, D. Derkacs, and C.G. Peterson. Practical free-space quantum key distribution over 10 km in daylight and at night. *New J. Phys.*, 4:43, 2002.

[6] B. Schumacher. Quantum coding. *Phys. Rev. A*, 51:2738–2747, 1995.

[7] R. Jozsa and B. Schumacher. A new proof of the quantum noiseless coding theorem. *J. Mod. Opt.*, 41:2343, 1994.

[8] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W.K. Wootters. Classical information capacity of a quantum channel. *Phys. Rev. A*, 54:4707, 1996.

[9] C.E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27:379–423, 623–656, 1948.

[10] S.L. Braunstein, C.A. Fuchs, D. Gottesmann, and H.-K. Lo. A quantum analog of huffman coding. *IEEE International Symposium on Information Theory*, 1998. quant-ph/9805080.

[11] B. Schumacher and M.D. Westmoreland. Indeterminate-length quantum coding. quant-ph/0011014, 2000.

[12] T.M. Cover and J.A. Thomas. *Elements of information theory*. Wiley, New York, 1991.

[13] A. Ekert. Quantum cryptography based on bell's theorem. *Phys. Rev. Lett.*, 67:661–663, 1991.

[14] C.H. Bennett, G. Brassard, and N.D. Mermin. Quantum cryptography without bell's theorem. *Phys. Rev. Lett.*, 68(5):557, 1991.

[15] C. Bennett and S.J. Wiesner. Communication via one- and two-particle operators on einstein-podolski-rosen states. *Phys. Rev. Lett.*, 69:2881, 1992.

[16] A. Beige, B.-G. Englert, C. Kurtsiefer, and H. Weinfurter. Secure communication with a publicly known key. *Acta Phys. Pol. A*, 101:357, 2002.

[17] A. Wojcik. Eavesdropping on the "ping-pong" quantum communication protocol. *Phys. Rev. Lett.*, 90(15):157901, 2003.

[18] Q. Cai. The "ping-pong" protocol can be attacked without eavesdropping. *Phys. Rev. Lett.*, 91(10):109801, 2003.

[19] D.J.C. MacKay. Information theory, inference, and learning algorithms. http://wol.ra.phy.cam.ac.uk/mackay/itprnn/book.html, 1995–2000.

[20] J. Preskill. Lecture notes. http://www.theory.caltech.edu/people/preskill/ph219/, 1997–1999.

[21] W. Heisenberg. über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik. *Zeitschrift für Physik*, 43:172–198, 1927.

[22] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780, 1935.

[23] K. Kraus. *States, effects and operations: Fundamental notions of quantum theory.* Springer, Heidelberg, 1983.

[24] A.S. Holevo. *Statistical problems in quantum physics*, pages 104–119. Springer, Berlin, 1973. in "Proceedings of the second Japan–USSR Symposium on Probability Theory".

[25] A.S. Holevo. Some estimates on the information transmitted by quantum communication channels. *Problems of Information Transmission*, 9:177–183, 1973.

[26] A.S. Holevo. The capacity of the quantum channel with general signal states. quant-ph/9611023, 1996.

[27] Paul Hausladen, Richard Jozsa, Benjamin Schumacher, Michael Westmoreland, and William K. Wootters. Classical information capacity of a quantum channel. *Phys. Rev. A*, 54(3):1869–1876, 1996.

[28] B. Schumacher and M.D. Westmoreland. Sending classical information via a noisy quantum channel. *Phys. Rev. A*, 56:131–138, 1997.

[29] T. Felbinger. Qmatrix – a mathematica package for quantum information. http://www.quantum.physik.uni-potsdam.de/Timo_Felbinger/qmatrix, 2001.

[30] C.E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28:656–715, 1949.

[31] D. Dieks. Communication by epr devices. *Phys. Lett. A*, 92(6):271–272, 1982.

[32] W.K. Wootters and W.H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982.

[33] M.A. Nielsen and I.L. Chuang. *Quantum computation and quantum information*. University Press, Cambridge, 2000.

[34] C. Kurtsiefer, P. Zarda, M. Halder, H. Weinfurter, P.M. Gorman, P.R. Tapster, and J.G. Rarity. A step towards global key distribution. *Nature*, 419:450, 2002.

[35] D. Bruss. Optimal eavesdropping in quantum cryptography with six states. *Phys. Rev. Lett.*, 81:3018–3021, 1998.

[36] H. Aschauer and H.-J. Briegel. Private entanglement over arbitrary distances, even using a noisy apparatus. *Phys. Rev. Lett.*, 88:047902, 2002. quant-ph/0008051.

[37] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden. Quantum cryptography. *Rev. Mod. Phys.*, 74:145–195, 2002.

[38] W.F. Stinespring. Positive functions on c^*-algebras. *Proc. Amer. Math. Soc.*, 6:211, 1955.

[39] G. Brassard, N. Luetkenhaus, T. Mor, and B.C. Sanders. Limitations on practical quantum cryptography. *Phys. Rev. Lett.*, 85(6):1330, 2000.

[40] E. Biham and T. Mor. Bounds on information and the security of quantum cryptography. *Phys. Rev. Lett.*, 79:4034, 1997.

[41] N. Luetkenhaus. Bounds on information and the security of quantum cryptography. *Phys. Rev. A*, 61:052304, 2000.

[42] N. Lütkenhaus. Security against individual attacks for realistic quantum key distribution. *Phys. Rev. A*, 61:052304, 2000.

[43] D. Mayers. Unconditional security in quantum cryptography. *J. of ACM*, 48:351–406, 2001.

[44] H. Inamori, N. Luetkenhaus, and Dominic Mayers. Unconditional security of practical quantum key distribution. *quant-ph/9802025*, 2001.

[45] Luke Rallan and Vlatko Vedral. Energy requirements for quantum data compression and 1-1 coding. *Phys. Rev. A*, 68(7), 2003.

Acknowledgements

I would like to thank everybody who supported me during my PhD period or contributed in any way to this thesis. Specifically, I would like to thank Professor Martin Wilkens for giving me the opportunity to work on the fascinating field of quantum information theory and who enriched my mind through various discussions, ideas and hints concerning the topics of the thesis and beyond. My special thanks also belong to Jens Eisert with whom I had excellent discussions that gave me precious insights into the nature of quantum mechanics and who also supported me as a friend and collegue throughtout the studies. I am very grateful to Timo Felbinger who contributed very much to this work and who was a valuable collegue and co-author. I also would like to thank my collegues Janet Anders, Alexander Albus, Axel Friedenauer, Markus Cramer, Holger Hoffman, Carsten Henkel, Antonio Negretti and Geesche Boedecker for having created a friendly, open-minded team atmosphere which is not only important but indispensable for a fruitful research period. Many many thanks to my parents who were a great support all the time and who never stopped believing in me. And a big amorous thank-you to my girlfriend Oana Nistor who enriched this period with being there by my side. Thank you all.

This work has been supported by the International Max Planck Research School, by the Deutsche Forschungsgemeinschaft ("Schwerpunktprogramm Quanteninformationsverarbeitung") coordinated by G. Leuchs, and by the University of Potsdam.

Index