ELSEVIER

Physics Reports 369 (2002) 431–548

PHYSICS REPORTS

www.elsevier.com/locate/physrep

Fundamentals of quantum information theory

Michael Keyl

TU-Braunschweig, Institute of Mathematical Physics, Mendelssohnstraße 3, D-38106 Braunschweig, Germany

Received 3 June 2002
editor: J. Eichler

Abstract

In this paper we give a self-contained introduction to the conceptional and mathematical foundations of quantum information theory. In the first part we introduce the basic notions like entanglement, channels, teleportation, etc. and their mathematical description. The second part is focused on a presentation of the quantitative aspects of the theory. Topics discussed in this context include: entanglement measures, channel capacities, relations between both, additivity and continuity properties and asymptotic rates of quantum operations. Finally, we give an overview on some recent developments and open questions.

PACS: 03.67.−a; 03.65.−w

Contents

E-mail address: m.keyl@tu-bs.de (M. Keyl).

PII: S0370-1573(02)00266-1

1. Introduction

Quantum information and quantum computation have recently attracted a lot of interest. The promise of new technologies like safe cryptography and new "super computers", capable of handling otherwise untractable problems, has excited not only researchers from many different fields like physicists, mathematicians and computer scientists, but also a large public audience. On a practical level all these new visions are based on the ability to control the quantum states of (a small number of) microsystems individually and to use them for information transmission and processing. From a more fundamental point of view the crucial point is a reconsideration of the foundations of quantum mechanics in an information theoretical context. The purpose of this work is to follow the second path and to guide physicists into the theoretical foundations of quantum information and some of the most relevant topics of current research.

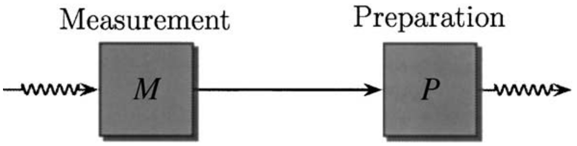
To this end the outline of this paper is as follows: The rest of this introduction is devoted to a rough and informal overview of the field, discussing some of its tasks and experimental realizations. Afterwards, in Section 2, we will consider the basic formalism which is necessary to present more detailed results. Typical keywords in this context are: systems, states, observables, correlations, entanglement and quantum channels. We then clarify these concepts (in particular, entanglement and channels) with several examples in Section 3, and in Section 4 we discuss the most important tasks of quantum information in greater detail. The last three sections are devoted to a more quantitative analysis, where we make closer contact to current research: In Section 5 we will discuss how entanglement can be measured. The topic of Section 6 are channel capacities, i.e. we are looking at the amount of information which can maximally be transmitted over a noisy channel and in Section 7 we consider state estimation, optimal cloning and related tasks.

Quantum information is a rapidly developing field and the present work can of course reflect only a small part of it. An incomplete list of other general sources the reader should consult is: the books of Lo [111], Gruska [76], Nielsen and Chuang [122], Bouwmeester et al. [23] and Alber et al. [3], the lecture notes of Preskill [130] and the collection of references by Cabello [37] which particularly contains many references to other reviews.

1.1. What is quantum information?

Classical information is, roughly speaking, everything which can be transmitted from a sender to a receiver with "letters" from a "classical alphabet" e.g. the two digits "0" and "1" or any other finite set of symbols. In the context of classical information theory, it is completely irrelevant which type of physical system is used to perform the transmission. This abstract approach is successful because it is easy to transform information between different types of carriers like electric currents in a wire, laser pulses in an optical fiber, or symbols on a piece of paper without loss of data; and even if there are losses they are well understood and it is known how to deal with them. However, quantum information theory breaks with this point of view. It studies, loosely speaking, that kind of information ("quantum information") which is transmitted by microparticles from a preparation device (sender) to a measuring apparatus (receiver) in a quantum mechanical experiment—in other words, the distinction between carriers of classical and quantum information becomes essential. This approach is justified by the observation that a lossless conversion of quantum information into classical information is in the above sense not possible. Therefore, quantum information is a *new kind of information*.

In order to explain why there is no way from quantum to classical information and back, let us discuss how such a conversion would look like. To convert quantum to classical information we need a device which takes quantum systems as input and produces classical information as output—this is nothing else than a measuring apparatus. The converse translation from classical to quantum information can be rephrased similarly as "parameter-dependent preparation", i.e. the classical input to such a device is used to control the state (and possibly the type of system) in



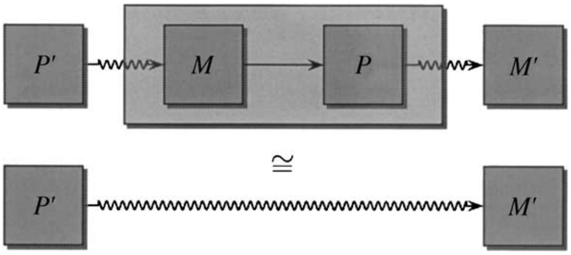


Fig. 1.1. Schematic representation of classical teleportation. Here and in the following diagrams a curly arrow stands for quantum systems and a straight one for the flow of classical information.

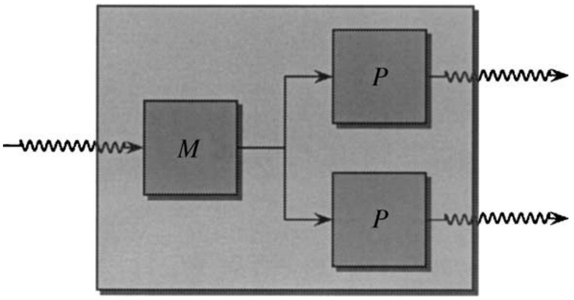
Fig. 1.2. A teleportation process should not affect the results of a statistical experiment with quantum systems. A more precise explanation of the diagram is given in the text.

which the microparticles should be prepared. A combination of these two elements can be done in two ways. Let us first consider a device which goes from classical to quantum to classical information. This is a possible task and in fact technically realized already. A typical example is the transmission of classical information via an optical fiber. The information transmitted through the fiber is carried by microparticles (photons) and is therefore quantum information (in the sense of our preliminary definition). To send classical information we have to prepare first photons in a certain state send them through the channel and measure an appropriate observable at the output side. This is exactly the combination of a classical → quantum with a quantum → classical device just described.

The crucial point is now that the converse composition—performing the measurement M first and the preparation P afterwards (cf. Fig. 1.1)—is more problematic. Such a process is called *classical teleportation*, if the particles produced by P are "indistinguishable" from the input systems. We will show the impossibility of such a device via a hierarchy of other "impossible machines" which traces the problem back to the fundamental structure of quantum mechanics. This finally will prove our statement that quantum information is a new kind of information.[1]

To start with, we have to clarify the precise meaning of "indistinguishable" in this context. This has to be done in a statistical way, because the only possibility to compare quantum mechanical systems is in terms of *statistical experiments*. Hence, we need an additional preparation device P' and an additional measuring apparatus M'. Indistinguishable now means that it does not matter whether we perform M' measurements directly on P' outputs or whether we switch a teleportation device in between; cf. Fig. 1.2. In both cases we should get the same *distribution of measuring results* for a large number of repetitions of the corresponding experiment. This requirement should hold for any preparation P' and any measurement M', but for fixed M and P. The latter means that we are not allowed to use a priori knowledge about P' or M' to adopt the teleportation process (otherwise we can choose in the most extreme case always P' for P and the whole discussion becomes meaningless).

[1] The following chain of arguments is taken from [168], where it is presented in greater detail. This concerns, in particular, the construction of Bell's telephone from a joint measurement, which we have omitted here.



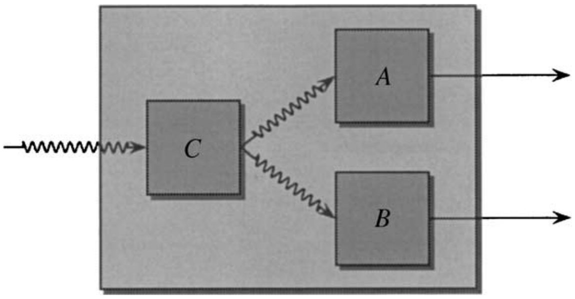


Fig. 1.3. Constructing a quantum copying machine from a teleportation device.

Fig. 1.4. Constructing a joint measurement for the observables A and B from a quantum copying machine.

The second impossible machine we have to consider is a *quantum copying machine*. This is a device C which takes one quantum system p as input and produces two systems p_1, p_2 of the same type as output. The limiting condition on C is that p_1 and p_2 are indistinguishable from the input, where "indistinguishable" has to be understood in the same way as above: Any statistical experiment performed with one of the output particles (i.e. always with p_1 or always with p_2) yields the same result as applied directly to the input p. To get such a device from teleportation is easy: We just have to perform an M measurement on p, make two copies of the classical data obtained, and run the preparation P on each of them; cf. Fig. 1.3. Hence if teleportation is possible copying is possible as well.

According to the "no-cloning theorem" of Wootters and Zurek [173], however, a quantum copy machine does not exist and this basically concludes our proof. However, we will give an easy argument for this theorem in terms of a third impossible machine—a *joint measuring device* M_{AB} for two arbitrary observables A and B. This is a measuring apparatus which produces each time it is invoked a pair (a, b) of classical outputs, where a is a possible output of A and b a possible output of B. The crucial requirement for M_{AB} again is of statistical nature: The statistics of the a outcomes is the same as for device A, and similarly for B. It is known from elementary quantum mechanics that many quantum observables are not jointly measurable in this way. The most famous examples are position and momentum or different components of angular momentum. Nevertheless, a device M_{AB} could be constructed for arbitrary A and B from a quantum copy machine C. We simply have to operate with C on the input system p producing two outputs p_1 and p_2 and to perform an A measurement on p_1 and a B measurement on p_2; cf. Fig. 1.4. Since the outputs p_1, p_2 are, by assumption, indistinguishable from the input p the overall device constructed this way would give a joint measurement for A and B. Hence, a quantum copying machine cannot exist, as stated by the no-cloning theorem. This in turn implies that classical teleportation is impossible, and therefore we cannot transform quantum information lossless into classical information and back. This concludes our chain of arguments.

1.2. Tasks of quantum information

So we have seen that quantum information is something new, but what can we do with it? There are three answers to this question which we want to present here. First of all let us remark that in fact all information in a modern data processing environment is carried by microparticles (e.g. electrons or photons). Hence, quantum information comes automatically into play. Currently, it is safe to ignore this and to use classical information theory to describe all relevant processes. If the size of the structures on a typical circuit decreases below a certain limit, however, this is no longer true and quantum information will become relevant.

This leads us to the second answer. Although it is far too early to say which concrete technologies will emerge from quantum information in the future, several interesting proposals show that devices based on quantum information can solve certain practical tasks much better than classical ones. The most well known and exciting one is, without a doubt, quantum computing. The basic idea is, roughly speaking, that a quantum computer can operate not only on one number per register but on *superpositions of numbers*. This possibility leads to an "exponential speedup" for some computations which makes problems feasible which are considered intractable by any classical algorithm. This is most impressively demonstrated by Shor's factoring algorithm [139,140]. A second example which is quite close to a concrete practical realization (i.e. outside the laboratory; see next section) is quantum cryptography. The fact that it is impossible to perform a quantum mechanical measurement without disturbing the state of the measured system is used here for the secure transmission of a cryptographic key (i.e. each eavesdropping attempt can be detected with certainty). Together with a subsequent application of a classical encryption method known as the "one-time" pad this leads to a cryptographic scheme with provable security—in contrast to currently used public key systems whose security relies on possibly doubtful assumptions about (pseudo) random number generators and prime numbers. We will come back to both subjects, quantum computing and quantum cryptography, in Sections 4.5 and 4.6.

The third answer to the above question is of more fundamental nature. The discussion of questions from information theory in the context of quantum mechanics leads to a deeper and in many cases to more quantitative understanding of quantum theory. Maybe the most relevant example for this statement is the study of entanglement, i.e. non-classical correlations between quantum systems, which lead to violations of the Bell inequalities.[2] Entanglement is a fundamental aspect of quantum mechanics and demonstrates the differences between quantum and classical physics in the most drastical way—this can be seen from Bell-type experiments, like the one of Aspect et al. [5], and the discussion about. Nevertheless, for a long time it was only considered as an exotic feature of the foundations of quantum mechanics which is not so relevant from a practical point of view. Since quantum information attained broader interest, however, this has changed completely. It has turned out that entanglement is an essential resource whenever classical information processing is outperformed by quantum devices. One of the most remarkable examples is the experimental realization of "entanglement enhanced" teleportation [24,22]. We have argued in Section 1.1 that *classical* teleportation, i.e. transmission of quantum information through a classical information channel, is impossible. If sender and receiver share, however, an entangled pair of particles (which can be used as an additional resource) the impossible task becomes, most surprisingly, possible [11]! (We will discuss this fact in detail in Section 4.1.) The study of entanglement and in particular the question *how it can be quantified* is therefore a central topic within quantum information theory (cf. Section 5). Further examples for fields where quantum information has led to a deeper and in particular more quantitative insight include "capacities" of quantum information channels and "quantum

[2] This is only a very rough characterization. A more precise one will be given in Section 2.2.

cloning". A detailed discussion of these topics will be given in Sections 6 and 7. Finally, let us remark that classical information theory benefits in a similar way from the synthesis with quantum mechanics. Beside the just mentioned channel capacities this concerns, for example, the theory of computational complexity which analyzes the scaling behavior of time and space consumed by an algorithm in dependence of the size of the input data. Quantum information challenges here, in particular, the fundamental Church–Turing hypotheses [45,152] which claims that each computation can be simulated "efficiently" on a Turing machine; we come back to this topic in Section 4.5.

1.3. Experimental realizations

Although this is a theoretical paper, it is of course necessary to say something about experimental realizations of the ideas of quantum information. Let us consider quantum computing first. Whatever way we go here, we need systems which can be prepared very precisely in few distinct states (i.e. we need "qubits"), which can be manipulated afterwards individually (we have to realize "quantum gates") and which can finally be measured with an appropriate observable (we have to "read out" the result).

One of the most far developed approaches to quantum computing is the ion trap technique (see Sections 4.3 and 5.3 in [23] and Section 7.6 of [122] for an overview and further references). A "quantum register" is realized here by a string of ions kept by electromagnetic fields in high vacuum inside a Paul trap, and two long-living states of each ion are chosen to represent "0" and "1". A single ion can be manipulated by laser beams and this allows the implementation of all "one-qubit gates". To get two-qubit gates as well (for a quantum computer we need at least one two qubit gate together with all one-qubit operations; cf. Section 4.5) the collective motional state of the ions has to be used. A "program" on an ion trap quantum computer starts now with a preparation of the register in an initial state—usually the ground state of the ions. This is done by optical pumping and laser cooling (which is in fact one of the most difficult parts of the whole procedure, in particular if many ions are involved). Then the "network" of quantum gates is applied, in terms of a (complicated) sequence of laser pulses. The readout finally is done by laser beams which illuminate the ions subsequently. The beams are tuned to a fast transition which affects only one of the qubit states and the fluorescent light is detected. Concrete implementations (see e.g. [118,102]) are currently restricted to two qubits; however, there is some hope that we will be able to control up to 10 or 12 qubits in the not too distant future.

A second quite successful technique is NMR quantum computing (see Section 5.4 of [23] and Section 7.7 of [122] together with the references therein for details). NMR stands for "nuclear magnetic resonance" and it is the study of transitions between Zeeman levels of an atomic nucleus in a magnetic field. The qubits are in this case different spin states of the nuclei in an appropriate molecule and quantum gates are realized by high-frequency oscillating magnetic fields in pulses of controlled duration. In contrast to ion traps, however, we do not use *one* molecule but a whole cup of liquid containing some 10^{20} of them. This causes a number of problems, concerning in particular the preparation of an initial state, fluctuations in the free time evolution of the molecules and the readout. There are several ways to overcome these difficulties and we refer the reader again to [23,122] for details. Concrete implementations of NMR quantum computers are capable to use up to five qubits [113]. Other realizations include the implementation of several known quantum algorithms on two and three qubits; see e.g. [44,96,109].

The fundamental problem of the two methods for quantum computation discussed so far is their lack of scalability. It is realistic to assume that NMR and ion-trap quantum computer with up to tens of qubits will exist somewhere in the future but not with thousands of qubits which are necessary for "real-world" applications. There are, however, many other alternative proposals available and some of them might be capable to avoid this problem. The following is a small (not at all exhaustive) list: atoms in optical lattices [28], semiconductor nanostructures such as quantum dots (there are many works in this area, some recent are [149,30,21,29]) and arrays of Josephson junctions [112].

A second circle of experiments we want to mention here is grouped around quantum communication and quantum cryptography (for a more detailed overview let us refer to [163,69]). Realizations of quantum cryptography are fairly far developed and it is currently possible to span up to 50 km with optical fibers (e.g. [93]). Potentially greater distances can be bridged by "free space cryptography" where the quantum information is transmitted through the air (e.g [34]). With this technology satellites can be used as some sort of "relays", thus enabling quantum key distribution over arbitrary distances. In the meantime there are quite a lot of successful implementations. For a detailed discussion we will refer the reader to the review of Gisin et al. [69] and the references therein. Other experiments concern the usage of entanglement in quantum communication. The creation and detection of entangled photons is here a fundamental building block. Nowadays this is no problem and the most famous experiment in this context is the one of Aspect et al. [5], where the maximal violation of Bell inequalities was demonstrated with polarization correlated photons. Another spectacular experiment is the creation of entangled photons over a distance of 10 km using standard telecommunication optical fibers by the Geneva group [151]. Among the most exciting applications of entanglement is the realization of entanglement based quantum key distribution [95], the first successful "teleportation" of a photon [24,22] and the implementation of "dense coding" [115]; cf. Section 4.1.

2. Basic concepts

After we have got a first, rough impression of the basic ideas and most relevant subjects of quantum information theory, let us start with a more detailed presentation. First, we have to introduce the fundamental notions of the theory and their mathematical description. Fortunately, much of the material we should have to present here, like Hilbert spaces, tensor products and density matrices, is known already from quantum mechanics and we can focus our discussion to those concepts which are less familiar like POV measures, completely positive maps and entangled states.

2.1. Systems, states and effects

As classical probability theory quantum mechanics is a *statistical theory*. Hence, its predictions are of probabilistic nature and can only be tested if the same experiment is repeated very often and the relative frequencies of the outcomes are calculated. In more operational terms this means: The experiment has to be repeated according to the same *procedure* as it can be set out in a detailed laboratory manual. If we consider a somewhat idealized model of such a *statistical experiment* we get, in fact, two different types of procedures: first *preparation procedures* which prepare a certain kind of physical system in a distinguished *state* and second *registration procedures* measuring a particular *observable*.

A mathematical description of such a setup basically consists of two sets $\mathscr{S}$ and $\mathscr{E}$ and a map $\mathscr{S} \times \mathscr{E} \ni (\rho, A) \to \rho(A) \in [0,1]$. The elements of $\mathscr{S}$ describe the states, i.e. preparations, while the $A \in \mathscr{E}$ represent all yes/no measurements (*effects*) which can be performed on the system. The probability (i.e. the relative frequency for a large number of repetitions) to get the result "yes", if we are measuring the effect A on a system prepared in the state ρ, is given by $\rho(A)$. This is a very general scheme applicable not only to quantum mechanics but also to a very broad class of statistical models, containing, in particular, classical probability. In order to make use of it we have to specify, of course, the precise structure of the sets $\mathscr{S}$ and $\mathscr{E}$ and the map $\rho(A)$ for the types of systems we want to discuss.

2.1.1. Operator algebras

Throughout this paper we will encounter three different kinds of systems: Quantum and classical systems and hybrid systems which are half classical, half quantum (cf. Section 2.2.2). In this subsection we will describe a general way to define states and effects which is applicable to all three cases and which therefore provides a handy way to discuss all three cases simultaneously (this will become most useful in Sections 2.2 and 2.3).

The scheme we are going to discuss is based on an algebra $\mathscr{A}$ of bounded operators acting on a Hilbert space $\mathscr{H}$. More precisely, $\mathscr{A}$ is a (closed) linear subspace of $\mathscr{B}(\mathscr{H})$, the algebra of bounded operates on $\mathscr{H}$, which contains the identity ($\mathbb{1} \in \mathscr{A}$) and is closed under products ($A, B \in \mathscr{A} \Rightarrow AB \in \mathscr{A}$) and adjoints ($A \in \mathscr{A} \Rightarrow A^* \in \mathscr{A}$). For simplicity we will refer to each such $\mathscr{A}$ as an *observable algebra*. The key observation is now that each type of system we will study in the following can be *completely characterized* by its observable algebra $\mathscr{A}$, i.e. once $\mathscr{A}$ is known there is a systematic way to derive the sets $\mathscr{S}$ and $\mathscr{E}$ and the map $(\rho, A) \mapsto \rho(A)$ from it. We frequently make use of this fact by referring to systems in terms of their observable algebra $\mathscr{A}$, or even by identifying them with their algebra and saying that *$\mathscr{A}$ is the system.*

Although $\mathscr{A}$ and $\mathscr{H}$ can be infinite dimensional in general, we will consider only finite-dimensional Hilbert spaces, as long as nothing else is explicitly stated. Since most research in quantum information is done up to now for finite-dimensional systems (the only exception in this work is the discussion of Gaussian systems in Section 3.3) this is not a too severe loss of generality. Hence we can choose $\mathscr{H} = \mathbb{C}^d$ and $\mathscr{B}(\mathscr{H})$ is just the algebra of complex $d \times d$ matrices. Since $\mathscr{A}$ is a subalgebra of $\mathscr{B}(\mathscr{H})$ it operates naturally on $\mathscr{H}$ and it inherits from $\mathscr{B}(\mathscr{H})$ the *operator norm* $\|A\| = \sup_{\|\psi\|=1} \|A\psi\|$ and the *operator ordering* $A \geqslant B \Leftrightarrow \langle \psi, A\psi \rangle \geqslant \langle \psi, B\psi \rangle \; \forall \psi \in \mathscr{H}$. Now we can define

$$\mathscr{S}(\mathscr{A}) = \{\rho \in \mathscr{A}^* \mid \rho \geqslant 0, \; \rho(\mathbb{1}) = 1\} , \tag{2.1}$$

where $\mathscr{A}^*$ denotes the *dual space* of $\mathscr{A}$, i.e. the set of all linear functionals on $\mathscr{A}$, and $\rho \geqslant 0$ means $\rho(A) \geqslant 0$, $\forall A \geqslant 0$. Elements of $\mathscr{S}(\mathscr{A})$ describe the states of the system in question while effects are given by

$$\mathscr{E}(\mathscr{A}) = \{A \in \mathscr{A} \mid A \geqslant 0, \; A \leqslant \mathbb{1}\} . \tag{2.2}$$

The probability to measure the effect A in the state ρ is $\rho(A)$. More generally, we can look at $\rho(A)$ for an arbitrary A as the *expectation value* of A in the state ρ. Hence, the idea behind Eq. (2.1) is to define states in terms of their expectation value functionals.

Both spaces are *convex*, i.e. $\rho, \sigma \in \mathcal{S}(\mathcal{A})$ and $0 \leqslant \lambda \leqslant 1$ implies $\lambda\rho + (1-\lambda)\sigma \in \mathcal{S}(\mathcal{A})$ and similarly for $\mathcal{E}(\mathcal{A})$. The *extremal points* of $\mathcal{S}(\mathcal{A})$, respectively, $\mathcal{E}(\mathcal{A})$, i.e. those elements which do not admit a proper convex decomposition ($x = \lambda y + (1-\lambda)z \Rightarrow \lambda = 1$ or $\lambda = 0$ or $y = z = x$), play a distinguished role: The extremal points of $\mathcal{S}(\mathcal{A})$ are *pure states* and those of $\mathcal{E}(\mathcal{A})$ are the *propositions* of the system in question. The latter represent those effects which register a property with certainty in contrast to non-extremal effects which admit some "fuzziness". As a simple example for the latter consider a detector which registers particles not with certainty but only with a probability which is smaller than one.

Finally, let us note that the complete discussion of this section can be generalized easily to infinite-dimensional systems, if we replace $\mathcal{H} = \mathbb{C}^d$ by an infinite-dimensional Hilbert space (e.g. $\mathcal{H} = \mathbf{L}^2(\mathbb{R})$). This would require, however, more material about C* algebras and measure theory than we want to use in this paper.

2.1.2. Quantum mechanics

For quantum mechanics we have

$$\mathcal{A} = \mathcal{B}(\mathcal{H}) , \tag{2.3}$$

where we have chosen again $\mathcal{H} = \mathbb{C}^d$. The corresponding systems are called *d-level systems* or *qubits* if $d = 2$ holds. To avoid clumsy notations we frequently write $\mathcal{S}(\mathcal{H})$ and $\mathcal{E}(\mathcal{H})$ instead of $\mathcal{S}[\mathcal{B}(\mathcal{H})]$ and $\mathcal{E}[\mathcal{B}(\mathcal{H})]$. From Eq. (2.2) we immediately see that an operator $A \in \mathcal{B}(\mathcal{H})$ is an effect iff it is positive and bounded from above by $\mathbb{1}$. An element $P \in \mathcal{E}(\mathcal{H})$ is a propositions iff P is a projection operator ($P^2 = P$).

States are described in quantum mechanics usually by density matrices, i.e. positive and normalized trace class[3] operators. To make contact to the general definition in Eq. (2.1) note first that $\mathcal{B}(\mathcal{H})$ is a Hilbert space with the Hilbert–Schmidt scalar product $\langle A, B\rangle = \mathrm{tr}(A^*B)$. Hence, each linear functional $\rho \in \mathcal{B}(\mathcal{H})^*$ can be expressed in terms of a (trace class) operator $\tilde{\rho}$ by[4] $A \mapsto \rho(A) = \mathrm{tr}(\tilde{\rho}A)$. It is obvious that each $\tilde{\rho}$ defines a unique functional ρ. If we start on the other hand with ρ we can recover the matrix elements of $\tilde{\rho}$ from ρ by $\tilde{\rho}_{kj} = \mathrm{tr}(\tilde{\rho}|j\rangle\langle k|) = \rho(|j\rangle\langle k|)$, where $|j\rangle\langle k|$ denotes the canonical basis of $\mathcal{B}(\mathcal{H})$ (i.e. $|j\rangle\langle k|_{ab} = \delta_{ja}\delta_{kb}$). More generally, we get for $\psi, \phi \in \mathcal{H}$ the relation $\langle\phi, \tilde{\rho}\psi\rangle = \rho(|\psi\rangle\langle\phi|)$, where $|\psi\rangle\langle\phi|$ now denotes the rank one operator which maps $\eta \in \mathcal{H}$ to $\langle\phi, \eta\rangle\psi$. In the following we drop the $\sim$ and use the same symbol for the operator and the functional whenever confusion can be avoided. Due to the same abuse of language we will interpret elements of $\mathcal{B}(\mathcal{H})^*$ frequently as (trace class) operators instead of linear functionals (and write $\mathrm{tr}(\rho A)$ instead of $\rho(A)$). However, we do not identify $\mathcal{B}(\mathcal{H})^*$ with $\mathcal{B}(\mathcal{H})$ in general, because the two different notations help to keep track of the distinction between spaces of states and spaces of observables. In addition, we equip $\mathcal{B}^*(\mathcal{H})$ with the trace-norm $\|\rho\|_1 = \mathrm{tr}\,|\rho|$ instead of the operator norm.

Positivity of the *functional* ρ implies positivity of the *operator* ρ due to $0 \leqslant \rho(|\psi\rangle\langle\psi|) = \langle\psi, \rho\psi\rangle$ and the same holds for normalization: $1 = \rho(\mathbb{1}) = \mathrm{tr}(\rho)$. Hence, we can identify the state space from

[3] On a finite-dimensional Hilbert space this attribute is of course redundant, since each operator is of trace class in this case. Nevertheless, we will frequently use this terminology, due to greater consistency with the infinite-dimensional case.

[4] If we consider infinite-dimensional systems this is not true. In this case the dual space of the observable algebra is much larger and Eq. (2.1) leads to states which are not necessarily given by trace class operators. Such "singular states" play an important role in theories which admit an infinite number of degrees of freedom like quantum statistics and quantum field theory; cf. [25,26]. For applications of singular states within quantum information see [97].

Eq. (2.1) with the set of density matrices, as expected for quantum mechanics. Pure states of a quantum system are the one-dimensional projectors. As usual, we will frequently identify the density matrix $|\psi\rangle\langle\psi|$ with the wave function ψ and call the latter in abuse of language a state.

To get a useful parameterization of the state space consider again the Hilbert–Schmidt scalar product $\langle\rho,\sigma\rangle=\mathrm{tr}(\rho^*\sigma)$, but now on $\mathscr{B}^*(\mathscr{H})$. The space of trace free matrices in $\mathscr{B}^*(\mathscr{H})$ (alternatively the functionals with $\rho(\mathbb{1})=0$) is the corresponding orthocomplement $\mathbb{1}^\perp$ of the unit operator. If we choose a basis $\sigma_1,\ldots,\sigma_{d^2-1}$ with $\langle\sigma_j,\sigma_k\rangle=2\delta_{jk}$ in $\mathbb{1}^\perp$ we can write each self-adjoint (trace class) operator ρ with $\mathrm{tr}(\rho)=1$ as

$$\rho=\frac{\mathbb{1}}{d}+\frac{1}{2}\sum_{j=1}^{d^2-1}x_j\sigma_j=:\frac{\mathbb{1}}{d}+\frac{1}{2}\vec{x}\cdot\vec{\sigma}\quad\text{with } \vec{x}\in\mathbb{R}^{d^2-1}\ . \tag{2.4}$$

If $d=2$ or $d=3$ holds, it is most natural to choose the Pauli matrices, respectively, the Gell–Mann matrices (cf. e.g. [48], Section 13.4) for the σ_j. In the qubit case it is easy to see that $\rho\geqslant0$ holds iff $|\vec{x}|\leqslant1$. Hence the state space $\mathscr{S}(\mathbb{C}^2)$ coincides with the *Bloch ball* $\{\vec{x}\in\mathbb{R}^3\mid|\vec{x}|\leqslant1\}$, and the set of pure states with its boundary, the *Bloch sphere* $\{\vec{x}\in\mathbb{R}^3\mid|\vec{x}|=1\}$. This shows in a very geometric way that the pure states are the extremal points of the convex set $\mathscr{S}(\mathscr{H})$. If ρ is more generally a pure state of a d-level system we get

$$1=\mathrm{tr}(\rho^2)=\frac{1}{d}+\frac{1}{2}|\vec{x}|^2\Rightarrow|\vec{x}|=\sqrt{2(1-1/d)}\ . \tag{2.5}$$

This implies that all states are contained in the ball with radius $2^{1/2}(1-1/d)^{1/2}$, however, not all operators in this set are positive. A simple example is $d^{-1}\mathbb{1}\pm2^{1/2}(1-1/d)^{1/2}\sigma_j$, which is positive only if $d=2$ holds.

2.1.3. *Classical probability*

Since the difference between classical and quantum systems is an important issue in this work let us reformulate classical probability theory according to the general scheme from Section 2.1.1. The restriction to finite-dimensional observable algebras leads now to the assumption that all systems we are considering admit a finite set X of *elementary events*. Typical examples are: throwing a dice $X=\{1,\ldots,6\}$, tossing a coin $X=\{\text{“head”},\text{“number”}\}$ or *classical bits* $X=\{0,1\}$. To simplify the notations we write (as in quantum mechanics) $\mathscr{S}(X)$ and $\mathscr{E}(X)$ for the spaces of states and effects.

The observable algebra $\mathscr{A}$ of such a system is the space

$$\mathscr{A}=\mathscr{C}(X)=\{f:X\to\mathbb{C}\} \tag{2.6}$$

of complex-valued functions on X. To interpret this as an operator algebra acting on a Hilbert space $\mathscr{H}$ (as indicated in Section 2.1.1) choose an arbitrary but fixed orthonormal basis $|x\rangle, x\in X$ in $\mathscr{H}$ and identify *the function* $f\in\mathscr{C}(X)$ with *the operator* $f=\sum_x f_x|x\rangle\langle x|\in\mathscr{B}(\mathscr{H})$ (we use the same symbol for the function and the operator, provided confusion can be avoided). Most frequently we have $X=\{1,\ldots,d\}$ and we can choose $\mathscr{H}=\mathbb{C}^d$ and the canonical basis for $|x\rangle$. Hence, $\mathscr{C}(X)$ becomes the algebra of *diagonal* $d\times d$ matrices. Using Eq. (2.2) we immediately see that $f\in\mathscr{C}(X)$ is an effect iff $0\leqslant f_x\leqslant1$, $\forall x\in X$. Physically, we can interpret f_x as the probability that the effect f registers the elementary event x. This makes the distinction between propositions and “fuzzy” effects very transparent: $P\in\mathscr{E}(X)$ is a proposition iff we have either $P_x=1$ or $P_x=0$ for all $x\in X$. Hence, the propositions $P\in\mathscr{C}(X)$ are in one-to-one correspondence with the subsets $\omega_P=\{x\in X\mid P_x=1\}\subset X$

which in turn describe the *events* of the system. Hence, P registers the event ω_P with certainty, while a fuzzy effect $f < P$ does this only with a probability less than one.

Since $\mathscr{C}(X)$ is finite dimensional and admits the distinguished basis $|x\rangle\langle x|, x \in X$ it is naturally isomorphic to its dual $\mathscr{C}^*(X)$. More precisely: each linear functional $\rho \in \mathscr{C}^*(X)$ defines and is uniquely defined by the function $x \mapsto \rho_x = \rho(|x\rangle\langle x|)$ and we have $\rho(f) = \sum_x f_x \rho_x$. As in the quantum case we will identify the function ρ with the linear functional and use the same symbol for both, although we keep the notation $\mathscr{C}^*(X)$ to indicate that we are talking about states rather than observables.

Positivity of $\rho \in \mathscr{C}^*(X)$ is given by $\rho_x \geqslant 0$ for all x and normalization leads to $1 = \rho(\mathbb{1}) = \rho(\sum_x |x\rangle\langle x|) = \sum_x \rho_x$. Hence to be a state $\rho \in \mathscr{C}^*(X)$ must be a *probability distribution* on X and ρ_x is the probability that the elementary event x occurs during statistical experiments with systems in the state ρ. More generally $\rho(f) = \sum_j \rho_j f_j$ is the probability to measure the effect f on systems in the state ρ. If P is in particular, a proposition, $\rho(P)$ gives the probability for the event ω_P. The pure states of the system are the *Dirac measures* δ_x, $x \in X$; with $\delta_x(|y\rangle\langle y|) = \delta_{xy}$. Hence, each $\rho \in \mathscr{S}(X)$ can be decomposed *in a unique way* into a convex linear combination of pure states.

2.1.4. Observables

Up to now we have discussed only effects, i.e. yes/no experiments. In this subsection we will have a first short look at more general observables. We will come back to this topic in Section 3.2.4 after we have introduced channels. We can think of an observable E taking its values in a finite set X as a map which associates to each possible outcome $x \in X$ the effect $E_x \in \mathscr{E}(\mathscr{A})$ (if $\mathscr{A}$ is the observable algebra of the system in question) which is true if x is measured and false otherwise. If the measurement is performed on systems in the state ρ we get for each $x \in X$ the probability $p_x = \rho(E_x)$ to measure x. Hence, the family of the p_x should be a probability distribution on X, and this implies that E should be a *positive operator-valued measure* (POV measure) on X.

Definition 2.1. Consider an observable algebra $\mathscr{A} \subset \mathscr{B}(\mathscr{H})$ and a finite[5] set X. A family $E = (E_x)_{x \in X}$ of effects in $\mathscr{A}$ (i.e. $0 \leqslant E_x \leqslant \mathbb{1}$) is called a POV measure on X if $\sum_{x \in X} E_x = \mathbb{1}$ holds. If all E_x are projections, E is called *projection-valued measure* (*PV measure*).

From basic quantum mechanics we know that observables are described by self-adjoint operators on a Hilbert space $\mathscr{H}$. But, how does this point of view fit into the previous definition? The answer is given by the spectral theorem [134, Theorem VIII.6]: Each self-adjoint operator A on a finite-dimensional Hilbert space $\mathscr{H}$ has the form $A = \sum_{\lambda \in \sigma(A)} \lambda P_\lambda$ where $\sigma(A)$ denotes the *spectrum* of A, i.e. the set of eigenvalues and P_λ denotes the projection onto the corresponding eigenspace. Hence, there is a unique PV measure $P = (P_\lambda)_{\lambda \in \sigma(A)}$ associated to A which is called the *spectral measure* of A. It is uniquely characterized by the property that the *expectation value* $\sum_\lambda \lambda \rho(P_\lambda)$ of P in the state ρ is given for any state ρ by $\rho(A) = \mathrm{tr}(\rho A)$; as it is well known from quantum mechanics. Hence, the traditional way to define observables within quantum mechanics perfectly fits into the scheme just outlined, however it only covers the projection-valued case and therefore admits no fuzziness. For this reason POV measures are sometimes called *generalized observables*.

[5] This is of course an artificial restriction and in many situations not justified (cf. in particular the discussion of quantum state estimation in Section 4.2 and Section 7). However, it helps us to avoid measure theoretical subtleties; cf. Holevo's book [79] for a more general discussion.

Finally, note that the eigenprojections P_λ of A are elements of an observable algebra $\mathscr{A}$ iff $A \in \mathscr{A}$. This shows two things: First of all we can consider self-adjoint elements of *any* *-subalgebra $\mathscr{A}$ of $\mathscr{B}(\mathscr{H})$ as observables of $\mathscr{A}$-systems, and this is precisely the reason why we have called $\mathscr{A}$ *observable* algebra. Secondly, we see why it is essential that $\mathscr{A}$ is really a subalgebra of $\mathscr{B}(\mathscr{H})$: if it is only a linear subspace of $\mathscr{B}(\mathscr{H})$ the relation $A \in \mathscr{A}$ does not imply $P_\lambda \in \mathscr{A}$.

2.2. Composite systems and entangled states

Composite systems occur in many places in quantum information theory. A typical example is a register of a quantum computer, which can be regarded as a system consisting of N qubits (if N is the length of the register). The crucial point is that this opens the possibility for correlations and entanglement between subsystems. In particular, entanglement is of great importance, because it is a central resource in many applications of quantum information theory like entanglement enhanced teleportation or quantum computing—we already discussed this in Section 1.2 of the Introduction. To explain entanglement in greater detail and to introduce some necessary formalism we have to complement the scheme developed in the last section by a procedure which allows us to construct states and observables of the composite system from its subsystems. In quantum mechanics this is done, of course, in terms of tensor products, and we will review in the following some of the most relevant material.

2.2.1. Tensor products

Consider two (finite dimensional) Hilbert spaces $\mathscr{H}$ and $\mathscr{K}$. To each pair of vectors $\psi_1 \in \mathscr{H}$, $\psi_2 \in \mathscr{K}$ we can associate a bilinear form $\psi_1 \otimes \psi_2$ called the *tensor product* of ψ_1 and ψ_2 by $\psi_1 \otimes \psi_2(\phi_1, \phi_2) = \langle \psi_1, \phi_1 \rangle \langle \psi_2, \phi_2 \rangle$. For two product vectors $\psi_1 \otimes \psi_2$ and $\eta_1 \otimes \eta_2$ their scalar product is defined by $\langle \psi_1 \otimes \psi_2, \eta_1 \otimes \eta_2 \rangle = \langle \psi_1, \eta_1 \rangle \langle \psi_2, \eta_2 \rangle$ and it can be shown that this definition extends in a unique way to the span of all $\psi_1 \otimes \psi_2$ which therefore defines the tensor product $\mathscr{H} \otimes \mathscr{K}$. If we have more than two Hilbert spaces $\mathscr{H}_j$, $j = 1, \dots, N$ their tensor product $\mathscr{H}_1 \otimes \cdots \otimes \mathscr{H}_N$ can be defined similarly.

The tensor product $A_1 \otimes A_2$ of two bounded operators $A_1 \in \mathscr{B}(\mathscr{H})$, $A_2 \in \mathscr{B}(\mathscr{K})$ is defined first for product vectors $\psi_1 \otimes \psi_2 \in \mathscr{H} \otimes \mathscr{K}$ by $A_1 \otimes A_2(\psi_1 \otimes \psi_2) = (A_1\psi_1) \otimes (A_2\psi_2)$ and then extended by linearity. The space $\mathscr{B}(\mathscr{H} \otimes \mathscr{K})$ coincides with the span of all $A_1 \otimes A_2$. If $\rho \in \mathscr{B}(\mathscr{H} \otimes \mathscr{K})$ is not of product form (and of trace class for infinite-dimensional $\mathscr{H}$ and $\mathscr{K}$) there is nevertheless a way to define "restrictions" to $\mathscr{H}$, respectively, $\mathscr{K}$ called the *partial trace* of ρ. It is defined by the equation

$$\operatorname{tr}[\operatorname{tr}_{\mathscr{K}}(\rho)A] = \operatorname{tr}(\rho A \otimes \mathbb{1}) \quad \forall A \in \mathscr{B}(\mathscr{H}) , \tag{2.7}$$

where the trace on the left-hand side is over $\mathscr{H}$ and on the right-hand side over $\mathscr{H} \otimes \mathscr{K}$.

If two orthonormal bases $\phi_1, \dots, \phi_n$ and $\psi_1, \dots, \psi_m$ are given in $\mathscr{H}$, respectively, $\mathscr{K}$ we can consider the product basis $\phi_1 \otimes \psi_1, \dots, \phi_n \otimes \psi_m$ in $\mathscr{H} \otimes \mathscr{K}$, and we can expand each $\Psi \in \mathscr{H} \otimes \mathscr{K}$ as $\Psi = \sum_{jk} \Psi_{jk} \phi_j \otimes \psi_k$ with $\Psi_{jk} = \langle \phi_j \otimes \psi_k, \Psi \rangle$. This procedure works for an arbitrary number of tensor factors. However, if we have exactly a twofold tensor product, there is a more economic way to expand Ψ, called *Schmidt decomposition* in which only diagonal terms of the form $\phi_j \otimes \psi_j$ appear.

Proposition 2.2. *For each element Ψ of the twofold tensor product $\mathscr{H} \otimes \mathscr{K}$ there are orthonormal systems ϕ_j, $j=1,\dots,n$ and ψ_k, $k=1,\dots,n$ (not necessarily bases, i.e. n can be smaller than $\dim \mathscr{H}$ and $\dim \mathscr{K}$) of $\mathscr{H}$ and $\mathscr{K}$, respectively, such that $\Psi = \sum_j \sqrt{\lambda}_j \phi_j \otimes \psi_j$ holds. The ϕ_j and ψ_j are uniquely determined by Ψ. The expansion is called Schmidt decomposition and the numbers $\sqrt{\lambda}_j$ are the Schmidt coefficients.*

Proof. Consider the partial trace $\rho_1 = \operatorname{tr}_{\mathscr{K}}(|\Psi\rangle\langle\Psi|)$ of the one-dimensional projector $|\Psi\rangle\langle\Psi|$ associated to Ψ. It can be decomposed in terms of its eigenvectors ϕ_n and we get $\operatorname{tr}_{\mathscr{K}}(|\Psi\rangle\langle\Psi|) = \rho_1 = \sum_n \lambda_n |\phi_n\rangle\langle\phi_n|$. Now we can choose an orthonormal basis ψ'_k, $k=1,\dots,m$ in $\mathscr{K}$ and expand Ψ with respect to $\phi_j \otimes \psi'_k$. Carrying out the k summation we get a family of vectors $\psi''_j = \sum_k \langle \Psi, \phi_j \otimes \psi'_k\rangle \psi'_k$ with the property $\Psi = \sum_j \phi_j \otimes \psi''_j$. Now we can calculate the partial trace and get for any $A \in \mathscr{B}(\mathscr{H}_1)$:

$$\sum_j \lambda_j \langle \phi_j, A\phi_j\rangle = \operatorname{tr}(\rho_1 A) = \langle \Psi, (A \otimes \mathbb{1})\Psi\rangle = \sum_{j,k} \langle \phi_j, A\phi_k\rangle \langle \psi''_j, \psi''_k\rangle \ . \tag{2.8}$$

Since A is arbitrary we can compare the left- and right-hand side of this equation term by term and we get $\langle \psi''_j, \psi''_k\rangle = \delta_{jk}\lambda_j$. Hence, $\psi_j = \lambda_j^{-1/2} \psi''_j$ is the desired orthonormal system. □

As an immediate application of this result we can show that each mixed state $\rho \in \mathscr{B}^*(\mathscr{H})$ (of the quantum system $\mathscr{B}(\mathscr{H})$) can be regarded as a pure state on a larger Hilbert space $\mathscr{H} \otimes \mathscr{H}'$. We just have to consider the eigenvalue expansion $\rho = \sum_j \lambda_j |\phi_j\rangle\langle\phi_j|$ of ρ and to choose an arbitrary orthonormal system ψ_j, $j=1,\dots n$ in $\mathscr{H}'$. Using Proposition 2.2 we get

Corollary 2.3. *Each state $\rho \in \mathscr{B}^*(\mathscr{H})$ can be extended to a pure state Ψ on a larger system with Hilbert space $\mathscr{H} \otimes \mathscr{H}'$ such that $\operatorname{tr}_{\mathscr{H}'} |\Psi\rangle\langle\Psi| = \rho$ holds.*

2.2.2. Compound and hybrid systems

To discuss the composition of two arbitrary (i.e. classical or quantum) systems it is very convenient to use the scheme developed in Section 2.1.1 and to talk about the two subsystems in terms of their observable algebras $\mathscr{A} \subset \mathscr{B}(\mathscr{H})$ and $\mathscr{B} \subset \mathscr{B}(\mathscr{K})$. The observable algebra of the composite system is then simply given by the tensor product of $\mathscr{A}$ and $\mathscr{B}$, i.e.

$$\mathscr{A} \otimes \mathscr{B} := \operatorname{span}\{A \otimes B \,|\, A \in \mathscr{A},\ B \in \mathscr{B}\} \subset \mathscr{B}(\mathscr{K} \otimes \mathscr{H}) \ . \tag{2.9}$$

The dual of $\mathscr{A} \otimes \mathscr{B}$ is generated by product states, $(\rho \otimes \sigma)(A \otimes B) = \rho(A)\sigma(B)$ and we therefore write $\mathscr{A}^* \otimes \mathscr{B}^*$ for $(\mathscr{A} \otimes \mathscr{B})^*$.

The interpretation of the composed system $\mathscr{A} \otimes \mathscr{B}$ in terms of states and effects is straightforward and therefore postponed to the next subsection. We will consider first the special cases arising from different choices for $\mathscr{A}$ and $\mathscr{B}$. If both systems are quantum ($\mathscr{A} = \mathscr{B}(\mathscr{H})$ and $\mathscr{B} = \mathscr{B}(\mathscr{K})$) we get

$$\mathscr{B}(\mathscr{H}) \otimes \mathscr{B}(\mathscr{K}) = \mathscr{B}(\mathscr{H} \otimes \mathscr{K}) \tag{2.10}$$

as expected. For two classical systems $\mathscr{A} = \mathscr{C}(X)$ and $\mathscr{B} = \mathscr{C}(Y)$ recall that elements of $\mathscr{C}(X)$ (respectively, $\mathscr{C}(Y)$) are complex-valued functions on X (on Y). Hence, the tensor product $\mathscr{C}(X) \otimes \mathscr{C}(Y)$ consists of complex-valued functions on $X \times Y$, i.e. $\mathscr{C}(X) \otimes \mathscr{C}(Y) = \mathscr{C}(X \times Y)$. In other words, states and observables of the composite system $\mathscr{C}(X) \otimes \mathscr{C}(Y)$ are, in accordance with classical probability theory, given by probability distributions and random variables on the Cartesian product $X \times Y$.

If only one subsystem is classical and the other is quantum; e.g. a microparticle interacting with a classical measuring device we have a hybrid system. The elements of its observable algebra $\mathscr{C}(X) \otimes \mathscr{B}(\mathscr{H})$ can be regarded as operator-valued functions on X, i.e. $X \ni x \mapsto A_x \in \mathscr{B}(\mathscr{H})$ and A is an effect iff $0 \leqslant A_x \leqslant \mathbb{1}$ holds for all $x \in X$. The elements of the dual $\mathscr{C}^*(X) \otimes \mathscr{B}^*(\mathscr{H})$ are in a similar way $\mathscr{B}^*(X)$-valued functions $X \ni x \mapsto \rho_x \in \mathscr{B}^*(\mathscr{H})$ and ρ is a state iff each ρ_x is a positive trace class operator on $\mathscr{H}$ and $\sum_x \rho_x = 1$. The probability to measure the effect A in the state ρ is $\sum_x \rho_x(A_x)$.

2.2.3. Correlations and entanglement

Let us now consider two effects $A \in \mathscr{A}$ and $B \in \mathscr{B}$ then $A \otimes B$ is an effect of the composite system $\mathscr{A} \otimes \mathscr{B}$. It is interpreted as the joint measurement of A on the first and B on the second subsystem, where the "yes" outcome means "both effects give yes". In particular, $A \otimes \mathbb{1}$ means to measure A on the first subsystem and to ignore the second one completely. If ρ is a state of $\mathscr{A} \otimes \mathscr{B}$ we can define its *restrictions* by $\rho^{\mathscr{A}}(A) = \rho(A \otimes \mathbb{1})$ and $\rho^{\mathscr{B}}(A) = \rho(\mathbb{1} \otimes A)$. If both systems are quantum the restrictions of ρ are the partial traces, while in the classical case we have to sum over the $\mathscr{B}$, respectively $\mathscr{A}$, variables. For two states $\rho_1 \in \mathscr{S}(\mathscr{A})$ and $\rho_2 \in \mathscr{S}(\mathscr{B})$ there is always a state ρ of $\mathscr{A} \otimes \mathscr{B}$ such that $\rho_1 = \rho^{\mathscr{A}}$ and $\rho_2 = \rho^{\mathscr{B}}$ holds: We just have to choose the product state $\rho_1 \otimes \rho_2$. However, in general, we have $\rho \neq \rho^{\mathscr{A}} \otimes \rho^{\mathscr{B}}$ which means nothing else then ρ also contains *correlations* between the two subsystems.

Definition 2.4. A state ρ of a bipartite system $\mathscr{A} \otimes \mathscr{B}$ is called *correlated* if there are some $A \in \mathscr{A}$, $B \in \mathscr{B}$ such that $\rho(A \otimes B) \neq \rho^{\mathscr{A}}(A)\rho^{\mathscr{B}}(B)$ holds.

We immediately see that $\rho = \rho_1 \otimes \rho_2$ implies $\rho(A \otimes B) = \rho_1(A)\rho_2(B) = \rho^{\mathscr{A}}(A)\rho^{\mathscr{B}}(B)$ hence ρ is not correlated. If on the other hand $\rho(A \otimes B) = \rho^{\mathscr{A}}(A)\rho^{\mathscr{B}}(B)$ holds we get $\rho = \rho^{\mathscr{A}} \otimes \rho^{\mathscr{B}}$. Hence, the definition of correlations just given perfectly fits into our intuitive considerations.

An important issue in quantum information theory is the comparison of correlations between quantum systems on the one hand and classical systems on the other. Hence, let us have a closer look on the state space of a system consisting of at least one classical subsystem.

Proposition 2.5. *Each state ρ of a composite system $\mathscr{A} \otimes \mathscr{B}$ consisting of a classical ($\mathscr{A} = \mathscr{C}(X)$) and an arbitrary system ($\mathscr{B}$) has the form*

$$\rho = \sum_{j \in X} \lambda_j \rho_j^{\mathscr{A}} \otimes \rho_j^{\mathscr{B}} \tag{2.11}$$

with positive weights $\lambda_j > 0$ and $\rho_j^{\mathscr{A}} \in \mathscr{S}(\mathscr{A})$, $\rho_j^{\mathscr{B}} \in \mathscr{S}(\mathscr{B})$.

Proof. Since $\mathscr{A} = \mathscr{C}(X)$ is classical, there is a basis $|j\rangle\langle j| \in \mathscr{A}$, $j \in X$ of mutually orthogonal one-dimensional projectors and we can write each $A \in \mathscr{A}$ as $\sum_j a_j |j\rangle\langle j|$ (cf. Subsection 2.1.3). For each state $\rho \in \mathscr{S}(\mathscr{A} \otimes \mathscr{B})$ we can now define $\rho_j^{\mathscr{A}} \in \mathscr{S}(\mathscr{A})$ with $\rho_j^{\mathscr{A}}(A) = \operatorname{tr}(A|j\rangle\langle j|) = a_j$ and

$\rho_j^{\mathscr{B}} \in \mathscr{S}(\mathscr{B})$ with $\rho_j^{\mathscr{B}}(B) = \lambda_j^{-1}\rho(|j\rangle\langle j| \otimes B)$ and $\lambda_j = \rho(|j\rangle\langle j| \otimes \mathbb{1})$. Hence we get $\rho = \sum_{j\in X} \lambda_j \rho_j^{\mathscr{A}} \otimes \rho_j^{\mathscr{B}}$ with positive λ_j as stated. □

If $\mathscr{A}$ and $\mathscr{B}$ are two quantum systems it is still possible for them to be correlated in the way just described. We can simply prepare them with a classical random generator which triggers two preparation devices to produce systems in the states ρ_j^A, ρ_j^B with probability λ_j. The overall state produced by this setup is obviously the ρ from Eq. (2.11). However, the crucial point is that *not all* correlations of quantum systems are of this type! This is an immediate consequence of the definition of pure states $\rho = |\Psi\rangle\langle\Psi| \in \mathscr{S}(\mathscr{H})$: Since there is no proper convex decomposition of ρ, it can be written as in Proposition 2.5 iff Ψ is a product vector, i.e. $\Psi = \phi \otimes \psi$. This observation motivates the following definition.

Definition 2.6. A state ρ of the composite system $\mathscr{B}(\mathscr{H}_1) \otimes \mathscr{B}(\mathscr{H}_2)$ is called *separable* or *classically correlated* if it can be written as

$$\rho = \sum_j \lambda_j \rho_j^{(1)} \otimes \rho_j^{(2)} \tag{2.12}$$

with states $\rho_j^{(k)}$ of $\mathscr{B}(\mathscr{H}_k)$ and weights $\lambda_j > 0$. Otherwise ρ is called *entangled*. The set of all separable states is denoted by $\mathscr{D}(\mathscr{H}_1 \otimes \mathscr{H}_2)$ or just $\mathscr{D}$ if $\mathscr{H}_1$ and $\mathscr{H}_2$ are understood.

2.2.4. Bell inequalities

We have just seen that it is quite easy for pure states to check whether they are entangled or not. In the mixed case however this is a much bigger, and in general unsolved, problem. In this subsection we will have a short look at the Bell inequalities, which are maybe the oldest criterion for entanglement (for a more detailed review see [169]). Today more powerful methods, most of them based on positivity properties, are available. We will postpone the corresponding discussion to the end of the following section, after we have studied (completely) positive maps (cf. Section 2.4).

Bell inequalities are traditionally discussed in the framework of “local hidden variable theories”. More precisely we will say that a state ρ of a bipartite system $\mathscr{B}(\mathscr{H} \otimes \mathscr{K})$ admits a hidden variable model, if there is a probability space (X, μ) and (measurable) response functions $X \ni x \mapsto F_A(x,k)$, $F_B(x,l) \in \mathbb{R}$ for all discrete PV measures $A = A_1, \dots, A_N \in \mathscr{B}(\mathscr{H})$, respectively $B = B_1, \dots, B_M \in \mathscr{B}(\mathscr{K})$, such that

$$\int_X F_A(x,k) F_B(x,l) \mu(\mathrm{d}x) = \mathrm{tr}(\rho A_k \otimes B_l) \tag{2.13}$$

holds for all, k, l and A, B. The value of the functions $F_A(x,k)$ is interpreted as the probability to get the value k during an A measurement with known “hidden parameter” x. The set of states admitting a hidden variable model is a convex set and as such it can be described by an (infinite) hierarchy of correlation inequalities. Any one of these inequalities is usually called (generalized) Bell inequality. The most well-known one is those given by Clauser et al. [47]: The state ρ satisfies the CHSH-inequality if

$$\rho(A \otimes (B + B') + A' \otimes (B - B')) \leqslant 2 \tag{2.14}$$

holds for all $A, A' \in \mathscr{B}(\mathscr{H})$, respectively $B, B' \in \mathscr{B}(\mathscr{K})$, with $-\mathbb{1} \leqslant A, A' \leqslant \mathbb{1}$ and $-\mathbb{1} \leqslant B, B' \leqslant \mathbb{1}$. For the special case of two dichotomic observables the CHSH inequalities are sufficient to characterize the states with a hidden variable model. In the general case the CHSH inequalities are a necessary but not a sufficient condition and a complete characterization is not known.

It is now easy to see that each separable state $\rho = \sum_{j=1}^{n} \lambda_j \rho_j^{(1)} \otimes \rho_j^{(2)}$ admits a hidden variable model: we have to choose $X = 1, \ldots, n$, $\mu(\{j\}) = \lambda_j$, $F_A(x,k) = \rho_x^{(1)}(A_k)$ and F_B analogously. Hence, we immediately see that each state of a composite system with at least one classical subsystem satisfies the Bell inequalities (in particular the CHSH version) while this is not the case for pure quantum systems. The most prominent examples are "maximally entangled states" (cf. Subsection 3.1.1) which violate the CHSH inequality (for appropriately chosen A, A', B, B') with a maximal value of $2\sqrt{2}$. This observation is the starting point for many discussions concerning the interpretation of quantum mechanics, in particular because the maximal violation of $2\sqrt{2}$ was observed in 1982 experimentally by Aspect and coworkers [5]. We do not want to follow this path (see [169] and the references therein instead). Interesting for us is the fact that Bell inequalities, in particular the CHSH case in Eq. (2.14), provide a *necessary condition* for a state ρ to be separable. However, there exist entangled states admitting a hidden variable model [165]. Hence, Bell inequalities are not sufficient for separability.

2.3. Channels

Assume now that we have a number of quantum systems, e.g. a string of ions in a trap. To "process" the quantum information they carry we have to perform, in general, many steps of a quite different nature. Typical examples are: free time evolution, controlled time evolution (e.g. the application of a "quantum gate" in a quantum computer), preparations and measurements. The purpose of this section is to provide a unified framework for the description of all these different operations. The basic idea is to represent each processing step by a "channel", which converts input systems, described by an observable algebra $\mathscr{A}$ into output systems described by a possibly different algebra $\mathscr{B}$. Henceforth we will call $\mathscr{A}$ the *input* and $\mathscr{B}$ the *output algebra*. If we consider e.g. the free time evolution, we need quantum systems of the same type on the input and the output side; hence, in this case we have $\mathscr{A} = \mathscr{B} = \mathscr{B}(\mathscr{H})$ with an appropriately chosen Hilbert space $\mathscr{H}$. If on the other hand, we want to describe a measurement we have to map quantum systems (the measured system) to classical information (the measuring result). Therefore, we need in this example $\mathscr{A} = \mathscr{B}(\mathscr{H})$ for the input and $\mathscr{B} = \mathscr{C}(X)$ for the output algebra, where X is the set of possible outcomes of the measurement (cf. Section 2.1.4).

Our aim is now to get a mathematical object which can be used to describe a channel. To this end consider an effect $A \in \mathscr{B}$ of the output system. If we invoke first a channel which transforms $\mathscr{A}$ systems into $\mathscr{B}$ systems, and measure A afterwards on the output systems, we end up with a measurement of an effect $T(A)$ on the input systems. Hence, we get a map $T : \mathscr{E}(\mathscr{B}) \to \mathscr{E}(\mathscr{A})$ which *completely describes the channel.*[6] Alternatively, we can look at the states and interpret a channel as a map $T^* : \mathscr{S}(\mathscr{A}) \to \mathscr{S}(\mathscr{B})$ which transforms $\mathscr{A}$ systems in the state $\rho \in \mathscr{S}(\mathscr{A})$ into $\mathscr{B}$ systems in the state $T^*(\rho)$. To distinguish between both maps we can say that T describes the channel in the *Heisenberg picture* and T^* in the *Schrödinger picture*. On the level of the statistical

[6] Note that the direction of the mapping arrow is reversed compared to the natural ordering of processing.

interpretation both points of view should coincide of course, i.e. the probabilities[7] $(T^*\rho)(A)$ and $\rho(TA)$ to get the result "yes" during an A measurement on $\mathscr{B}$ systems in the state $T^*\rho$, respectively, a TA measurement on $\mathscr{A}$ systems in the state ρ, should be the same. Since $(T^*\rho)(A)$ is linear in A we see immediately that T must be an *affine map*, i.e. $T(\lambda_1 A_1 + \lambda_2 A_2) = \lambda_1 T(A_1) + \lambda_2 T(A_2)$ for each convex linear combination $\lambda_1 A_1 + \lambda_2 A_2$ of effects in $\mathscr{B}$, and this in turn implies that T can be extended naturally to a *linear map*, which we will identify in the following with the channel itself, i.e. we say that T *is* the channel.

2.3.1. Completely positive maps

Let us change now slightly our point of view and start with a linear operator $T : \mathscr{A} \to \mathscr{B}$. To be a channel, T must map effects to effects, i.e. T has to be positive: $T(A) \geqslant 0\ \forall A \geqslant 0$ and bounded from above by $\mathbb{1}$, i.e. $T(\mathbb{1}) \leqslant \mathbb{1}$. In addition it is natural to require that two channels in parallel are again a channel. More precisely, if two channels $T : \mathscr{A}_1 \to \mathscr{B}_1$ and $S : \mathscr{A}_2 \to \mathscr{B}_2$ are given we can consider the map $T \otimes S$ which associates to each $A \otimes B \in \mathscr{A}_1 \otimes \mathscr{A}_2$ the tensor product $T(A) \otimes S(B) \in \mathscr{B}_1 \otimes \mathscr{B}_2$. It is natural to assume that $T \otimes S$ is a channel which converts composite systems of type $\mathscr{A}_1 \otimes \mathscr{A}_2$ into $\mathscr{B}_1 \otimes \mathscr{B}_2$ systems. Hence $S \otimes T$ should be positive as well [125].

Definition 2.7. Consider two observable algebras $\mathscr{A}$, $\mathscr{B}$ and a linear map $T : \mathscr{A} \to \mathscr{B} \subset \mathscr{B}(\mathscr{H})$.

1. T is called *positive* if $T(A) \geqslant 0$ holds for all positive $A \in \mathscr{A}$.
2. T is called *completely positive* (cp) if $T \otimes \mathrm{Id} : \mathscr{A} \otimes \mathscr{B}(\mathbb{C}^n) \to \mathscr{B}(\mathscr{H}) \otimes \mathscr{B}(\mathbb{C}^n)$ is positive for all $n \in \mathbb{N}$. Here Id denotes the identity map on $\mathscr{B}(\mathbb{C}^n)$.
3. T is called *unital* if $T(\mathbb{1}) = \mathbb{1}$ holds.

Consider now the map $T^* : \mathscr{B}^* \to \mathscr{A}^*$ which is *dual* to T, i.e. $T^*\rho(A) = \rho(TA)$ for all $\rho \in \mathscr{B}^*$ and $A \in \mathscr{A}$. It is called the Schrödinger picture representation of the channel T, since it maps states to states provided T is unital. (Complete) positivity can be defined in the Schrödinger picture as in the Heisenberg picture and we immediately see that T is (completely) positive iff T^* is.

It is natural to ask whether the distinction between positivity and complete positivity is really necessary, i.e. whether there are positive maps which are not completely positive. If at least one of the algebras $\mathscr{A}$ or $\mathscr{B}$ is classical the answer is no: each positive map is completely positive in this case. If both algebras are quantum, however, complete positivity is *not implied* by positivity alone. We will discuss explicit examples in Section 2.4.2.

If item 2 holds only for a fixed $n \in \mathbb{N}$ the map T is called *n-positive*. This is obviously a weaker condition than complete positivity. However, n-positivity implies m-positivity for all $m \leqslant n$, and for $\mathscr{A} = \mathscr{B}(\mathbb{C}^d)$ complete positivity is implied by n-positivity, provided $n \geqslant d$ holds.

Let us consider now the question whether a channel should be unital or not. We have already mentioned that $T(\mathbb{1}) \leqslant \mathbb{1}$ must hold since effects should be mapped to effects. If $T(\mathbb{1})$ is not equal to $\mathbb{1}$ we get $\rho(T\mathbb{1}) = T^*\rho(\mathbb{1}) < 1$ for the probability to measure the effect $\mathbb{1}$ on systems in the state $T^*\rho$, but this is impossible for channels *which produce an output with certainty*, because $\mathbb{1}$ is the

[7] To keep notations more readable we will follow frequently the usual convention to drop the parenthesis around arguments of linear operators. Hence, we will write TA and $T^*\rho$ instead of $T(A)$ and $T^*(\rho)$. Similarly, we will simply write TS instead of $T \circ S$ for compositions.

effect which is always true. In other words: If a cp map is not unital it describes a channel which sometimes produces no output at all and $T(\mathbb{1})$ is the effect which measures whether we have got an output. We will assume in the future that channels are unital if nothing else is explicitly stated.

2.3.2. The Stinespring theorem

Consider now channels between quantum systems, i.e. $\mathscr{A} = \mathscr{B}(\mathscr{H}_1)$ and $\mathscr{B} = \mathscr{B}(\mathscr{H}_2)$. A fairly simple example (not necessarily unital) is given in terms of an operator $V : \mathscr{H}_1 \to \mathscr{H}_2$ by $\mathscr{B}(\mathscr{H}_1) \ni A \mapsto VAV^* \in \mathscr{B}(\mathscr{H}_2)$. A second example is the restriction to a subsystem, which is given in the Heisenberg picture by $\mathscr{B}(\mathscr{H}) \ni A \mapsto A \otimes \mathbb{1}_{\mathscr{K}} \in \mathscr{B}(\mathscr{H} \otimes \mathscr{K})$. Finally, the composition $S \circ T = ST$ of two channels is again a channel. The following theorem, which is the most fundamental structural result about cp maps,[8] says that each channel can be represented as a composition of these two examples [147].

Theorem 2.8 (Stinespring dilation theorem). *Every completely positive map $T : \mathscr{B}(\mathscr{H}_1) \to \mathscr{B}(\mathscr{H}_2)$ has the form*

$$T(A) = V^*(A \otimes \mathbb{1}_{\mathscr{K}})V \ , \tag{2.15}$$

with an additional Hilbert space $\mathscr{K}$ and an operator $V : \mathscr{H}_2 \to \mathscr{H}_1 \otimes \mathscr{K}$. Both (i.e. $\mathscr{K}$ and V) can be chosen such that the span of all $(A \otimes \mathbb{1})V\phi$ with $A \in \mathscr{B}(\mathscr{H}_1)$ and $\phi \in \mathscr{H}_2$ is dense in $\mathscr{H}_1 \otimes \mathscr{K}$. This particular decomposition is unique (up to unitary equivalence) and called the minimal decomposition. If $\dim \mathscr{H}_1 = d_1$ *and* $\dim \mathscr{H}_2 = d_2$ *the minimal $\mathscr{K}$ satisfies* $\dim \mathscr{K} \leqslant d_1^2 d_2$.

By introducing a family $|\chi_j\rangle\langle\chi_j|$ of one-dimensional projectors with $\sum_j |\chi_j\rangle\langle\chi_j| = \mathbb{1}$ we can define the "Kraus operators" $\langle\psi, V_j\phi\rangle = \langle\psi \otimes \chi_j, V\phi\rangle$. In terms of them we can rewrite Eq. (2.15) in the following form [105]:

Corollary 2.9 (Kraus form). *Every completely positive map $T : \mathscr{B}(\mathscr{H}_1) \to \mathscr{B}(\mathscr{H}_2)$ can be written in the form*

$$T(A) = \sum_{j=1}^{N} V_j^* A V_j \tag{2.16}$$

with operators $V_j : \mathscr{H}_2 \to \mathscr{H}_1$ and $N \leqslant \dim(\mathscr{H}_1)\dim(\mathscr{H}_2)$.

2.3.3. The duality lemma

We will consider a fundamental relation between positive maps and bipartite systems, which will allow us later on to translate properties of entangled states to properties of channels and vice versa. The basic idea originates from elementary linear algebra: A bilinear form ϕ on a d-dimensional vector space V can be represented by a $d \times d$-matrix, just as an operator on V. Hence, we can transform ϕ into an operator simply by reinterpreting the matrix elements. In our situation things

[8] Basically, there is a more general version of this theorem which works with arbitrary output algebras. It needs however some material from representation theory of C*-algebras which we want to avoid here. See e.g. [125,83].

are more difficult, because the positivity constraints for states and channels should match up in the right way. Nevertheless, we have the following theorem.

Theorem 2.10. *Let ρ be a density operator on $\mathscr{H} \otimes \mathscr{H}_1$. Then there is a Hilbert space $\mathscr{K}$ a pure state σ on $\mathscr{H} \otimes \mathscr{K}$ and a channel $T : \mathscr{B}(\mathscr{H}_1) \to \mathscr{B}(\mathscr{K})$ with*

$$\rho = (\mathrm{Id} \otimes T^*)\sigma \ , \tag{2.17}$$

where Id *denotes the identity map on $\mathscr{B}^*(\mathscr{H})$. The pure state σ can be chosen such that $\mathrm{tr}_{\mathscr{H}}(\sigma)$ has no zero eigenvalue. In this case T and σ are uniquely determined* (*up to unitary equivalence*) *by Eq.* (2.17); *i.e. if $\tilde{\sigma}$, $\tilde{T}$ with $\rho = (\mathrm{Id} \otimes \tilde{T}^*)\tilde{\sigma}$ are given, we have $\tilde{\sigma} = (\mathbb{1} \otimes U)^* \sigma (\mathbb{1} \otimes U)$ and $\tilde{T}(\cdot) = U^* T(\cdot) U$ with an appropriate unitary operator U.*

Proof. The state σ is obviously the purification of $\mathrm{tr}_{\mathscr{H}_1}(\rho)$. Hence if λ_j and ψ_j are eigenvalues and eigenvectors of $\mathrm{tr}_{\mathscr{H}_1}(\rho)$ we can set $\sigma = |\Psi\rangle\langle\Psi|$ with $\Psi = \sum_j \sqrt{\lambda_j}\psi_j \otimes \phi_j$ where ϕ_j is an (arbitrary) orthonormal basis in $\mathscr{K}$. It is clear that σ is uniquely determined up to a unitary. Hence, we only have to show that a unique T exists if Ψ is given. To satisfy Eq. (2.17) we must have

$$\rho(|\psi_j \otimes \eta_k\rangle\langle\psi_l \otimes \eta_l|) = \langle \Psi, (\mathrm{Id} \otimes T)(|\psi_j \otimes \eta_k\rangle\langle\psi_l \otimes \eta_l|)\Psi\rangle \ , \tag{2.18}$$

$$= \langle \Psi, |\psi_j\rangle\langle\psi_l| \otimes T(|\eta_k\rangle\langle\eta_p|)\Psi\rangle \ , \tag{2.19}$$

$$= \sqrt{\lambda_j \lambda_l}\langle \phi_j, T(|\eta_k\rangle\langle\eta_p|)\phi_l\rangle \tag{2.20}$$

where η_k is an (arbitrary) orthonormal basis in $\mathscr{H}_1$. Hence T is uniquely determined by ρ in terms of its matrix elements and we only have to check complete positivity. To this end it is useful to note that the map $\rho \mapsto T$ is linear if the λ_j are fixed. Hence, it is sufficient to consider the case $\rho = |\chi\rangle\langle\chi|$. Inserting this into Eq. (2.20) we immediately see that $T(A) = V^*AV$ with $\langle V\phi_j, \eta_k\rangle = \lambda_j^{-1/2}\langle\psi_j \otimes \eta_k, \chi\rangle$ holds. Hence T is completely positive. Since normalization $T(\mathbb{1}) = \mathbb{1}$ follows from the choice of the λ_j the theorem is proved. □

2.4. Separability criteria and positive maps

We have already stated in Section 2.3.1 that positive but not completely positive maps exist, whenever input and output algebra are quantum. No such map represents a valid quantum operation, nevertheless they are of great importance in quantum information theory, due to their deep relations to entanglement properties. Hence, this section is a continuation of the study of separability criteria which we have started in Section 2.2.4. In contrast to the rest of this section, all maps are considered in the Schrödinger rather than in the Heisenberg picture.

2.4.1. Positivity

Let us consider now an arbitrary positive, but not necessarily completely positive map $T^* : \mathscr{B}^*(\mathscr{H}) \to \mathscr{B}^*(\mathscr{K})$. If Id again denotes the identity map, it is easy to see that $(\mathrm{Id} \otimes T^*)(\sigma_2 \otimes \sigma_2) = \sigma_1 \otimes T^*(\sigma_2) \geqslant 0$ holds for each product state $\sigma_1 \otimes \sigma_2 \in \mathscr{S}(\mathscr{H} \otimes \mathscr{K})$. Hence $(\mathrm{Id} \otimes T^*)\rho \geqslant 0$ for each positive T^* is a necessary condition for ρ to be separable. The following theorem proved in [86] shows that sufficiency holds as well.

Theorem 2.11. *A state $\rho \in \mathscr{B}^*(\mathscr{H} \otimes \mathscr{K})$ is separable iff for any positive map $T^* : \mathscr{B}^*(\mathscr{K}) \to \mathscr{B}^*(\mathscr{H})$ the operator $(\mathrm{Id} \otimes T^*)\rho$ is positive.*

Proof. We will only give a sketch of the proof, see [86] for details. The condition is obviously necessary since $(\mathrm{Id} \otimes T^*)\rho_1 \otimes \rho_2 \geqslant 0$ holds for any product state provided T^* is positive. The proof of sufficiency relies on the fact that it is always possible to separate a point ρ (an entangled state) from a convex set $\mathscr{D}$ (the set of separable states) by a hyperplane. A precise formulation of this idea leads to the following proposition.

Proposition 2.12. *For any entangled state $\rho \in \mathscr{S}(\mathscr{H} \otimes \mathscr{K})$ there is an operator A on $\mathscr{H} \otimes \mathscr{K}$ called entanglement witness for ρ, with the property $\rho(A) < 0$ and $\sigma(A) \geqslant 0$ for all separable $\sigma \in \mathscr{S}(\mathscr{H} \otimes \mathscr{K})$.*

Proof. Since $\mathscr{D} \subset \mathscr{B}^*(\mathscr{H} \otimes \mathscr{K})$ is a closed convex set, for each $\rho \in \mathscr{S} \subset \mathscr{B}^*(\mathscr{H} \otimes \mathscr{K})$ with $\rho \notin \mathscr{D}$ there exists a linear functional α on $\mathscr{B}^*(\mathscr{H} \otimes \mathscr{K})$, such that $\alpha(\rho) < \gamma \leqslant \alpha(\sigma)$ for each $\sigma \in \mathscr{D}$ with a constant γ. This holds as well in infinite-dimensional Banach spaces and is a consequence of the Hahn–Banach theorem (cf. [135, Theorem 3.4]). Without loss of generality, we can assume that $\gamma = 0$ holds. Otherwise we just have to replace α by $\alpha - \gamma\,\mathrm{tr}$. Hence, the result follows from the fact that each linear functional on $\mathscr{B}^*(\mathscr{H} \otimes \mathscr{K})$ has the form $\alpha(\sigma) = \mathrm{tr}(A\sigma)$ with $A \in \mathscr{B}(\mathscr{H} \otimes \mathscr{K})$. □

To continue the proof of Theorem 2.11 associate now to any operator $A \in \mathscr{B}(\mathscr{H} \otimes \mathscr{K})$ the map $T_A^* : \mathscr{B}^*(\mathscr{K}) \to \mathscr{B}^*(\mathscr{H})$ with

$$\mathrm{tr}(A\rho_1 \otimes \rho_2) = \mathrm{tr}(\rho_1^{\mathrm{T}} T_A^*(\rho_2)) , \tag{2.21}$$

where $(\cdot)^{\mathrm{T}}$ denotes the transposition in an arbitrary but fixed orthonormal basis $|j\rangle$, $j = 1, \ldots, d$. It is easy to see that T_A^* is positive if $\mathrm{tr}(A\rho_1 \otimes \rho_2) \geqslant 0$ for all product states $\rho_1 \otimes \rho_2 \in \mathscr{S}(\mathscr{H} \otimes \mathscr{K})$ [94]. A straightforward calculation [86] shows in addition that

$$\mathrm{tr}(A\rho) = \mathrm{tr}(|\Psi\rangle\langle\Psi|(\mathrm{Id} \otimes T_A^*)(\rho)) \tag{2.22}$$

holds, where $\Psi = d^{-1/2} \sum_j |j\rangle \otimes |j\rangle$. Assume now that $(\mathrm{Id} \otimes T^*)\rho \geqslant 0$ for all positive T^*. Since T_A^* is positive this implies that the left-hand side of (2.22) is positive; hence $\mathrm{tr}(A\rho) \geqslant 0$ provided $\mathrm{tr}(A\sigma) \geqslant 0$ holds for all separable σ, and the statement follows from Proposition 2.12. □

2.4.2. The partial transpose

The most typical example for a positive non-cp map is the transposition $\Theta A = A^{\mathrm{T}}$ of $d \times d$ matrices, which we have just used in the proof of Theorem 2.11. Θ is obviously a positive map, but the *partial transpose*

$$\mathscr{B}^*(\mathscr{H} \otimes \mathscr{K}) \ni \rho \mapsto (\mathrm{Id} \otimes \Theta)(\rho) \in \mathscr{B}^*(\mathscr{H} \otimes \mathscr{K}) \tag{2.23}$$

is not. The latter can be easily checked with the maximally entangled state (cf. Section 3.1.1).

$$\Psi = \frac{1}{\sqrt{d}} \sum_j |j\rangle \otimes |j\rangle , \tag{2.24}$$

where $|j\rangle \in \mathbb{C}^d$, $j = 1, \ldots, d$ denote the canonical basis vectors. In low dimensions the transposition is basically the only positive map which is not cp. Due to results of Størmer [148] and Woronowicz

[174] we have: $\dim \mathscr{H} = 2$ and $\dim \mathscr{K} = 2, 3$ imply that each positive map $T^* : \mathscr{B}^*(\mathscr{H}) \to \mathscr{B}^*(\mathscr{K})$ has the form $T^* = T_1^* + T_2^* \Theta$ with two cp maps T_1^*, T_2^* and the transposition on $\mathscr{B}(\mathscr{H})$. This immediately implies that positivity of the partial transpose is necessary *and sufficient* for separability of a state $\rho \in \mathscr{S}(\mathscr{H} \otimes \mathscr{K})$ (cf. [86]):

Theorem 2.13. *Consider a bipartite system* $\mathscr{B}(\mathscr{H} \otimes \mathscr{K})$ *with* $\dim \mathscr{H} = 2$ *and* $\dim \mathscr{K} = 2, 3$. *A state* $\rho \in \mathscr{S}(\mathscr{H} \otimes \mathscr{K})$ *is separable iff its partial transpose is positive.*

To use positivity of the partial transpose as a separability criterion was proposed for the first time by Peres [127], and he conjectured that it is a necessary and sufficient condition in arbitrary finite dimension. Although it has turned out in the meantime that this conjecture is wrong in general (cf. Section 3.1.5), partial transposition has become a crucial tool within entanglement theory and we define:

Definition 2.14. A state $\rho \in \mathscr{B}^*(\mathscr{H} \otimes \mathscr{K})$ of a bipartite quantum system is called *ppt-state* if $(\mathrm{Id} \otimes \Theta)\rho \geqslant 0$ holds and *npt-state* otherwise (ppt = "positive partial transpose" and npt = "negative partial transpose").

2.4.3. The reduction criterion

Another frequently used example of a non-cp but positive map is $\mathscr{B}^*(\mathscr{H}) \ni \rho \mapsto T^*(\rho) = (\mathrm{tr}\,\rho)\mathbb{1} - \rho \in \mathscr{B}^*(\mathscr{H})$. The eigenvalues of $T^*(\rho)$ are given by $\mathrm{tr}\,\rho - \lambda_i$, where λ_i are the eigenvalues of ρ. If $\rho \geqslant 0$ we have $\lambda_i \geqslant 0$ and therefore $\sum_j \lambda_j - \lambda_k \geqslant 0$. Hence T^* is positive. That T^* is not completely positive follows if we consider again the example $|\Psi\rangle\langle\Psi|$ from Eq. (2.24); hence we get

$$\mathbb{1} \otimes \mathrm{tr}_2(\rho) - \rho \geqslant 0, \quad \mathrm{tr}_1(\rho) \otimes \mathbb{1} - \rho \geqslant 0 \tag{2.25}$$

for any separable state $\rho \in \mathscr{B}^*(\mathscr{H} \otimes \mathscr{K})$. These equations are another non-trivial separability criterion, which is called the *reduction criterion* [85,42]. It is closely related to the ppt criterion, due to the following proposition (see [85] for a proof).

Proposition 2.15. *Each ppt-state* $\rho \in \mathscr{S}(\mathscr{H} \otimes \mathscr{K})$ *satisfies the reduction criterion. If* $\dim \mathscr{H} = 2$ *and* $\dim \mathscr{K} = 2, 3$ *both criteria are equivalent.*

Hence we see with Theorem 2.13 that a state ρ in 2×2 or 2×3 dimensions is separable iff it satisfies the reduction criterion.

3. Basic examples

After the somewhat abstract discussion in the last section we will become more concrete now. In the following, we will present a number of examples which help on the one hand to understand the structures just introduced, and which are of fundamental importance within quantum information on the other.

3.1. Entanglement

Although our definition of entanglement (Definition 2.6) is applicable in arbitrary dimensions, detailed knowledge about entangled states is available only for low-dimensional systems or for states with very special properties. In this section we will discuss some of the most basic examples.

3.1.1. Maximally entangled states

Let us start with a look on pure states of a composite systems $\mathscr{A}\otimes\mathscr{B}$ and their possible correlations. If one subsystem is classical, i.e. $\mathscr{A}=\mathscr{C}(\{1,\dots,d\})$, the state space is given according to Section 2.2.2 by $\mathscr{S}(\mathscr{B})^d$ and $\rho\in\mathscr{S}(\mathscr{B})^d$ is pure iff $\rho=(\delta_{j1}\tau,\dots,\delta_{jd}\tau)$ with $j=1,\dots,d$ and a pure state τ of the $\mathscr{B}$ system. Hence, the restrictions of ρ to $\mathscr{A}$, respectively, $\mathscr{B}$ are the Dirac measure $\delta_j\in\mathscr{S}(X)$ or $\tau\in\mathscr{S}(\mathscr{B})$, in other words both restrictions are pure. This is completely different if $\mathscr{A}$ and $\mathscr{B}$ are quantum, i.e. $\mathscr{A}\otimes\mathscr{B}=\mathscr{B}(\mathscr{H}\otimes\mathscr{K})$: Consider $\rho=|\Psi\rangle\langle\Psi|$ with $\Psi\in\mathscr{H}\otimes\mathscr{K}$ and Schmidt decomposition (Proposition 2.2) $\Psi=\sum_j\lambda_j^{1/2}\phi_j\otimes\psi_j$. Calculating the $\mathscr{A}$ restriction, i.e. the partial trace over $\mathscr{K}$ we get

$$\mathrm{tr}[\mathrm{tr}_{\mathscr{K}}(\rho)A]=\mathrm{tr}[|\Psi\rangle\langle\Psi|A\otimes\mathbb{1}]=\sum_{jk}\lambda_j^{1/2}\lambda_k^{1/2}\langle\phi_j,A\phi_k\rangle\delta_{jk}\ , \tag{3.1}$$

hence $\mathrm{tr}_{\mathscr{K}}(\rho)=\sum_j\lambda_j|\phi_j\rangle\langle\phi_j|$ is mixed iff Ψ is entangled. The most extreme case arises if $\mathscr{H}=\mathscr{K}=\mathbb{C}^d$ and $\mathrm{tr}_{\mathscr{K}}(\rho)$ is maximally mixed, i.e. $\mathrm{tr}_{\mathscr{K}}(\rho)=\mathbb{1}/d$. We get for Ψ

$$\Psi=\frac{1}{\sqrt{d}}\sum_{j=1}^{d}\phi_j\otimes\psi_j \tag{3.2}$$

with two orthonormal bases $\phi_1,\dots,\phi_d$ and $\psi_1,\dots,\psi_d$. In $2n\times 2n$ dimensions these states violate maximally the CHSH inequalities, with appropriately chosen operators A,A',B,B'. Such states are therefore called *maximally entangled*. The most prominent examples of maximally entangled states are the four "Bell states" for two qubit systems, i.e. $\mathscr{H}=\mathscr{K}=\mathbb{C}^2$, $|1\rangle,|0\rangle$ denotes the canonical basis and

$$\Phi_0=\frac{1}{\sqrt{2}}(|11\rangle+|00\rangle),\quad \Phi_j=i(\mathbb{1}\otimes\sigma_j)\Phi_0,\quad j=1,2,3\ , \tag{3.3}$$

where we have used the shorthand notation $|jk\rangle$ for $|j\rangle\otimes|k\rangle$ and the σ_j denote the Pauli matrices.

The Bell states, which form an orthonormal basis of $\mathbb{C}^2\otimes\mathbb{C}^2$, are the best studied and most relevant examples of entangled states within quantum information. A mixture of them, i.e. a density matrix $\rho\in\mathscr{S}(\mathbb{C}^2\otimes\mathbb{C}^2)$ with eigenvectors Φ_j and eigenvalues $0\leqslant\lambda_j\leqslant 1$, $\sum_j\lambda_j=1$, is called a *Bell diagonal state*. It can be shown [16] that ρ is entangled iff $\max_j\lambda_j>\frac{1}{2}$ holds. We omit the proof of this statement here, but we will come back to this point in Section 5 within the discussion of entanglement measures.

Let us come back to the general case now and consider an arbitrary $\rho\in\mathscr{S}(\mathscr{H}\otimes\mathscr{H})$. Using maximally entangled states, we can introduce another separability criterion in terms of the *maximally entangled fraction* (cf. [16])

$$\mathscr{F}(\rho)=\sup_{\Psi\ \mathrm{max.\,ent.}}\langle\Psi,\rho\Psi\rangle\ . \tag{3.4}$$

If ρ is separable the reduction criterion (2.25) implies $\langle \Psi, [\mathrm{tr}_1(\rho) \otimes \mathbb{1} - \rho]\Psi \rangle \geqslant 0$ for any maximally entangled state. Since the partial trace of $|\Psi\rangle\langle\Psi|$ is $d^{-1}\mathbb{1}$ we get

$$d^{-1} = \langle \Psi, \mathrm{tr}_1(\rho) \otimes \mathbb{1}\Psi \rangle \leqslant \langle \Psi, \rho\Psi \rangle , \tag{3.5}$$

hence $\mathscr{F}(\rho) \leqslant 1/d$. This condition is not very sharp however. Using the ppt criterion it can be shown that $\rho = \lambda|\Phi_1\rangle\langle\Phi_1| + (1-\lambda)|00\rangle\langle 00\rangle$ (with the Bell state Φ_1) is entangled for all $0 < \lambda \leqslant 1$ but a straightforward calculation shows that $\mathscr{F}(\rho) \leqslant 1/2$ holds for $\lambda \leqslant 1/2$.

Finally, we have to mention here a very useful parameterization of the set of pure states on $\mathscr{H} \otimes \mathscr{H}$ in terms of maximally entangled states: If Ψ is an arbitrary but fixed maximally entangled state, each $\phi \in \mathscr{H} \otimes \mathscr{H}$ admits (uniquely determined) operators X_1, X_2 such that

$$\phi = (X_1 \otimes \mathbb{1})\Psi = (\mathbb{1} \otimes X_2)\Psi \tag{3.6}$$

holds. This can be easily checked in a product basis.

3.1.2. Werner states

If we consider entanglement of mixed states rather than pure ones, the analysis becomes quite difficult, even if the dimensions of the underlying Hilbert spaces are low. The reason is that the state space $\mathscr{S}(\mathscr{H}_1 \otimes \mathscr{H}_2)$ of a two-partite system with $\dim \mathscr{H}_i = d_i$ is a geometric object in a $(d_1^2 d_2^2 - 1)$-dimensional space. Hence even in the simplest non-trivial case (two qubits) the dimension of the state space becomes very high (15 dimensions) and naive geometric intuition can be misleading. Therefore, it is often useful to look at special classes of model states, which can be characterized by only few parameters. A quite powerful tool is the study of symmetry properties; i.e. to investigate the set of states which is invariant under a group of local unitaries. A general discussion of this scheme can be found in [159]. In this paper we will present only three of the most prominent examples.

Consider first a state $\rho \in \mathscr{S}(\mathscr{H} \otimes \mathscr{H})$ (with $\mathscr{H} = \mathbb{C}^d$) which is invariant under the group of all $U \otimes U$ with a unitary U on $\mathscr{H}$; i.e. $[U \otimes U, \rho] = 0$ for all U. Such a ρ is usually called a *Werner state* [165,128] and its structure can be analyzed quite easily using a well-known result of group theory which goes back to Weyl [171] (see also [142, Theorem IX.11.5]), and which we will state in detail for later reference:

Theorem 3.1. *Each operator A on the N-fold tensor product $\mathscr{H}^{\otimes N}$ of the (finite dimensional) Hilbert space $\mathscr{H}$ which commutes with all unitaries of the form $U^{\otimes N}$ is a linear combination of permutation operators, i.e. $A = \sum_\pi \lambda_\pi V_\pi$, where the sum is taken over all permutations π of N elements, $\lambda_\pi \in \mathbb{C}$ and V_π is defined by*

$$V_\pi \phi_1 \otimes \cdots \otimes \phi_N = \phi_{\pi^{-1}(1)} \otimes \cdots \otimes \phi_{\pi^{-1}(N)} . \tag{3.7}$$

In our case ($N = 2$) there are only two permutations: the identity $\mathbb{1}$ and the flip $F(\psi \otimes \phi) = \phi \otimes \psi$. Hence $\rho = a\mathbb{1} + bF$ with appropriate coefficients a, b. Since ρ is a density matrix, a and b are not independent. To get a transparent way to express these constraints, it is reasonable to consider the eigenprojections $P_\pm$ of F rather than $\mathbb{1}$ and F; i.e. $FP_\pm\psi = \pm P_\pm\psi$ and $P_\pm = (\mathbb{1} \pm F)/2$. The $P_\pm$ are the projections on the subspaces $\mathscr{H}_\pm^{\otimes 2} \subset \mathscr{H} \otimes \mathscr{H}$ of symmetric, respectively antisymmetric, tensor products (Bose-, respectively, Fermi-subspace). If we write $d_\pm = d(d \pm 1)/2$ for the dimensions of

$\mathscr{H}_{\pm}^{\otimes 2}$ we get for each Werner state ρ

$$\rho = \frac{\lambda}{d_+}P_+ + \frac{(1-\lambda)}{d_-}P_-, \quad \lambda \in [0,1] \; . \tag{3.8}$$

On the other hand, it is obvious that each state of this form is $U \otimes U$ invariant, hence a Werner state.

If ρ is given, it is very easy to calculate the parameter λ from the expectation value of ρ and the flip $\mathrm{tr}(\rho F) = 2\lambda - 1 \in [-1, 1]$. Therefore, we can write for an *arbitrary* state $\sigma \in \mathscr{S}(\mathscr{H} \otimes \mathscr{H})$

$$P_{UU}(\sigma) = \frac{\mathrm{tr}(\sigma F) + 1}{2d_+}P_+ + \frac{(1 - \mathrm{tr}\,\sigma F)}{2d_-}P_- \tag{3.9}$$

and this defines a projection from the full state space to the set of Werner states which is called the *twirl operation*. In many cases it is quite useful that it can be written alternatively as a group average of the form

$$P_{UU}(\sigma) = \int_{U(d)} (U \otimes U)\sigma(U^* \otimes U^*)\,\mathrm{d}U \; , \tag{3.10}$$

where $\mathrm{d}U$ denotes the normalized, left invariant Haar measure on $U(d)$. To check this identity note first that its right-hand side is indeed $U \otimes U$ invariant, due to the invariance of the volume element $\mathrm{d}U$. Hence, we have to check only that the trace of F times the integral coincides with $\mathrm{tr}(F\sigma)$:

$$\mathrm{tr}\left[F \int_{U(d)} (U \otimes U)\sigma(U^* \otimes U^*)\,\mathrm{d}U\right] = \int_{U(d)} \mathrm{tr}[F(U \otimes U)\sigma(U^* \otimes U^*)]\,\mathrm{d}U \; , \tag{3.11}$$

$$= \mathrm{tr}(F\sigma)\int_{U(d)} \mathrm{d}U = \mathrm{tr}(F\sigma) \; , \tag{3.12}$$

where we have used the fact that F commutes with $U \otimes U$ and the normalization of $\mathrm{d}U$. We can apply P_{UU} obviously to arbitrary operators $A \in \mathscr{B}(\mathscr{H} \otimes \mathscr{H})$ and, as an integral over unitarily implemented operations, we get a channel. Substituting $U \to U^*$ in (3.10) and cycling the trace $\mathrm{tr}(AP_{UU}(\sigma))$ we find $\mathrm{tr}(P_{UU}(A)\rho) = \mathrm{tr}(AP_{UU}(\rho))$, hence P_{UU} has the same form in the Heisenberg and the Schrödinger picture (i.e. $P_{UU}^* = P_{UU}$).

If $\sigma \in \mathscr{S}(\mathscr{H} \otimes \mathscr{H})$ is a separable state the integrand of $P_{UU}(\sigma)$ in Eq. (3.10) consists entirely of separable states, hence $P_{UU}(\sigma)$ is separable. Since each Werner state ρ is the twirl of itself, we see that ρ is separable iff it is the twirl $P_{UU}(\sigma)$ of a separable state $\sigma \in \mathscr{S}(\mathscr{H} \otimes \mathscr{H})$. To determine the set of separable Werner states we therefore have to calculate only the set of all $\mathrm{tr}(F\sigma) \in [-1, 1]$ with separable σ. Since each such σ admits a convex decomposition into pure product states it is sufficient to look at

$$\langle \psi \otimes \phi, F\psi \otimes \phi\rangle = |\langle \psi, \phi\rangle|^2 \; , \tag{3.13}$$

which ranges from 0 to 1. Hence ρ from Eq. (3.8) is separable iff $\frac{1}{2} \leqslant \lambda \leqslant 1$ and entangled otherwise (due to $\lambda = (\mathrm{tr}(F\rho) + 1)/2$). If $\mathscr{H} = \mathbb{C}^2$ holds, each Werner state is Bell diagonal and we recover the result from Section 3.1.1 (separable if highest eigenvalue less or equal than 1/2).

3.1.3. Isotropic states

To derive a second class of states consider the partial transpose $(\mathrm{Id} \otimes \Theta)\rho$ (with respect to a distinguished base $|j\rangle \in \mathscr{H}$, $j = 1, \ldots, d$) of a Werner state ρ. Since ρ is, by definition, $U \otimes U$

invariant, it is easy to see that $(\mathrm{Id}\otimes\Theta)\rho$ is $U\otimes\bar{U}$ invariant, where $\bar{U}$ denotes componentwise complex conjugation in the base $|j\rangle$ (we just have to use that $U^* = \bar{U}^{\mathrm{T}}$ holds). Each state τ with this kind of symmetry is called an *isotropic state* [132], and our previous discussion shows that τ is a linear combination of $\mathbb{1}$ and the partial transpose of the flip, which is the rank one operator

$$\tilde{F} = (\mathrm{Id}\otimes\Theta)F = |\Psi\rangle\langle\Psi| = \sum_{jk=1}^{d} |jj\rangle\langle kk| \ , \tag{3.14}$$

where $\Psi = \sum_j |jj\rangle$ is, up to normalization a maximally entangled state. Hence, each isotropic τ can be written as

$$\tau = \frac{1}{d}\left(\lambda\frac{\mathbb{1}}{d} + (1-\lambda)\tilde{F}\right), \quad \lambda\in\left[0, \frac{d^2}{d^2-1}\right] \ , \tag{3.15}$$

where the bounds on λ follow from normalization and positivity. As above we can determine the parameter λ from the expectation value

$$\mathrm{tr}(\tilde{F}\tau) = \frac{1-d^2}{d}\lambda + d \ , \tag{3.16}$$

which ranges from 0 to d and this again leads to a twirl operation: For an arbitrary state $\sigma\in\mathcal{S}(\mathcal{H}\otimes\mathcal{H})$ we can define

$$P_{U\bar{U}}(\sigma) = \frac{1}{d(1-d^2)}([\mathrm{tr}(\tilde{F}\sigma) - d]\mathbb{1} + [1 - d\,\mathrm{tr}(\tilde{F}\sigma)]\tilde{F}) \tag{3.17}$$

and as for Werner states $P_{U\bar{U}}$ can be rewritten in terms of a group average

$$P_{U\bar{U}}(\sigma) = \int_{U(d)} (U\otimes\bar{U})\sigma(U^*\otimes\bar{U}^*)\,\mathrm{d}U \ . \tag{3.18}$$

Now we can proceed in the same way as above: $P_{U\bar{U}}$ is a channel with $P^*_{U\bar{U}} = P_{U\bar{U}}$, its fixed points $P_{U\bar{U}}(\tau) = \tau$ are exactly the isotropic states, and the image of the set of separable states under $P_{U\bar{U}}$ coincides with the set of separable isotropic states. To determine the latter we have to consider the expectation values (cf. Eq. (3.13))

$$\langle\psi\otimes\phi, \tilde{F}\psi\otimes\phi\rangle = \left|\sum_{j=1}^{d}\psi_j\phi_j\right| = |\langle\psi, \bar{\phi}\rangle|^2 \in [0,1] \ . \tag{3.19}$$

This implies that τ is separable iff

$$\frac{d(d-1)}{d^2-1} \leqslant \lambda \leqslant \frac{d^2}{d^2-1} \tag{3.20}$$

holds and entangled otherwise. For $\lambda = 0$ we recover the maximally entangled state. For $d = 2$, again we recover again the special case of Bell diagonal states encountered already in the last subsection.

3.1.4. OO-invariant states

Let us combine now Werner states with isotropic states, i.e. we look for density matrices ρ which can be written as $\rho = a\mathbb{1} + bF + c\tilde{F}$, or, if we introduce the three mutually orthogonal projection operators

$$p_0 = \frac{1}{d}\tilde{F}, \quad p_1 = \frac{1}{2}(\mathbb{1} - F), \quad \frac{1}{2}(\mathbb{1} + F) - \frac{1}{d}\tilde{F} \tag{3.21}$$

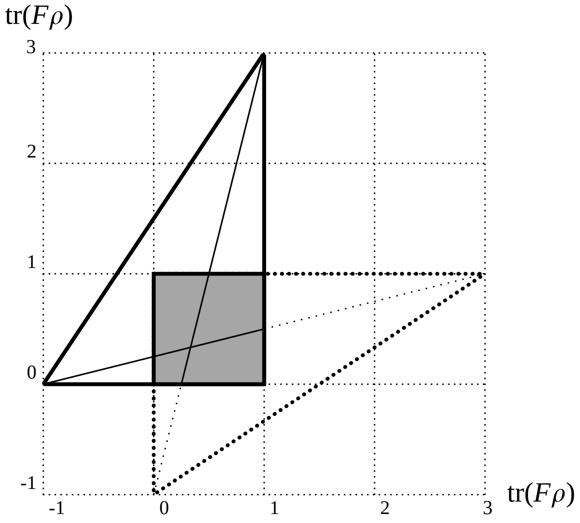


Fig. 3.1. State space of OO-invariant states (upper triangle) and its partial transpose (lower triangle) for $d=3$. The special cases of isotropic and Werner states are drawn as thin lines.

as a convex linear combination of $\mathrm{tr}(p_j)^{-1}p_j$, $j=0,1,2$:

$$\rho=(1-\lambda_1-\lambda_2)p_0+\lambda_1\frac{p_1}{\mathrm{tr}(p_1)}+\lambda_2\frac{p_2}{\mathrm{tr}(p_2)},\quad \lambda_1,\lambda_2\geqslant 0,\quad \lambda_1+\lambda_2\leqslant 1\ . \tag{3.22}$$

Each such operator is invariant under all transformations of the form $U\otimes U$ if U is a unitary with $U=\bar{U}$, in other words: U should be a real orthogonal matrix. A little bit representation theory of the orthogonal group shows that in fact all operators with this invariance property have the form given in (3.22); cf. [159]. The corresponding states are therefore called *OO-invariant*, and we can apply basically the same machinery as in Section 3.1.2 if we replace the unitary group $U(d)$ by the orthogonal group $O(d)$. This includes, in particular, the definition of a twirl operation as an average over $O(d)$ (for an arbitrary $\rho\in\mathscr{S}(\mathscr{H}\otimes\mathscr{H})$):

$$P_{OO}(\rho)=\int_{O(d)}U\otimes U\rho U\otimes U^*\,\mathrm{d}U\ , \tag{3.23}$$

which we can express alternatively in terms of the expectation values $\mathrm{tr}(F\rho)$, $\mathrm{tr}(\tilde{F}\rho)$ by

$$P_{OO}(\rho)=\frac{\mathrm{tr}(\tilde{F}\rho)}{d}\,p_0+\frac{1-\mathrm{tr}(F\rho)}{2\,\mathrm{tr}(p_1)}\,p_1+\left(\frac{1+\mathrm{tr}(F\rho)}{2}-\frac{\mathrm{tr}(\tilde{F}\rho)}{d}\right)\frac{p_2}{\mathrm{tr}(p_2)}\ . \tag{3.24}$$

The range of allowed values for $\mathrm{tr}(F\rho)$, $\mathrm{tr}(\tilde{F}\rho)$ is given by

$$-1\leqslant\mathrm{tr}(F\rho)\leqslant 1,\quad 0\leqslant\mathrm{tr}(\tilde{F}\rho)\leqslant d,\quad \mathrm{tr}(F\rho)\geqslant\frac{2\mathrm{tr}(\tilde{F}\rho)}{d}-1\ . \tag{3.25}$$

For $d=3$ this is the upper triangle in Fig. 3.1.

The values in the lower (dotted) triangle belong to partial transpositions of OO-invariant states. The intersection of both, i.e. the gray-shaded square $Q=[0,1]\times[0,1]$, represents therefore the set of OO-invariant ppt states, and at the same time the set of separable states, since each OO-invariant ppt state is separable. To see the latter note that separable OO-invariant states form a convex subset of Q. Hence, we only have to show that the corners of Q are separable. To do this note that (1) $P_{OO}(\rho)$ is separable whenever ρ is and (2) that $\operatorname{tr}(FP_{OO}(\rho))=\operatorname{tr}(F\rho)$ and $\operatorname{tr}(\tilde{F}P_{OO}(\rho))=\operatorname{tr}(F\rho)$ holds (cf. Eq. (3.12)). We can consider pure product states $|\phi\otimes\psi\rangle\langle\phi\otimes\psi|$ for ρ and get $(|\langle\phi,\psi\rangle|^2, \langle\phi,\bar{\psi}\rangle|^2)$ for the tuple $(\operatorname{tr}(F\rho),\operatorname{tr}(\tilde{F}\rho))$. Now the point $(1,1)$ in Q is obtained if $\psi=\phi$ is real, the point $(0,0)$ is obtained for real and orthogonal ϕ,ψ and the point $(1,0)$ belongs to the case $\psi=\phi$ and $\langle\phi,\bar{\phi}\rangle=0$. Symmetrically we get $(0,1)$ with the same ϕ and $\psi=\bar{\phi}$.

3.1.5. PPT states

We have seen in Theorem 2.13 that separable states and ppt states coincide in 2×2 and 2×3 dimensions. Another class of examples with this property are OO-invariant states just studied. Nevertheless, separability and a positive partial transpose are *not* equivalent. An easy way to produce such examples of states which are entangled and ppt is given in terms of *unextendible product bases* [14]. An orthonormal family $\phi_j\in\mathscr{H}_1\otimes\mathscr{H}_2$, $j=1,\dots,N<d_1d_2$ (with $d_k=\dim\mathscr{H}_k$) is called an unextendible product basis[9] (UPB) iff (1) all ϕ_j are product vectors and (2) there is no product vector orthogonal to all ϕ_j. Let us denote the projector to the span of all ϕ_j by E, its orthocomplement by $E^\perp$, i.e. $E^\perp=\mathbb{1}-E$, and define the state $\rho=(d_1d_2-N)^{-1}E^\perp$. It is entangled because there is by construction no product vector in the support of ρ, and it is ppt. The latter can be seen as follows: The projector E is a sum of the one-dimensional projectors $|\phi_j\rangle\langle\phi_j|$, $j=1,\dots,N$. Since all ϕ_j are product vectors the partial transposes of the $|\phi_j\rangle\langle\phi_j|$ are of the form $|\tilde{\phi}_j\rangle\langle\tilde{\phi}_j|$, with another UPB $\tilde{\phi}_j$, $j=1,\dots,N$ and the partial transpose $(\mathbb{1}\otimes\Theta)E$ of E is the sum of the $|\tilde{\phi}_j\rangle\langle\tilde{\phi}_j|$. Hence $(\mathbb{1}\otimes\Theta)E^\perp=\mathbb{1}-(\mathbb{1}\otimes\Theta)E$ is a projector and therefore positive.

To construct entangled ppt states we have to find UPBs. The following two examples are taken from [14]. Consider first the five vectors

$$\phi_j=N(\cos(2\pi j/5),\sin(2\pi j/5),h),\quad j=0,\dots,4 \tag{3.26}$$

with $N=2/\sqrt{5+\sqrt{5}}$ and $h=\frac{1}{2}\sqrt{1+\sqrt{5}}$. They form the apex of a regular pentagonal pyramid with height h. The latter is chosen such that non-adjacent vectors are orthogonal. It is now easy to show that the five vectors

$$\Psi_j=\phi_j\otimes\phi_{2j\bmod 5},\quad j=0,\dots,4 \tag{3.27}$$

form a UPB in the Hilbert space $\mathscr{H}\otimes\mathscr{H}$, $\dim\mathscr{H}=3$ (cf. [14]). A second example, again in (3×3)-dimensional Hilbert space are the following five vectors (called “Tiles” in [14]):

$$\frac{1}{\sqrt{2}}|0\rangle\otimes(|0\rangle-|1\rangle),\quad \frac{1}{\sqrt{2}}|2\rangle\otimes(|1\rangle-|2\rangle),\quad \frac{1}{\sqrt{2}}(|0\rangle-|1\rangle)\otimes|2\rangle\ ,$$

[9] This name is somewhat misleading because the ϕ_j are *not* a base of $\mathscr{H}_1\otimes\mathscr{H}_2$.

$$\frac{1}{\sqrt{2}}(|1\rangle - |2\rangle)\otimes|0\rangle, \quad \frac{1}{3}(|0\rangle + |1\rangle + |2\rangle)\otimes(|0\rangle + |1\rangle + |2\rangle)\,, \tag{3.28}$$

where $|k\rangle$, $k = 0,1,2$ denotes the standard basis in $\mathscr{H} = \mathbb{C}^3$.

3.1.6. Multipartite states

In many applications of quantum information rather big systems, consisting of a large number of subsystems, occur (e.g. a quantum register of a quantum computer) and it is necessary to study the corresponding correlation and entanglement properties. Since this is a fairly difficult task, there is not much known about—much less as in the two-partite case, which we mainly consider in this paper. Nevertheless, in this subsection we will give a rough outline of some of the most relevant aspects.

At the level of pure states the most significant difficulty is the lack of an analog of the Schmidt decomposition [126]. More precisely, there are elements in an N-fold tensor product $\mathscr{H}^{(1)}\otimes\cdots\otimes\mathscr{H}^{(N)}$ (with $N > 2$) which cannot be written as [10]

$$\Psi = \sum_{j=1}^{d} \lambda_j \phi_j^{(1)} \otimes \cdots \otimes \phi_j^{(N)} \tag{3.29}$$

with N orthonormal bases $\phi_1^{(k)},\ldots,\phi_d^{(k)}$ of $\mathscr{H}^{(k)}$, $k = 1,\ldots,N$. To get examples for such states in the tri-partite case, note first that any partial trace of $|\Psi\rangle\langle\Psi|$ with Ψ from Eq. (3.29) has separable eigenvectors. Hence, each purification (Corollary 2.3) of an entangled, two-partite, mixed state with inseparable eigenvectors (e.g. a Bell diagonal state) does not admit a Schmidt decomposition. This implies on the one hand that there are interesting new properties to be discovered, but on the other we see that many techniques developed for bipartite pure states can be generalized in a straightforward way only for states which are *Schmidt decomposable* in the sense of Eq. (3.29). The most well-known representative of this class for a tripartite qubit system is the GHZ state [73]

$$\Psi = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)\,, \tag{3.30}$$

which has the special property that contradictions between local hidden variable theories and quantum mechanics occur even for non-statistical predictions (as opposed to maximally entangled states of bipartite systems [73,117,116]).

A second new aspect arising in the discussion of multiparty entanglement is the fact that several different notions of separability occur. A state ρ of an N-partite system $\mathscr{B}(\mathscr{H}_1)\otimes\cdots\otimes\mathscr{B}(\mathscr{H}_N)$ is called *N-separable* if

$$\rho = \sum_J \lambda_J \rho_{j_1} \otimes \cdots \otimes \rho_{j_N} \tag{3.31}$$

with states $\rho_{j_k} \in \mathscr{B}^*(\mathscr{H}_k)$ and multiindices $J = (j_1,\ldots,j_k)$. Alternatively, however, we can decompose $\mathscr{B}(\mathscr{H}_1)\otimes\cdots\otimes\mathscr{B}(\mathscr{H}_N)$ into two subsystems (or even into M subsystems if $M < N$) and call ρ *biseparable* if it is separable with respect to this decomposition. It is obvious that N-separability implies

[10] There is, however, the possibility to choose the bases $\phi_1^{(k)},\ldots,\phi_d^{(k)}$ such that the number of summands becomes minimal. For tri-partite systems this “minimal canonical form” is study in [1].

biseparability with respect to all possible decompositions. The converse is—not very surprisingly—not true. One way to construct a corresponding counterexample is to use an unextendable product base (cf. Section 3.1.5). In [14] it is shown that the tripartite qubit state complementary to the UPB

$$|0,1,+\rangle, |1,+,0\rangle, |+,0,1\rangle, |-,-,-\rangle \text{ with } |\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle) \tag{3.32}$$

is entangled (i.e. tri-inseparable) but biseparable with respect to any decomposition into two subsystems (cf. [14] for details).

Another, maybe more systematic, way to find examples for multipartite states with interesting properties is the generalization of the methods used for Werner states (Section 3.1.2), i.e. to look for density matrices $\rho \in \mathscr{B}^*(\mathscr{H}^{\otimes N})$ which commute with all unitaries of the form $U^{\otimes N}$. Applying again Theorem 3.1 we see that each such ρ is a linear combination of permutation unitaries. Hence, the structure of the set of all $U^{\otimes N}$ invariant states can be derived from representation theory of the symmetric group (which can be tedious for large $N!$). For $N = 3$ this program is carried out in [61] and it turns out that the corresponding set of invariant states is a five-dimensional (real) manifold. We skip the details here and refer to [61] instead.

3.2. Channels

In Section 2.3 we have introduced channels as very general objects transforming arbitrary types of information (i.e. classical, quantum and mixtures of them) into one another. In the following, we will consider some of the most important special cases.

3.2.1. Quantum channnels

Many tasks of quantum information theory require the transmission of quantum information over long distances, using devices like optical fibers or storing quantum information in some sort of memory. Both situations can be described by a channel or quantum *operation* $T : \mathscr{B}(\mathscr{H}) \to \mathscr{B}(\mathscr{H})$, where $T^*(\rho)$ is the quantum information which will be received when ρ was sent, or alternatively: which will be read off the quantum memory when ρ was written. Ideally, we would prefer those channels which do not affect the information at all, i.e. $T = \mathbb{1}$, or, as the next best choice, a T whose action can be undone by a physical device, i.e. T should be invertible and T^{-1} is again a channel. The Stinespring Theorem (Theorem 2.8) immediately shows that this implies $T^*\rho = U\rho U^*$ with a unitary U; in other words, the systems carrying the information do not interact with the environment. We will call such a kind of channel an *ideal channel*. In real situations, however, interaction with the environment, i.e. additional, unobservable degrees of freedom, cannot be avoided. The general structure of such a *noisy channel* is given by

$$T^*(\rho) = \operatorname{tr}_{\mathscr{K}}(U(\rho \otimes \rho_0)U^*) , \tag{3.33}$$

where $U : \mathscr{H} \otimes \mathscr{K} \to \mathscr{H} \otimes \mathscr{K}$ is a unitary operator describing the common evolution of the system (Hilbert space $\mathscr{H}$) and the environment (Hilbert space $\mathscr{K}$) and $\rho_0 \in \mathscr{S}(\mathscr{K})$ is the initial state of the environment (cf. Fig. 3.2). It is obvious that the quantum information originally stored in $\rho \in \mathscr{S}(\mathscr{H})$ cannot be completely recovered from $T^*(\rho)$ if *only one system is available*. It is an easy consequence of the Stinepspring theorem that each channel can be expressed in this form

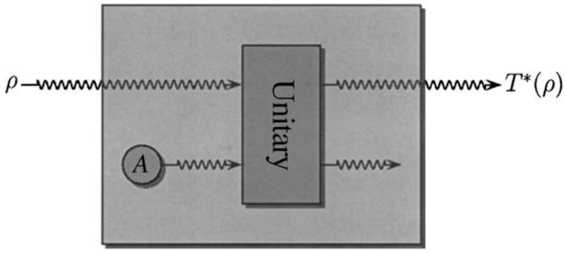


Fig. 3.2. Noisy channel.

Corollary 3.2 (Ancilla form). *Assume that* $T:\mathscr{B}(\mathscr{H})\to\mathscr{B}(\mathscr{H})$ *is a channel. Then there is a Hilbert space* $\mathscr{K}$, *a pure state* ρ_0 *and a unitary map* $U:\mathscr{H}\otimes\mathscr{K}\to\mathscr{H}\otimes\mathscr{K}$ *such that Eq.* (3.33) *holds. It is always possible*, *to choose* $\mathscr{K}$ *such that* $\dim(\mathscr{K})=\dim(\mathscr{H})^3$ *holds.*

Proof. Consider the Stinepspring form $T(A)=V^*(A\otimes\mathbb{1})V$ with $V:\mathscr{H}\to\mathscr{H}\otimes\mathscr{K}$ of T and choose a vector $\psi\in\mathscr{K}$ such that $U(\phi\otimes\psi)=V(\phi)$ can be extended to a unitary map $U:\mathscr{H}\otimes\mathscr{K}\to\mathscr{H}\otimes\mathscr{K}$ (this is always possible since T is unital and V therefore isometric). If $e_j\in\mathscr{H}$, $j=1,\dots,d_1$ and $f_k\in\mathscr{K}$, $k=1,\dots,d_2$ are orthonormal bases with $f_1=\psi$ we get

$$\operatorname{tr}[T(A)\rho]=\operatorname{tr}[\rho V^*(A\otimes\mathbb{1})V]=\sum_j\langle V\rho e_j,(A\otimes\mathbb{1})Ve_j\rangle \tag{3.34}$$

$$=\sum_{jk}\langle U(\rho\otimes|\psi\rangle\langle\psi|)(e_j\otimes f_k),(A\otimes\mathbb{1})U(e_j\otimes f_k)\rangle \tag{3.35}$$

$$=\operatorname{tr}[\operatorname{tr}_{\mathscr{K}}[U(\rho\otimes|\psi\rangle\langle\psi|)U^*]A]\ , \tag{3.36}$$

which proves the statement. □

Note that there are, in general, many ways to express a channel this way, e.g. if T is an ideal channel $\rho\mapsto T^*\rho=U\rho U^*$ we can rewrite it with an arbitrary unitary $U_0:\mathscr{K}\to\mathscr{K}$ by $T^*\rho=\operatorname{tr}_2(U\otimes U_0\rho\otimes\rho_0U^*\otimes U_0^*)$. This is the weakness of the ancilla form compared to the Stinespring representation of Theorem 2.8. Nevertheless, Corollary 3.2 shows that each channel which is not an ideal channel is noisy in the described way.

The most prominent example for a noisy channel is the *depolarizing channel* for d-level systems (i.e. $\mathscr{H}=\mathbb{C}^d$)

$$\mathscr{S}(\mathscr{H})\ni\rho\mapsto\vartheta\rho+(1-\vartheta)\frac{\mathbb{1}}{d}\in\mathscr{S}(\mathscr{H}),\quad 0\leqslant\vartheta\leqslant 1 \tag{3.37}$$

or in the Heisenberg picture

$$\mathscr{B}(\mathscr{H})\ni A\mapsto\vartheta A+(1-\vartheta)\frac{\operatorname{tr}(A)}{d}\mathbb{1}\in\mathscr{B}(\mathscr{H})\ . \tag{3.38}$$

A Stinespring dilation of T (not the minimal one—this can be checked by counting dimensions) is given by $\mathscr{K}=\mathscr{H}\otimes\mathscr{H}\oplus\mathbb{C}$ and $V:\mathscr{H}\to\mathscr{H}\otimes\mathscr{K}=\mathscr{H}^{\otimes 3}\oplus\mathscr{H}$ with

$$|j\rangle\mapsto V|j\rangle=\left[\sqrt{\frac{1-\vartheta}{d}}\sum_{k=1}^{d}|k\rangle\otimes|k\rangle\otimes|j\rangle\right]\oplus[\sqrt{\vartheta}|j\rangle]\ , \tag{3.39}$$

where $|k\rangle$, $k=1,\dots,d$ denotes again the canonical basis in $\mathscr{H}$. An ancilla form of T with the same $\mathscr{K}$ is given by the (pure) environment state

$$\psi = \left[\sqrt{\frac{1-\vartheta}{d}} \sum_{k=1}^{d} |k\rangle \otimes |k\rangle\right] \oplus [\sqrt{\vartheta}|0\rangle] \in \mathscr{K} \tag{3.40}$$

and the unitary operator $U : \mathscr{H} \otimes \mathscr{K} \to \mathscr{H} \otimes \mathscr{K}$ with

$$U(\phi_1 \otimes \phi_2 \otimes \phi_3 \oplus \chi) = \phi_2 \otimes \phi_3 \otimes \phi_1 \oplus \chi , \tag{3.41}$$

i.e. U is the direct sum of a permutation unitary and the identity.

3.2.2. Channels under symmetry

Similarly to the discussion in Section 3.1 it is often useful to consider channels with special symmetry properties. To be more precise, consider a group G and two unitary representations π_1, π_2 on the Hilbert spaces $\mathscr{H}_1$ and $\mathscr{H}_2$, respectively. A channel $T : \mathscr{B}(\mathscr{H}_1) \to \mathscr{B}(\mathscr{H}_2)$ is called *covariant* (with respect to π_1 and π_2) if

$$T[\pi_1(U)A\pi_1(U)^*] = \pi_2(U)T[A]\pi_2(U)^* \quad \forall A \in \mathscr{B}(\mathscr{H}_1) \quad \forall U \in G \tag{3.42}$$

holds. The general structure of covariant channels is governed by a fairly powerful variant of Stinespring's theorem which we will state below (and which will be very useful for the study of the cloning problem in Section 7). Before we do this let us have a short look on a particular class of examples which is closely related to OO-invariant states.

Hence consider a channel $T : \mathscr{B}(\mathscr{H}) \to \mathscr{B}(\mathscr{H})$ which is covariant with respect to the orthogonal group, i.e. $T(UAU^*) = UT(A)U^*$ for all unitaries U on $\mathscr{H}$ with $\bar{U} = U$ in a distinguished basis $|j\rangle$, $j=1,\dots,d$. The maximally entangled state $\psi = d^{-1/2} \sum_j |jj\rangle$ is OO-invariant, i.e. $U \otimes U\psi = \psi$ for all these U. Therefore, each state $\rho = (\mathrm{Id} \otimes T^*)|\psi\rangle\langle\psi|$ is OO-invariant as well and by the duality lemma (Theorem 2.10) T and ψ are uniquely determined (up to unitary equivalence) by ρ. This means we can use the structure of OO-invariant states derived in Section 3.1.4 to characterize all orthogonal covariant channels. As a first step consider the linear maps $X_1(A) = d\,\mathrm{tr}(A)\mathbb{1}$, $X_2(A) = dA^{\mathrm{T}}$ and $X_3(A) = dA$. They are not channels (they are not unital and X_2 is not cp) but they have the correct covariance property and it is easy to see that they correspond to the operators $\mathbb{1}, F, \tilde{F} \in \mathscr{B}(\mathscr{H} \otimes \mathscr{H})$, i.e.

$$(\mathrm{Id} \otimes X_1)|\psi\rangle\langle\psi| = \mathbb{1}, \quad (\mathrm{Id} \otimes X_2)|\psi\rangle\langle\psi| = F, \quad (\mathrm{Id} \otimes X_3)|\psi\rangle\langle\psi| = \tilde{F} . \tag{3.43}$$

Using Eq. (3.21), we can determine therefore the channels which belong to the three extremal OO-invariant states (the corners of the upper triangle in Fig. 3.1):

$$T_0(A) = A, \quad T_1(A) = \frac{\mathrm{tr}(A)\mathbb{1} - A^{\mathrm{T}}}{d-1} , \tag{3.44}$$

$$T_2(A) = \frac{2}{d(d+1)-2}\left[\frac{d}{2}(\mathrm{tr}(A)\mathbb{1} + A^{\mathrm{T}}) - A\right] . \tag{3.45}$$

Each OO-invariant channel is a convex linear combination of these three. Special cases are the channels corresponding to Werner and isotropic states. The latter leads to depolarizing channels

$T(A)=\vartheta A+(1-\vartheta)d^{-1}\mathrm{tr}(A)\mathbb{1}$ with $\vartheta\in[0,d^2/(d^2-1)]$; cf. Eq. (3.15), while Werner states correspond to

$$T(A)=\frac{\vartheta}{d+1}[\mathrm{tr}(A)\mathbb{1}+A^{\mathrm{T}}]+\frac{1-\vartheta}{d-1}[\mathrm{tr}(A)\mathbb{1}-A^{\mathrm{T}}],\quad \vartheta\in[0,1]\ ; \tag{3.46}$$

cf. Eq. (3.8).

Let us come back now to the general case. We will state here the covariant version of the Stinespring theorem (see [98] for a proof). The basic idea is that all covariant channels are parameterized by representations on the dilation space.

Theorem 3.3. *Let G be a group with finite-dimensional unitary representations $\pi_j : G\to U(\mathscr{H}_j)$ and $T:\mathscr{B}(\mathscr{H}_1)\to\mathscr{B}(\mathscr{H}_2)$ a π_1,π_2-covariant channel. Then there is a finite-dimensional unitary representation $\tilde{\pi}:G\to U(\mathscr{K})$ and an operator $V:\mathscr{H}_2\to\mathscr{H}_1\otimes\mathscr{K}$ with $V\pi_2(U)=\pi_1(U)\otimes\tilde{\pi}(U)$ and $T(A)=V^*A\otimes\mathbb{1}V$.*

To get an explicit example consider the dilation of a depolarizing channel given in Eq. (3.39). In this case we have $\pi_1(U)=\pi_2(U)=U$ and $\tilde{\pi}(U)=(U\otimes\bar{U})\oplus\mathbb{1}$. The check that the map V has indeed the intertwining property $V\pi_2(U)=\pi_1(U)\otimes\tilde{\pi}(U)$ stated in the theorem is left as an exercise to the reader.

3.2.3. Classical channels

The classical analog to a quantum operation is a channel $T:\mathscr{C}(X)\to\mathscr{C}(Y)$ which describes the transmission or manipulation of classical information. As we have mentioned already in Section 2.3.1 positivity and complete positivity are equivalent in this case. Hence, we have to assume only that T is positive and unital. Obviously, T is characterized by its matrix elements $T_{xy}=\delta_y(T|x\rangle\langle x|)$, where $\delta_y\in\mathscr{C}^*(X)$ denotes the Dirac measure at $y\in Y$ and $|x\rangle\langle x|\in\mathscr{C}(X)$ is the canonical basis in $\mathscr{C}(X)$ (cf. Section 2.1.3). Positivity and normalization of T imply that $0\leqslant T_{xy}\leqslant 1$ and

$$1=\delta_y(\mathbb{1})=\delta_y(T(\mathbb{1}))=\delta_y\left[T\left(\sum_x|x\rangle\langle x|\right)\right]=\sum_x T_{xy} \tag{3.47}$$

holds. Hence, the family $(T_{xy})_{x\in X}$ is a probability distribution on X and T_{xy} is therefore the probability to get the information $x\in X$ at the output side of the channel if $y\in Y$ was send. Each classical channel is uniquely determined by its matrix of *transition probabilities*. For $X=Y$ we see that the information is transmitted without error iff $T_{xy}=\delta_{xy}$, i.e. T is an ideal channel if $T=\mathrm{Id}$ holds and noisy otherwise.

3.2.4. Observables and preparations

Let us consider now a channel which transforms quantum information $\mathscr{B}(\mathscr{H})$ into classical information $\mathscr{C}(X)$. Since positivity and complete positivity are again equivalent, we just have to look at a positive and unital map $E:\mathscr{C}(X)\to\mathscr{B}(\mathscr{H})$. With the canonical basis $|x\rangle\langle x|$, $x\in X$ of $\mathscr{C}(X)$ we get a family $E_x=E(|x\rangle\langle x|)$, $x\in X$ of positive operators $E_x\in\mathscr{B}(\mathscr{H})$ with $\sum_{x\in X}E_x=\mathbb{1}$. Hence the E_x form a POV measure, i.e. an observable. If on the other hand a POV measure $E_x\in\mathscr{B}(\mathscr{H})$, $x\in X$ is given we can define a quantum to classical channel $E:\mathscr{C}(X)\to\mathscr{B}(\mathscr{H})$ by $E(f)=\sum_x f(x)E_x$.

This shows that the observable $E_x, x \in X$ and the channel E can be identified and we say *E is the observable*.

Keeping this interpretation in mind it is possible to have a short look at continuous observables without the need of abstract measure theory: We only have to define the classical algebra $\mathscr{C}(X)$ for a set X which is not finite or discrete. For simplicity, we assume that $X = \mathbb{R}$ holds; however, the generalization to other locally compact spaces is straightforward. We choose for $\mathscr{C}(\mathbb{R})$ the space of continuous, complex-valued functions vanishing at infinity, i.e. $|f(x)| < \varepsilon$ for each $\varepsilon > 0$ provided $|x|$ is large enough. $\mathscr{C}(\mathbb{R})$ can be equipped with the sup-norm and becomes an Abelian C*-algebra (cf. [25]). To interpret it as an operator algebra as assumed in Section 2.1.1 we have to identify $f \in \mathscr{C}(\mathbb{R})$ with the corresponding multiplication operator on $\mathrm{L}^2(\mathbb{R})$. An observable taking arbitrary real values can now be defined as a positive map $E : \mathscr{C}(\mathbb{R}) \to \mathscr{B}(\mathscr{H})$. The probability to get a result in the interval $[a, b] \subset \mathbb{R}$ during an E measurement on systems in the state ρ is[11]

$$\mu([a,b]) = \sup\{\operatorname{tr}(E(f)\rho) \mid f \in \mathscr{C}(\mathbb{R}),\ 0 \leqslant f \leqslant \mathbb{1},\ \operatorname{supp} f \subset [a,b]\}\ , \tag{3.48}$$

where supp denotes the *support* of f. The most well-known example for $\mathbb{R}$ valued observables are of course position Q and momentum P of a free particle in one dimension. In this case we have $\mathscr{H} = \mathrm{L}^2(\mathbb{R})$ and the channels corresponding to Q and P are (in position representation) given by $\mathscr{C}(\mathbb{R}) \ni f \mapsto E_Q(f) \in \mathscr{B}(\mathscr{H})$ with $E_Q(f)\psi = f\psi$, respectively, $\mathscr{C}(\mathbb{R}) \ni f \mapsto E_P(f) \in \mathscr{B}(\mathscr{H})$ with $E_P(f)\psi = (f\hat{\psi})^\vee$ where $\wedge$ and $\vee$ denote the Fourier transform and its inverse.

Let us return now to a finite set X and exchange the role of $\mathscr{C}(X)$ and $\mathscr{B}(\mathscr{H})$; in other words let us consider a channel $R : \mathscr{B}(\mathscr{H}) \to \mathscr{C}(X)$ with a classical input and a quantum output algebra. In the Schrödinger picture we get a family of density matrices $\rho_x := R^*(\delta_x) \in \mathscr{B}^*(\mathscr{H})$, $x \in X$, where $\delta_x \in \mathscr{C}^*(X)$ again denote the Dirac measures (cf. Section 2.1.3). Hence, we get a *parameter-dependent preparation* which can be used to encode the classical information $x \in X$ into the quantum information $\rho_x \in \mathscr{B}^*(\mathscr{H})$.

3.2.5. Instruments and parameter-dependent operations

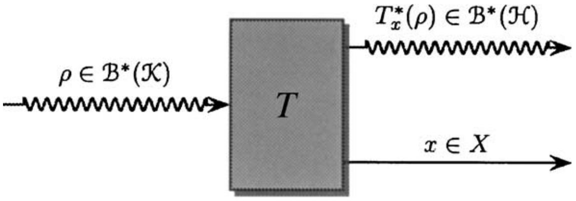
An observable describes only the statistics of measuring results, but does not contain information about the state of the system after the measurement. To get a description which fills this gap we have to consider channels which operates on quantum systems and produces hybrid systems as output, i.e. $T : \mathscr{B}(\mathscr{H}) \otimes \mathscr{M}(X) \to \mathscr{B}(\mathscr{K})$. Following Davies [50] we will call such an object an *instrument*. From T we can derive the subchannel

$$\mathscr{C}(X) \ni f \mapsto T(\mathbb{1} \otimes f) \in \mathscr{B}(\mathscr{K})\ , \tag{3.49}$$

which is the observable measured by T, i.e. $\operatorname{tr}[T(\mathbb{1} \otimes |x\rangle\langle x|)\rho]$ is the probability to measure $x \in X$ on systems in the state ρ. On the other hand, we get for each $x \in X$ a quantum channel (which is *not* unital)

$$\mathscr{B}(\mathscr{H}) \ni A \mapsto T_x(A) = T(A \otimes |x\rangle\langle x|) \in \mathscr{B}(\mathscr{K})\ . \tag{3.50}$$

[11] Due to the Riesz–Markov theorem (cf. [134, Theorem IV.18]) the set function μ extends in unique way to a probability measure on the real line.


$T_x^*(\rho) \in \mathcal{B}^*(\mathcal{H})$
$\rho \in \mathcal{B}^*(\mathcal{K})$
T
$x \in X$

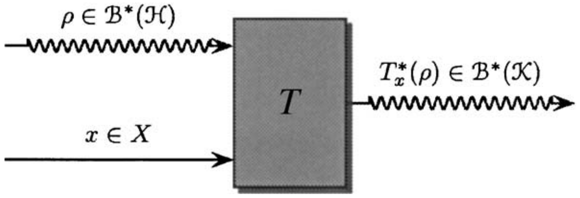

$\rho \in \mathcal{B}^*(\mathcal{H})$
$T_x^*(\rho) \in \mathcal{B}^*(\mathcal{K})$
T
$x \in X$

Fig. 3.3. Instrument.

Fig. 3.4. Parameter-dependent operation.

It describes the operation performed by the instrument T if $x \in X$ was measured. More precisely if a measurement on systems in the state ρ gives the result $x \in X$ we get (up to normalization) the state $T_x^*(\rho)$ *after the measurement* (cf. Fig. 3.3), while

$$\mathrm{tr}(T_x^*(\rho)) = \mathrm{tr}(T_x^*(\rho)\mathbb{1}) = \mathrm{tr}(\rho T(\mathbb{1} \otimes |x\rangle\langle x|)) \tag{3.51}$$

is (again) the probability to measure $x \in X$ on ρ. The instrument T can be expressed in terms of the operations T_x by

$$T(A \otimes f) = \sum_x f(x) T_x(A) \ ; \tag{3.52}$$

hence, we can identify T with the family T_x, $x \in X$. Finally, we can consider the second marginal of T

$$\mathscr{B}(\mathscr{H}) \ni A \mapsto T(A \otimes \mathbb{1}) = \sum_{x \in X} T_x(A) \in \mathscr{B}(\mathscr{K}) \ . \tag{3.53}$$

It describes the operation we get if the outcome of the measurement is ignored.

The most well-known example of an instrument is a *von Neumann–Lüders measurement* associated to a PV measure given by family of projections E_x, $x = 1, \ldots d$; e.g. the eigenprojections of a self-adjoint operator $A \in \mathscr{B}(\mathscr{H})$. It is defined as the channel

$$T : \mathscr{B}(\mathscr{H}) \otimes \mathscr{C}(X) \to \mathscr{B}(\mathscr{H}) \quad \text{with } X = \{1, \ldots, d\} \quad \text{and} \quad T_x(A) = E_x A E_x \ . \tag{3.54}$$

Hence, we get the final state $\mathrm{tr}(E_x \rho)^{-1} E_x \rho E_x$ if we measure the value $x \in X$ on systems initially in the state ρ—this is well known from quantum mechanics.

Let us change now the role of $\mathscr{B}(\mathscr{H}) \otimes \mathscr{C}(X)$ and $\mathscr{B}(\mathscr{K})$; in other words, consider a channel $T : \mathscr{B}(\mathscr{K}) \to \mathscr{B}(\mathscr{H}) \otimes \mathscr{C}(X)$ with hybrid input and quantum output. It describes a device which changes the state of a system depending on additional classical information. As for an instrument, T decomposes into a family of (unital!) channels $T_x : \mathscr{B}(\mathscr{K}) \to \mathscr{B}(\mathscr{H})$ such that we get $T^*(\rho \otimes p) = \sum_x p_x T_x^*(\rho)$ in the Schrödinger picture. Physically T describes a *parameter-dependent operation*: depending on the classical information $x \in X$ the quantum information $\rho \in \mathscr{B}(\mathscr{K})$ is transformed by the operation T_x (cf. Fig. 3.4).

Finally, we can consider a channel $T : \mathscr{B}(\mathscr{H}) \otimes \mathscr{C}(X) \to \mathscr{B}(\mathscr{K}) \otimes \mathscr{C}(Y)$ with hybrid input and output to get a *parameter-dependent instrument* (cf. Fig. 3.5): Similar to the discussion in the last paragraph we can define a family of instruments $T_y : \mathscr{B}(\mathscr{H}) \otimes \mathscr{C}(X) \to \mathscr{B}(\mathscr{K})$, $y \in Y$ by the equation $T^*(\rho \otimes p) = \sum_y p_y T_y^*(\rho)$. Physically, T describes the following device: It receives

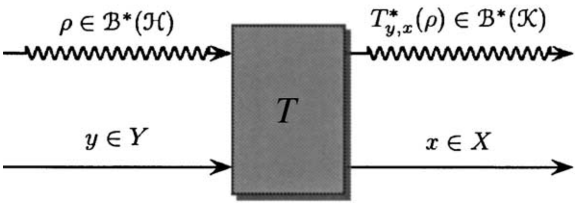

$\rho \in \mathcal{B}^*(\mathcal{H})$
$T^*_{y,x}(\rho) \in \mathcal{B}^*(\mathcal{K})$
T
$y \in Y$
$x \in X$

Fig. 3.5. Parameter-dependent instrument.

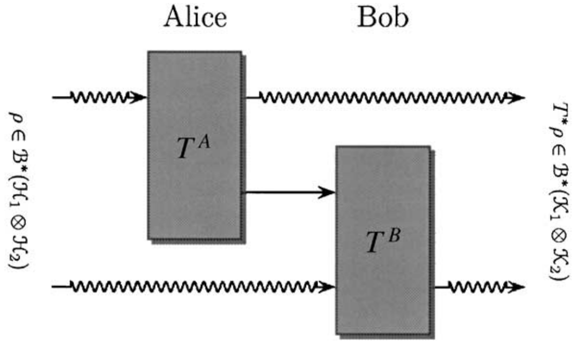

Alice
Bob
$\rho \in \mathcal{B}^*(\mathcal{H}_1 \otimes \mathcal{H}_2)$
T^A
T^B
$T^*\rho \in \mathcal{B}^*(\mathcal{K}_1 \otimes \mathcal{K}_2)$

Fig. 3.6. One-way LOCC operation; cf. Fig. 3.7 for an explanation.

the classical information $y \in Y$ and a quantum system in the state $\rho \in \mathscr{B}^*(\mathscr{K})$ as input. Depending on y a measurement with the instrument T_y is performed, which in turn produces the measuring value $x \in X$ and leaves the quantum system in the state (up to normalization) $T^*_{y,x}(\rho)$; with $T_{y,x}$ given as in Eq. (3.50) by $T_{y,x}(A) = T_y(A \otimes |x\rangle\langle x|)$.

3.2.6. LOCC and separable channels

Let us consider now channels acting on finite-dimensional bipartite systems: $T : \mathscr{B}(\mathscr{H}_1 \otimes \mathscr{K}_2) \to \mathscr{B}(\mathscr{K}_1 \otimes \mathscr{K}_2)$. In this case we can ask the question whether a channel preserves separability. Simple examples are *local operations* (LOs), i.e. $T = T^A \otimes T^B$ with two channels $T^{A,B} : \mathscr{B}(\mathscr{H}_j) \to \mathscr{B}(\mathscr{K}_j)$. Physically, we think of such a T in terms of two physicists Alice and Bob both performing operations on their own particle but without information transmission neither classical nor quantum. The next difficult step are LOs with *one-way classical communications* (one way LOCC). This means Alice operates on her system with an instrument, communicates the classical measuring result $j \in X = \{1, \ldots, N\}$ to Bob and he selects an operation depending on these data. We can write such a channel as a composition $T = (T^A \otimes \mathrm{Id})(\mathrm{Id} \otimes T^B)$ of the instrument $T^A : \mathscr{B}(\mathscr{H}_1) \otimes \mathscr{C}(X_1) \to \mathscr{B}(\mathscr{K}_1)$ and the parameter-dependent operation $T^B : \mathscr{B}(\mathscr{H}_2) \to \mathscr{C}(X_1) \otimes \mathscr{B}(\mathscr{K}_2)$ (cf. Fig. 3.6)

$$\mathscr{B}(\mathscr{H}_1 \otimes \mathscr{H}_2) \overset{\mathrm{Id} \otimes T^B}{\longrightarrow} \mathscr{B}(\mathscr{H}_1) \otimes \mathscr{C}(X) \otimes \mathscr{B}(\mathscr{K}_2) \overset{T^A \otimes \mathrm{Id}}{\longrightarrow} \mathscr{B}(\mathscr{K}_1 \otimes \mathscr{K}_2) . \tag{3.55}$$

It is of course possible to continue the chain in Eq. (3.55), i.e. instead of just operating on his system, Bob can invoke a parameter-dependent instrument depending on Alice's data $j_1 \in X_1$, send the corresponding measuring results $j_2 \in X_2$ to Alice and so on. To write down the corresponding chain of maps (as in Eq. (3.55)) is simple but not very illuminating and therefore omitted; cf. Fig. 3.7 instead. If we allow Alice and Bob to drop some of their particles, i.e. the operations they perform need not to be unital, we get an *LOCC channel* ("local operations and classical communications"). It represents the most general physical process which can be performed on a two partite system if only classical communication (in both directions) is available.

The LOCC channels play a significant role in entanglement theory (we will see this in Section 4.3), but they are difficult to handle. Fortunately, it is often possible to replace them by closely

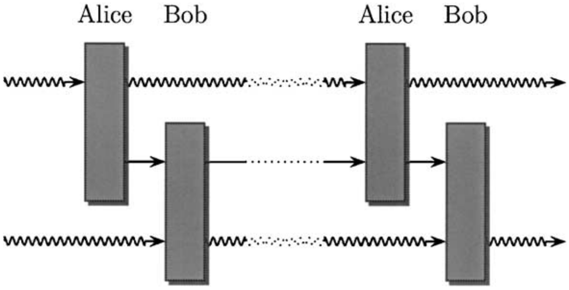


Fig. 3.7. LOCC operation. The upper and lower curly arrows represent Alice's respectively Bob's, quantum system, while the straight arrows in the middle stand for the classical information Alice and Bob exchange. The boxes symbolize the channels applied by Alice and Bob.

related operations with a more simple structure: A *not necessarily unital* channel $T:\mathscr{B}(\mathscr{H}_1\otimes\mathscr{K}_2)\to\mathscr{B}(\mathscr{K}_1\otimes\mathscr{K}_2)$ is called *separable*, if it is a sum of (in general non-unital) local operations, i.e.

$$T=\sum_{j=1}^{N}T_j^A\otimes T_j^B. \tag{3.56}$$

It is easy to see that a separable T maps separable states to separable states (up to normalization) and that each LOCC channel is separable (cf. [13]). The converse, however, is (somewhat surprisingly) not true: there are separable channels which are not LOCC, see [13] for a concrete example.

3.3. Quantum mechanics in phase space

Up to now we have considered only finite-dimensional systems and even in this extremely idealized situation it is not easy to get non-trivial results. At a first look the discussion of continuous quantum systems seems therefore to be hopeless. If we restrict our attention however to small classes of states and channels, with sufficiently simple structure, many problems become tractable. Phase space quantum mechanics, which will be reviewed in this section (see [79, Chapter 5] for details), provides a very powerful tool in this context.

Before we start let us add some remarks to the discussion of Section 2 which we have restricted to finite-dimensional Hilbert spaces. Basically, most of the material considered there can be generalized in a straightforward way, as long as topological issues like continuity and convergence arguments are treated carefully enough. There are of course some caveats (cf. in particular, footnote 4 of Section 2); however, they do not lead to problems in the framework we are going to discuss and can therefore be ignored.

3.3.1. Weyl operators and the CCR

The kinematical structure of a quantum system with d degrees of freedom is usually described by a separable Hilbert space $\mathscr{H}$ and $2d$ self-adjoint operators $Q_1,\ldots,Q_d,P_1,\ldots,P_d$ satisfying the canonical commutation relations $[Q_j,Q_k]=0$, $[P_j,P_k]=0$, $[Q_j,P_k]=i\delta_{jk}\mathbb{1}$. The latter can be rewritten in a more compact form as

$$R_{2j-1}=Q_j, R_{2j}=P_j, \quad j=1,\ldots,d,\ [R_j,R_k]=-\mathrm{i}\sigma_{jk}\ . \tag{3.57}$$

Here σ denotes the *symplectic matrix*

$$\sigma = \mathrm{diag}(J, \ldots, J), \quad J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \tag{3.58}$$

which plays a crucial role for the geometry of classical mechanics. We will call the pair (V, σ) consisting of σ and the $2d$-dimensional real vector space $V = \mathbb{R}^{2d}$ henceforth the *classical phase space*.

The relations in Eq. (3.57) are, however, not sufficient to fix the operators R_j up to unitary equivalence. The best way to remove the remaining physical ambiguities is the study of the unitaries

$$W(x) = \exp(\mathrm{i} x \cdot \sigma \cdot R), \quad x \in V, \quad x \cdot \sigma \cdot R = \sum_{jk=1}^{2d} x_j \sigma_{jk} R_k \tag{3.59}$$

instead of the R_j directly. If the family $W(x)$, $x \in V$ is *irreducible* (i.e. $[W(x), A] = 0$, $\forall x \in V$ implies $A = \lambda \mathbb{1}$ with $\lambda \in \mathbb{C}$) and satisfies [12]

$$W(x)W(x') = \exp\left(-\frac{i}{2} x \cdot \sigma \cdot x'\right) W(x + x'), \tag{3.60}$$

it is called an (irreducible) representation of the *Weyl relations* (on (V, σ)) and the operators $W(x)$ are called *Weyl operators*. By the well-known Stone–von Neumann uniqueness theorem all these representations are mutually unitarily equivalent, i.e. if we have two of them $W_1(x), W_2(x)$, there is a unitary operator U with $UW_1(x)U^* = W_2(x)$ $\forall x \in V$. This implies that it does not matter from a physical point of view which representation we use. The most well-known one is of course the *Schrödinger representation* where $\mathscr{H} = \mathrm{L}^2(\mathbb{R}^d)$ and Q_j, P_k are the usual position and momentum operators.

3.3.2. Gaussian states

A density operator $\rho \in \mathscr{S}(\mathscr{H})$ has *finite second moments* if the expectation values $\mathrm{tr}(\rho Q_j^2)$ and $\mathrm{tr}(\rho P_j^2)$ are finite for all $j = 1, \ldots, d$. In this case we can define the *mean* $m \in \mathbb{R}^{2d}$ and the *correlation matrix* α by

$$m_j = \mathrm{tr}(\rho R_j), \quad \alpha_{jk} + \mathrm{i}\sigma_{jk} = 2\,\mathrm{tr}[(R_j - m_j)\rho(R_k - m_k)]. \tag{3.61}$$

The mean m can be arbitrary, but the correlation matrix α must be real and symmetric and the positivity condition

$$\alpha + \mathrm{i}\sigma \geqslant 0 \tag{3.62}$$

must hold (this is an easy consequence of the canonical commutation relations (3.57)).

Our aim is now to distinguish exactly one state among all others with the same mean and correlation matrix. This is the point where the Weyl operators come into play. Each state $\rho \in \mathscr{S}(\mathscr{H})$ can be characterized uniquely by its *quantum characteristic function* $X \ni x \mapsto \mathrm{tr}[W(x)\rho] \in \mathbb{C}$ which

[12] Note that the CCR (3.57) are implied by the Weyl relations (3.60) but the converse is, in contrast to popular believe, not true: There are representations of the CCR which are unitarily inequivalent to the Schrödinger representation; cf. [134, Section VIII.5] for particular examples. Hence, uniqueness can only be achieved on the level of Weyl operators—which is one major reason to study them.

should be regarded as the quantum Fourier transform of ρ and is in fact the Fourier transform of the Wigner function of ρ [164]. We call ρ *Gaussian* if

$$\mathrm{tr}[W(x)\rho] = \exp(im \cdot x - \tfrac{1}{4}x \cdot \alpha \cdot x) \tag{3.63}$$

holds. By differentiation it is easy to check that ρ has indeed mean m and covariance matrix α.

The most prominent examples for Gaussian states are the ground state ρ_0 of a system of d harmonic oscillators (where the mean is 0 and α is given by the corresponding classical Hamiltonian) and its phase space translates $\rho_m = W(m)\rho W(-m)$ (with mean m and the same α as ρ_0), which are known from quantum optics as *coherent states.* ρ_0 and ρ_m are pure states and it can be shown that a Gaussian state is pure iff $\sigma^{-1}\alpha = -\mathbb{1}$ holds (see [79, Chapter 5]). Examples for mixed Gaussians are temperature states of harmonic oscillators. In one degree of freedom this is

$$\rho_N = \frac{1}{N+1}\sum_{n=0}^{\infty}\left(\frac{N}{N+1}\right)^n |n\rangle\langle n| \ , \tag{3.64}$$

where $|n\rangle\langle n|$ denotes the number basis and N is the mean photon number. The characteristic function of ρ_N is

$$\mathrm{tr}[W(x)\rho_N] = \exp[-\tfrac{1}{2}(N + \tfrac{1}{2})|x|^2] \tag{3.65}$$

and its correlation matrix is simply $\alpha = 2(N + 1/2)\mathbb{1}$

3.3.3. Entangled Gaussians

Let us now consider bipartite systems. Hence the phase space (V, σ) decomposes into a direct sum $V = V_A \oplus V_B$ (where A stands for "Alice" and B for "Bob") and the symplectic matrix $\sigma = \sigma_A \oplus \sigma_B$ is block diagonal with respect to this decomposition. If $W_A(x)$, respectively $W_B(y)$, denote Weyl operators, acting on the Hilbert spaces $\mathscr{H}_A$, $\mathscr{H}_B$, and corresponding to the phase spaces V_A and V_B, it is easy to see that the tensor product $W_A(x) \otimes W_B(y)$ satisfies the Weyl relations with respect to (V, σ). Hence by the Stone–von Neumann uniqueness theorem we can identify $W(x \oplus y)$, $x \oplus y \in V_a \oplus V_B = V$ with $W_A(x) \otimes W_A(y)$. This immediately shows that a state ρ on $\mathscr{H} = \mathscr{H}_A \otimes \mathscr{H}_B$ is a product state iff its characteristic function factorizes. Separability[13] is characterized as follows (we omit the proof, see [170] instead).

Theorem 3.4. *A Gaussian state with covariance matrix α is separable iff there are covariance matrices α_A, α_B such that*

$$\alpha \geqslant \begin{bmatrix} \alpha_A & 0 \\ 0 & \alpha_B \end{bmatrix} \tag{3.66}$$

holds.

This theorem is somewhat similar to Theorem 2.1: It provides a useful criterion as long as abstract considerations are concerned, but not for explicit calculations. In contrast to finite-dimensional

[13] In infinite dimensions we have to define separable states (in slight generalization to Definition 2.5) as a trace-norm convergent convex sum of product states.

systems, however, separability of Gaussian states can be decided by an operational criterion in terms of nonlinear maps between matrices [65]. To state it we have to introduce some terminology first. The key tool is a sequence of $2n + 2m \times 2n + 2m$ matrices α_N, $N \in \mathbb{N}$, written in block matrix notation as

$$\alpha_N = \begin{bmatrix} A_N & C_N \\ C_N^{\mathrm{T}} & B_N \end{bmatrix} . \tag{3.67}$$

Given α_0 the other α_N are recursively defined by

$$A_{N+1} = B_{N+1} = A_N - \mathrm{Re}(X_N) \quad \text{and} \quad C_{N+1} = -\mathrm{Im}(X_N) \tag{3.68}$$

if $\alpha_N - \mathrm{i}\sigma \geqslant 0$ and $\alpha_{N+1} = 0$ otherwise. Here we have set $X_N = C_N(B_N - \mathrm{i}\sigma_B)^{-1}C_N^{\mathrm{T}}$ and the inverse denotes the *pseudoinverse* [14] if $B_N - \mathrm{i}\sigma_B$ is not invertible. Now we can state the following theorem (see [65] for a proof).

Theorem 3.5. *Consider a Gaussian state ρ of a bipartite system with correlation matrix α_0 and the sequence α_N, $N \in \mathbb{N}$ just defined.*

1. *If for some $N \in \mathbb{N}$ we have $A_N - \mathrm{i}\sigma_A \ngeqslant 0$ then ρ is not separable.*
2. *If there is, on the other hand an $N \in \mathbb{N}$ such that $A_N - \|C_N\|\mathbb{1} - \mathrm{i}\sigma_A \geqslant 0$, then the state ρ is separable ($\|C_N\|$ denotes the operator norm of C_N).*

To check whether a Gaussian state ρ is separable or not we have to iterate through the sequence α_N until either condition 1 or 2 holds. In the first case we know that ρ is entangled and separable in the second. Hence, only the question remains whether the whole procedure terminates after a finite number of iterations. This problem is treated in [65] and it turns out that the set of ρ for which separability is decidable after a finite number of steps is the complement of a measure zero set (in the set of all separable states). Numerical calculations indicate in addition that the method converges usually very fast (typically less than five iterations).

To consider ppt states we first have to characterize the transpose for infinite-dimensional systems. There are different ways to do that. We will use the fact that the adjoint of a matrix can be regarded as transposition followed by componentwise complex conjugation. Hence, we define for any (possibly unbounded) operator $A^{\mathrm{T}} = CA^*C$, where $C : \mathscr{H} \to \mathscr{H}$ denotes complex conjugation of the wave function in position representation. This implies $Q_j^{\mathrm{T}} = Q_j$ for position and $P_j^{\mathrm{T}} = -P_j$ for momentum operators. If we insert the *partial* transpose of a bipartite state ρ into Eq. (3.61) we see that the correlation matrix $\tilde{\alpha}_{jk}$ of ρ^{T} picks up a minus sign whenever one of the indices belongs to one of Alice's momentum operators. To be a state $\tilde{\alpha}$ should satisfy $\tilde{\alpha} + \mathrm{i}\sigma \geqslant 0$, but this is equivalent to $\alpha + \mathrm{i}\tilde{\sigma} \geqslant 0$, where in $\tilde{\sigma}$ the corresponding components are reversed i.e. $\tilde{\sigma} = (-\sigma_A) \oplus \sigma_B$. Hence we have shown

[14] A^{-1} is the pseudoinverse of a matrix A if $AA^{-1} = A^{-1}A$ is the projector onto the range of A. If A is invertible A^{-1} is the usual inverse.

Proposition 3.6. *A Gaussian state is ppt iff its correlation matrix α satisfies*

$$\alpha + \mathrm{i}\tilde{\sigma} \geqslant 0 \quad \text{with } \tilde{\sigma} = \begin{bmatrix} -\sigma_A & 0 \\ 0 & \sigma_B \end{bmatrix} . \tag{3.69}$$

The interesting question is now whether the ppt criterion is (for a given number of degrees of freedom) equivalent to separability or not. The following theorem which was proved in [144] for 1×1 systems and in [170] in $1 \times d$ case gives a complete answer.

Theorem 3.7. *A Gaussian state of a quantum system with $1 \times d$ degrees of freedom* (*i.e.* $\dim X_A = 2$ *and* $\dim X_B = 2d$) *is separable iff it is ppt*; *in other words iff the condition of Proposition* 3.6 *holds.*

For other kinds of systems the ppt criterion may fail which means that there are entangled Gaussian states which are ppt. A systematic way to construct such states can be found in [170]. Roughly speaking, it is based on the idea to go to the boundary of the set of ppt covariance matrices, i.e. α has to satisfy Eqs. (3.62) and (3.69) and it has to be a minimal matrix with this property. Using this method explicit examples for ppt and entangled Gaussians are constructed for 2×2 degrees of freedom (cf. [170] for details).

3.3.4. Gaussian channels

Finally, we want to give a short review on a special class of channels for infinite-dimensional quantum systems (cf. [84] for details). To explain the basic idea firstly note that each finite set of Weyl operators ($W(x_j)$, $j = 1, \ldots, N$, $x_j \neq x_k$ for $j \neq k$) is linear independent. This can be checked easily using expectation values of $\sum_j \lambda_j W(x_j)$ in Gaussian states. Hence, linear maps on the space of finite linear combinations of Weyl operators can be defined by $T[W(x)] = f(x)W(Ax)$ where f is a complex-valued function on V and A is a $2d \times 2d$ matrix. If we choose A and f carefully enough, such that some continuity properties match T can be extended in a unique way to a linear map on $\mathscr{B}(\mathscr{H})$—which is, however, in general not completely positive.

This means we have to consider special choices for A and f. The most easy case arises if $f \equiv 1$ and A is a symplectic isomorphism, i.e. $A^{\mathrm{T}}\sigma A = \sigma$. If this holds the map $V \ni x \mapsto W(Ax)$ is a representation of the Weyl relations and therefore unitarily equivalent to the representation we have started with. In other words, there is a unitary operator U with $T[W(x)] = W(Ax) = UW(x)U^*$, i.e. T is unitarily implemented, hence completely positive and, in fact, well known as *Bogolubov transformation.*

If A does not preserve the symplectic matrix, $f \equiv 1$ is no option. Instead, we have to choose f such that the matrices

$$M_{jk} = f(x_j - x_k)\exp\left(-\frac{\mathrm{i}}{2}x_j \cdot \sigma x_k + \frac{\mathrm{i}}{2}Ax_j \cdot \sigma A x_k\right) \tag{3.70}$$

are positive. Complete positivity of the corresponding T is then a standard result of abstract C*-algebra theory (cf. [51]). If the factor f is in addition a Gaussian, i.e. $f(x) = \exp(-\frac{1}{2}x \cdot \beta x)$ for a positive definite matrix β the cp-map T is called a *Gaussian channel.*

A simple way to construct a Gaussian channel is in terms of an ancilla representation. More precisely, if $A : V \to V$ is an arbitrary linear map we can extend it to a symplectic map $V \ni x \mapsto Ax \oplus A'x \in V \oplus V'$, where the symplectic vector space (V', σ') now refers to the environment. Consider now the Weyl operator $W(x) \otimes W'(x') = W(x, x')$ on the Hilbert space $\mathscr{H} \otimes \mathscr{H}'$ associated to the phase space element $x \oplus x' \in V \oplus V'$. Since $A \oplus A'$ is symplectic it admits a unitary Bogolubov transformation $U : \mathscr{H} \otimes \mathscr{H}' \to \mathscr{H} \otimes \mathscr{H}'$ with $U^* W(x, x')U = W(Ax, A'x)$. If ρ' denotes now a Gaussian density matrix on $\mathscr{H}'$ describing the initial state of the environment we get a Gaussian channel by

$$\mathrm{tr}[T^*(\rho)W(x)] = \mathrm{tr}[\rho \otimes \rho' U^* W(x, x')U] = \mathrm{tr}[\rho W(Ax)]\mathrm{tr}[\rho' W(A'x)] \ . \tag{3.71}$$

Hence $T[W(x)] = f(x)W(Ax)$ with $f(x) = \mathrm{tr}[\rho' W(A'x)]$.

Particular examples for Gaussian channels in the case of one degree of freedom are attenuation and amplification channels [81,84]. They are given in terms of a real parameter $k \neq 1$ by $\mathbb{R}^2 \ni x \mapsto Ax = kx \in \mathbb{R}^2$

$$\mathbb{R}^2 \ni x \mapsto A'x = \sqrt{1 - k^2}x \in \mathbb{R}^2 < 1 \tag{3.72}$$

for $k < 1$ and

$$\mathbb{R}^2 \ni (q, p) \mapsto A'(q, p) = (\kappa q, -\kappa p) \in \mathbb{R}^2 \quad \text{with } \kappa = \sqrt{k^2 - 1} \tag{3.73}$$

for $k > 1$. If the environment is initially in a thermal state $\rho_{\tilde{N}}$ (cf. Eq. (3.64)) this leads to

$$T[W(x)] = \exp\left[\frac{1}{2}\left(\frac{|k^2 - 1|}{2} + N_c\right)x^2\right] W(kx) \ , \tag{3.74}$$

where we have set $N_c = |k^2 - 1|\tilde{N}$. If we start initially with a thermal state ρ_N it is mapped by T again to a thermal state $\rho_{N'}$ with mean photon number N' given by

$$N' = k^2 N + \max\{0, k^2 - 1\} + N_c \ . \tag{3.75}$$

If $N_c = 0$ this means that T amplifies ($k > 1$) or damps ($k < 1$) the mean photon number, while $N_c > 0$ leads to additional classical, Gaussian noise. We will reconsider this channel in greater detail in Section 6.

4. Basic tasks

After we have discussed the conceptual foundations of quantum information we will now consider some of its basic tasks. The spectrum ranges here from elementary processes, like teleportation 4.1 or error correction 4.4, which are building blocks for more complex applications, up to possible future technologies like quantum cryptography 4.6 and quantum computing 4.5.

4.1. Teleportation and dense coding

Maybe the most striking feature of entanglement is the fact that otherwise impossible machines become possible if entangled states are used as an additional resource. The most prominent examples are teleportation and dense coding which we want to discuss in this section.

4.1.1. Impossible machines revisited: classical teleportation

We have already pointed out in the introduction that *classical* teleportation, i.e. transmission of quantum information over a classical information channel is impossible. With the material introduced in the last two chapters it is now possible to reconsider this subject in a slightly more mathematical way, which makes the following treatment of *entanglement' enhanced* teleportation more transparent. To "teleport" the state $\rho \in \mathscr{B}^*(\mathscr{H})$ Alice performs a measurement (described by a POV measure $E_1, \dots, E_N \in \mathscr{B}(\mathscr{H})$) on her system and gets a value $x \in X = \{1, \dots, N\}$ with probability $p_x = \operatorname{tr}(E_x \rho)$. These data she communicates to Bob and he prepares a $\mathscr{B}(\mathscr{H})$ system in the state ρ_x. Hence the overall state Bob gets if the experiment is repeated many times is: $\tilde{\rho} = \sum_{x \in X} \operatorname{tr}(E_x \rho)\rho_x$ (cf. Fig. 1.1). The latter can be rewritten as the *composition*

$$\mathscr{B}^*(\mathscr{H}) \xrightarrow{E^*} \mathscr{C}(X)^* \xrightarrow{D^*} \mathscr{B}^*(\mathscr{H})^* \tag{4.1}$$

of the channels

$$\mathscr{C}(X) \ni f \mapsto E(f) = \sum_{x \in X} f(x) E_x \in \mathscr{B}(\mathscr{H}) \tag{4.2}$$

and

$$\mathscr{C}^*(X) \ni p \mapsto D^*(p) = \sum_{x \in X} p_x \rho_x \in \mathscr{B}^*(\mathscr{H}) , \tag{4.3}$$

i.e. $\tilde{\rho} = D^* E^*(\rho)$ and this equation makes sense even if X is not finite. The teleportation is successful if the output state $\tilde{\rho}$ cannot be distinguished from the input state ρ by any statistical experiment, i.e. if $D^* E^*(\rho) = \rho$. Hence the impossibility of classical teleportation can be rephrased simply as $ED \neq \mathrm{Id}$ for all observables E and all preparations D.

4.1.2. Entanglement enhanced teleportation

Let us now change our setup slightly. Assume that Alice wants to send a quantum state $\rho \in \mathscr{B}^*(\mathscr{H})$ to Bob and that she shares an entangled state $\sigma \in \mathscr{B}^*(\mathscr{K} \otimes \mathscr{K})$ and an ideal classical communication channel $\mathscr{C}(X) \to \mathscr{C}(X)$ with him. Alice can perform a measurement $E : \mathscr{C}(X) \to \mathscr{B}(\mathscr{H} \otimes \mathscr{K})$ on the composite system $\mathscr{B}(\mathscr{H} \otimes \mathscr{K})$ consisting of the particle to teleport ($\mathscr{B}(\mathscr{H})$) and her part of the entangled system ($\mathscr{B}(\mathscr{K})$). Then she communicates the classical data $x \in X$ to Bob and he operates with the parameter-dependent operation $D : \mathscr{B}(\mathscr{H}) \to \mathscr{B}(\mathscr{K}) \otimes \mathscr{C}(X)$ appropriately on his particle (cf. Fig. 4.1). Hence, the overall procedure can be described by the channel $T = (E \otimes \mathrm{Id})D$,

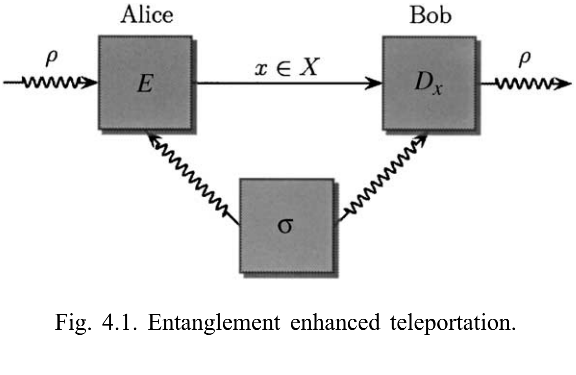


Fig. 4.1. Entanglement enhanced teleportation.

or in analogy to (4.1)

$$\mathscr{B}^*(\mathscr{H}\otimes\mathscr{K}^{\otimes 2})\xrightarrow{E^*\otimes\mathrm{Id}}\mathscr{C}^*(X)\otimes\mathscr{B}^*(\mathscr{K})\xrightarrow{D^*}\mathscr{B}^*(\mathscr{H})\,. \tag{4.4}$$

The teleportation of ρ is successful if

$$T^*(\rho\otimes\sigma):=D^*((E^*\otimes\mathrm{Id})(\rho\otimes\sigma))=\rho \tag{4.5}$$

holds, in other words if there is no statistical measurement which can distinguish the final state $T^*(\rho\otimes\sigma)$ of Bob's particle from the initial state ρ of Alice's input system. The two channels E and D and the entangled state σ form a *teleportation scheme* if Eq. (4.5) holds for all states ρ of the $\mathscr{B}(\mathscr{H})$ system, i.e. if each state of a $\mathscr{B}(\mathscr{H})$ system can be teleported without loss of quantum information.

Assume now that $\mathscr{H}=\mathscr{K}=\mathbb{C}^d$ and $X=\{0,\dots,d^2-1\}$ holds. In this case we can define a teleportation scheme as follows: The entangled state shared by Alice and Bob is a maximally entangled state $\sigma=|\Omega\rangle\langle\Omega|$ and Alice performs a measurement which is given by the one-dimensional projections $E_j=|\Phi_j\rangle\langle\Phi_j|$, where $\Phi_j\in\mathscr{H}\otimes\mathscr{H}$, $j=0,\dots,d^2-1$ is a basis of maximally entangled vectors. If her result is $j=0,\dots,d^2-1$ Bob has to apply the operation $\tau\mapsto U_j^*\tau U_j$ on his partner of the entangled pair, where the $U_j\in\mathscr{B}(\mathscr{H})$, $j=0,\dots,d^2-1$ are an orthonormal family of unitary operators, i.e. $\mathrm{tr}(U_j^*U_k)=d\delta_{jk}$. Hence, the parameter-dependent operation D has the form (in the Schrödinger picture):

$$\mathscr{C}^*(X)\otimes\mathscr{B}^*(\mathscr{H})\ni(p,\tau)\mapsto D^*(p,\tau)=\sum_{j=0}^{d^2-1}p_jU_j^*\tau U_j\in\mathscr{B}^*(\mathscr{H})\,. \tag{4.6}$$

Therefore, we get for $T^*(\rho\otimes\sigma)$ from Eq. (4.5)

$$\mathrm{tr}[T^*(\rho\otimes\sigma)A]=\mathrm{tr}[(E\otimes\mathrm{Id})^*(\rho\otimes\sigma)D(A)] \tag{4.7}$$

$$=\mathrm{tr}\left[\sum_{j=0}^{d^2-1}\mathrm{tr}_{12}[|\Phi_j\rangle\langle\Phi_j|(\rho\otimes\sigma)]U_j^*AU_j\right]. \tag{4.8}$$

$$=\sum_{j=0}^{d^2-1}\mathrm{tr}[(\rho\otimes\sigma)|\Phi_j\rangle\langle\Phi_j|\otimes(U_j^*AU_j)]. \tag{4.9}$$

Here tr_{12} denotes the partial trace over the first two tensor factors (= Alice's qubits). If Ω, the Φ_j and the U_j are related by the equation

$$\Phi_j=(U_j\otimes\mathbb{1})\Omega\,, \tag{4.10}$$

it is a straightforward calculation to show that $T^*(\rho\otimes\sigma)=\rho$ holds as expected [167]. If $d=2$ there is basically a unique choice: the Φ_j, $j=0,\dots,3$ are the four Bell states (cf. Eq. (3.3), $\Omega=\Phi_0$ and the U_j are the identity and the three Pauli matrices. In this way, we recover the standard example for teleportation, published for the first time in [11]. The first experimental realizations are [24,22].

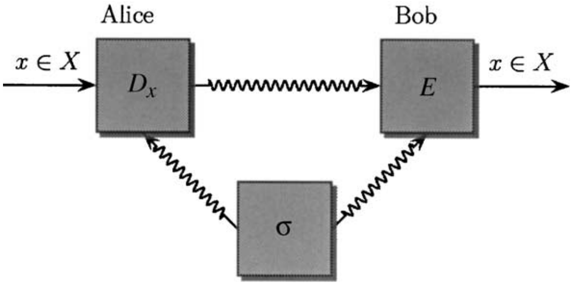


Fig. 4.2. Dense coding.

4.1.3. Dense coding

We have just shown how quantum information can be transmitted via a classical channel, if entanglement is available as an additional resource. Now we are looking at the dual procedure: transmission of classical information over a quantum channel. To send the classical information $x \in X = \{1, \ldots, n\}$ to Bob, Alice can prepare a d-level quantum system in the state $\rho_x \in \mathscr{B}^*(\mathscr{H})$, sends it to Bob and he measures an observable given by positive operators $E_1, \ldots, E_m$. The probability for Bob to receive the signal $y \in X$ if Alice has sent $x \in X$ is $\operatorname{tr}(\rho_x E_y)$ and this defines a classical information channel by (cf. Section 3.2.3)

$$\mathscr{C}^*(X) \ni p \mapsto \left(\sum_{x \in X} p(x) \operatorname{tr}(\rho_x E_1), \ldots, \sum_{x \in X} p(x) \operatorname{tr}(\rho_x E_m) \right) \in \mathscr{C}^*(X) . \tag{4.11}$$

To get an ideal channel we just have to choose mutually orthogonal pure states $\rho_x = |\psi_x\rangle\langle\psi_x|$, $x = 1, \ldots, d$ on Alice's side and the corresponding one-dimensional projections $E_y = |\psi_y\rangle\langle\psi_y|$, $y = 1, \ldots, d$ on Bob's. If $d = 2$ and $\mathscr{H} = \mathbb{C}^2$ it is possible to send one bit classical information via one qubit quantum information. The crucial point is now that the amount of classical information *can be increased* (doubled in the qubit case) if Alice shares an entangled state $\sigma \in \mathscr{S}(\mathscr{H} \otimes \mathscr{H})$ with Bob. To send the classical information $x \in X = \{1, \ldots, n\}$ to Bob, Alice operates on her particle with an operation $D_x : \mathscr{B}(\mathscr{H}) \to \mathscr{B}(\mathscr{H})$, sends it through an (ideal) quantum channel to Bob and he performs a measurement $E_1, \ldots, E_n \in \mathscr{B}(\mathscr{H} \otimes \mathscr{H})$ on *both* particles. The probability for Bob to measure $y \in X$ if Alice has send $x \in X$ is given by

$$\operatorname{tr}[(D_x \otimes \mathrm{Id})^*(\sigma) E_y] \tag{4.12}$$

and this defines the transition matrix of a classical communication channel T. If T is an ideal channel, i.e. if the transition matrix (4.12) is the identity, we will call E, D and σ a *dense coding scheme* (cf. Fig. 4.2).

In analogy to Eq. (4.4) we can rewrite the channel T defined by (4.12) in terms of the composition

$$\mathscr{C}^*(X) \otimes \mathscr{B}^*(\mathscr{H}) \otimes \mathscr{B}^*(\mathscr{H}) \xrightarrow{D^* \otimes \mathrm{Id}} \mathscr{B}^*(\mathscr{H}) \otimes \mathscr{B}^*(\mathscr{H}) \xrightarrow{E^*} \mathscr{C}^*(X) \tag{4.13}$$

of the parameter-dependent operation

$$D : \mathscr{C}^*(X) \otimes \mathscr{B}^*(\mathscr{H}) \to \mathscr{B}^*(\mathscr{H}), \quad p \otimes \tau \mapsto \sum_{j=1}^{n} p_j D_j(\tau) \tag{4.14}$$

and the observable

$$E : \mathscr{C}(X) \to \mathscr{B}(\mathscr{H} \otimes \mathscr{H}), \quad p \mapsto \sum_{j=1}^{n} p_j E_j \ , \tag{4.15}$$

i.e. $T^*(p) = E^* \circ (D^* \otimes \mathrm{Id})(p \otimes \sigma)$. The advantage of this point of view is that it works as well for infinite-dimensional Hilbert spaces and continuous observables.

Finally, let us again consider the case where $\mathscr{H} = \mathbb{C}^d$ and $X = \{1, \ldots, d^2\}$. If we choose as in the last paragraph a maximally entangled vector $\Omega \in \mathscr{H} \otimes \mathscr{H}$, an orthonormal base $\Phi_x \in \mathscr{H} \otimes \mathscr{H}$, $x = 1, \ldots, d^2$ of maximally entangled vectors and an orthonormal family $U_x \in \mathscr{B}(\mathscr{H} \otimes \mathscr{H})$, $x = 1, \ldots, d^2$ of unitary operators, we can construct a dense coding scheme as follows: $E_x = |\Phi_x\rangle\langle\Phi_x|$, $D_x(A) = U_x^* A U_x$ and $\sigma = |\Omega\rangle\langle\Omega|$. If Ω, the Φ_x and the U_x are related by Eq. (4.10) it is easy to see that we really get a dense coding scheme [167]. If $d = 2$ holds, we have to set again the Bell basis for the Φ_x, $\Omega = \Phi_0$ and the identity and the Pauli matrices for the U_x. We recover in this case the standard example of dense coding proposed in [19] and we see that we can transfer two bits via one qubit, as stated above.

4.2. Estimating and copying

The impossibility of classical teleportation can be rephrased as follows: It is impossible to get complete information about the state ρ of a quantum system by *one* measurement on *one* system. However, if we have *many systems*, say N, all prepared in the same state ρ it should be possible to get (with a clever measuring strategy) as much information on ρ as possible, provided N is large enough. In this way, we can circumvent the impossibility of devices like classical teleportation or quantum copying at least in an approximate way.

4.2.1. Quantum state estimation

To discuss this idea in a more detailed way consider a number N of d-level quantum systems, all of them prepared in the same (unknown) state $\rho \in \mathscr{B}^*(\mathscr{H})$. Our aim is to *estimate* the state ρ by measurements on the compound system $\rho^{\otimes N}$. This is described in terms of an observable $E^N : \mathscr{C}(X_N) \to \mathscr{B}(\mathscr{H}^{\otimes N})$ with values in a finite subset[15] $X_N \subset \mathscr{S}(\mathscr{H})$ of the quantum state space $\mathscr{S}(\mathscr{H})$. According to Section 3.2.4 each such E^N is given in terms of a tuple E^N_σ, $\sigma \in X_N$, by $E(f) = \sum_\sigma f(\sigma) E^N_\sigma$; hence, we get for the expectation value of an E_N measurement on systems in the state $\rho^{\otimes N}$ the density matrix $\hat{\rho}_N \in \mathscr{S}(\mathscr{H})$ with matrix elements

$$\langle \phi, \hat{\rho}_N \psi \rangle = \sum_{x \in X_N} \langle \phi, \sigma \psi \rangle E^N_\sigma \ . \tag{4.16}$$

We will call the channel E^N an *estimator* and the criterion for a good estimator E^N is that for any one-particle density operator ρ, the value measured on a state $\rho^{\otimes N}$ is likely to be close to ρ,

[15] This is a severe restriction at this point and physically not very well motivated. There might be more general (i.e. continuous) observables taking their values in the whole state space $\mathscr{S}(\mathscr{H})$ which lead to much better estimates. However, we do not discuss this possibility in order to keep mathematics more elementary.

i.e. that the probability

$$K^N(\omega):=\mathrm{tr}(E^N(\omega)\rho^{\otimes N}) \quad \text{with } E^N(\omega)=\sum_{\sigma\in X_N\cap\omega} E^N_\sigma \tag{4.17}$$

is small if $\omega \subset \mathscr{S}(\mathscr{H})$ is the complement of a small ball around ρ. Of course, we will look at this problem for large N. So the task is to find a whole sequence of observables E^N, $N=1,2,\ldots$, making error probabilities like (4.17) go to zero as $N\to\infty$.

The most direct way to get a family E^N, $N\in\mathbb{N}$ of estimators with this property is to perform a sequence of measurements on each of the N input systems separately. A finite set of observables which leads to a successful estimation strategy is usually called a "quorum" (cf. e.g. [107,162]). E.g. for $d=2$ we can perform alternating measurements of the three spin components. If $\rho=\frac{1}{2}(\mathbb{1}+\vec{x}\cdot\vec{\sigma})$ is the Bloch representation of ρ (cf. Section 2.1.2) we see that the expectation values of these measurements are given by $\frac{1}{2}(1+x_j)$. Hence we get an arbitrarily good estimate if N is large enough. A similar procedure is possible for arbitrary d if we consider the generalized Bloch representation for ρ (see again Section 2.1.2). There are however more efficient strategies based on "entangled" measurements (i.e. the $E_N(\sigma)$ cannot be decomposed into pure tensor products) on the whole input system $\rho^{\otimes N}$ (e.g. [156,99]). Somewhat in between are "adaptive schemes" [63] consisting of separate measurements but the jth measurement depend on the results of $(j-1)$th. We will reconsider this circle of questions in a more quantitative way in Section 7.

4.2.2. Approximate cloning

By virtue of the no-cloning theorem [173], it is impossible to produce M perfect copies of a d-level quantum system if $N<M$ input systems in the common (unknown) state $\rho^{\otimes N}$ are given. More precisely there is no channel $T_{MN}:\mathscr{B}(\mathscr{H}^{\otimes M})\to\mathscr{B}(\mathscr{H}^{\otimes N})$ such that $T^*_{MN}(\rho^{\otimes N})=\rho^{\otimes M}$ holds for all $\rho\in\mathscr{S}(\mathscr{H})$. Using state estimation, however, it is easy to find a device T_{MN} which produces at least approximate copies which become exact in the limit $N,M\to\infty$: If $\rho^{\otimes N}$ is given, we measure the observable E^N and get *the classical data* $\sigma\in X_N\subset\mathscr{S}(\mathscr{H})$, which we use subsequently to prepare M systems in the state $\sigma^{\otimes M}$. In other words, T_{MN} has the form

$$\mathscr{B}^*(\mathscr{H}^{\otimes N})\ni\tau\mapsto\sum_{\sigma\in X_N}\mathrm{tr}(E^N_\sigma\tau)\sigma^{\otimes M}\in\mathscr{B}^*(\mathscr{H}^{\otimes M})\,. \tag{4.18}$$

We immediately see that the probability to get wrong copies coincides exactly with the error probability of the estimator given in Eq. (4.17). This shows first that we get exact copies in the limit $N\to\infty$ and second that the quality of the copies does not depend on the number M of output systems, i.e. the asymptotic rate $\lim_{N,M\to\infty}M/N$ of output systems per input system can be arbitrary large.

The fact that we get classical data at an intermediate step allows a further generalization of this scheme. Instead of just preparing M systems in the state σ detected by the estimator, we can apply first an *arbitrary transformation* $F:\mathscr{S}(\mathscr{H})\to\mathscr{S}(\mathscr{H})$ on the density matrix σ and prepare $F(\sigma)^{\otimes M}$ instead of $\sigma^{\otimes M}$. In this way, we get the channel (cf. Fig. 4.3)

$$\mathscr{B}^*(\mathscr{H}^{\otimes N})\ni\tau\mapsto\sum_{\sigma\in X_N}\mathrm{tr}(E^N_\sigma\tau)F(\sigma)^{\otimes M}\in\mathscr{B}^*(\mathscr{H}^{\otimes M})\,, \tag{4.19}$$

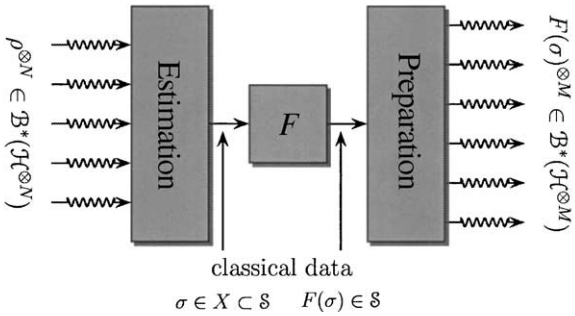

$\rho^{\otimes N} \in \mathcal{B}^*(\mathcal{H}^{\otimes N})$
Estimation
F
Preparation
$F(\sigma)^{\otimes M} \in \mathcal{B}^*(\mathcal{H}^{\otimes M})$
classical data
$\sigma \in X \subset \mathcal{S}$
$F(\sigma) \in \mathcal{S}$

Fig. 4.3. Approximating the impossible machine F by state estimation.

i.e. a *physically realizable* device which approximates the *impossible machine F*. The probability to get a bad approximation of the state $F(\rho)^{\otimes M}$ (if the input state was $\rho^{\otimes N}$) is again given by the error probability of the estimator and we get a perfect realization of F at arbitrary rate as $M, N \to \infty$.

There are in particular two interesting tasks which become possible this way: The first is the "universal not gate" which associates to each pure state of a qubit the unique pure state orthogonal to it [36]. This is a special example of a antiunitarily implemented symmetry operation and therefore not completely positive. The second example is the purification of states [46,100]. Here it is assumed that the input states were once pure but have passed later on a depolarizing channel $|\phi\rangle\langle\phi| \mapsto \vartheta|\phi\rangle\langle\phi| + (1-\vartheta)\mathbb{1}/d$. If $\vartheta > 0$ this map is invertible but its inverse does not describe an allowed quantum operation because it maps some density operators to operators with negative eigenvalues. Hence the reversal of noise is not possible with a one-shot operation but can be done with high accuracy if enough input systems are available. We rediscuss this topic in Section 7.

4.3. Distillation of entanglement

Let us now return to entanglement. We have seen in Section 4.1 that maximally entangled states play a crucial role for processes like teleportation and dense coding. In practice however entanglement is a rather fragile property: If Alice produces a pair of particles in a maximally entangled state $|\Omega\rangle\langle\Omega| \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and distributes one of them over a great distance to Bob, both end up with a mixed state ρ which contains much less entanglement then the original and which cannot be used any longer for teleportation. The latter can be seen quite easily if we try to apply the qubit teleportation scheme (Section 4.1.2) with a non-maximally entangled isotropic state (Eq. (3.15) with $\lambda > 0$) instead of Ω.

Hence the question arises, whether it is possible to recover $|\Omega\rangle\langle\Omega|$ from ρ, or, following the reasoning from the last section, at least a small number of (almost) maximally entangled states from a large number N of copies of ρ. However, since the distance between Alice and Bob is big (and quantum communication therefore impossible) only LOCC operations (Section 3.2.6) are available for this task (Alice and Bob can only operate on their respective particles, drop some of them and communicate classically with one another). This excludes procedures like the purification scheme just sketched, because we would need "entangled" measurements to get an asymptotically exact estimate for the state ρ. Hence, we need a sequence of LOCC channels

$$T_N : \mathscr{B}(\mathbb{C}^{d_N} \otimes \mathbb{C}^{d_N}) \to \mathscr{B}(\mathscr{H}_A^{\otimes N} \otimes \mathscr{H}_B^{\otimes N}) \tag{4.20}$$

such that

$$\| T_N^*(\rho^{\otimes N}) - |\Omega_N\rangle\langle\Omega_N| \|_1 \to 0 \quad \text{for } N \to \infty \tag{4.21}$$

holds, with a sequence of maximally entangled vectors $\Omega_N \in \mathbb{C}^{d_N} \otimes \mathbb{C}^{d_N}$. Note that we have to use here the natural isomorphism $\mathscr{H}_A^{\otimes N} \otimes \mathscr{H}_B^{\otimes N} \cong (\mathscr{H}_A \otimes \mathscr{H}_B)^{\otimes N}$, i.e. we have to reshuffle $\rho^{\otimes N}$ such that the first N tensor factors belong to Alice ($\mathscr{H}_A$) and the last N to Bob ($\mathscr{H}_B$). If confusion can be avoided we will use this isomorphism in the following without a further note. We will call a sequence of LOCC channels, T_N satisfying (4.21) with a state $\rho \in \mathscr{S}(\mathscr{H}_A \otimes \mathscr{H}_B)$ a *distillation scheme* for ρ and ρ is called *distillable* if it admits a distillation scheme. The *asymptotic rate* with which maximally entangled states can be distilled with a given protocol is

$$\liminf_{n\to\infty} \log_2(d_N)/N \; . \tag{4.22}$$

This quantity will become relevant in the framework of entanglement measures (Section 5).

4.3.1. Distillation of pairs of qubits

Concrete distillation protocols are in general rather complicated procedures. We will sketch in the following how any pair of entangled qubits can be distilled. The first step is a scheme proposed for the first time by Bennett et al. [12]. It can be applied if the maximally entangled fraction $\mathscr{F}$ (Eq. (3.4)) is greater than 1/2. As indicated above, we assume that Alice and Bob share a large amount of pairs in the state ρ, so that the total state is $\rho^{\otimes N}$. To obtain a smaller number of pairs with a higher $\mathscr{F}$ they proceed as follows:

1. First they take two pairs (let us call them pairs 1 and 2), i.e. $\rho \otimes \rho$ and apply to each of them the twirl operation $P_{U\bar{U}}$ associated to isotropic states (cf. Eq. (3.18)). This can be done by LOCC operations in the following way: Alice selects at random (respecting the Haar measure on $U(2)$) a unitary operator U applies it to her qubits and sends to Bob which transformation she has chosen; then he applies $\bar{U}$ to his particles. They end up with two isotropic states $\tilde{\rho} \otimes \tilde{\rho}$ with the same maximally entangled fraction as ρ.
2. Each party performs the unitary transformation
$$U_{\mathrm{XOR}} : |a\rangle \otimes |b\rangle \mapsto |a\rangle \otimes |a + b \bmod 2\rangle \tag{4.23}$$
on his/her members of the pairs.
3. Finally, Alice and Bob perform local measurements in the basis $|0\rangle, |1\rangle$ on pair 1 and discards it afterwards. If the measurements agree, pair 2 is kept and has a higher $\mathscr{F}$. Otherwise pair 2 is discarded as well.

If this procedure is repeated over and over again, it is possible to get states with an arbitrarily high $\mathscr{F}$, but we have to sacrifice more and more pairs and the asymptotic rate is zero. To overcome this problem we can apply the scheme above until $\mathscr{F}(\rho)$ is high enough such that $1 + \mathrm{tr}(\rho \ln \rho) \geqslant 0$ holds and then we continue with another scheme called hashing [16] which leads to a non-vanishing rate.

If finally $\mathscr{F}(\rho) \leqslant 1/2$ but ρ is entangled, Alice and Bob can increase $\mathscr{F}$ for some of their particles by *filtering operations* [9,67]. The basic idea is that Alice applies an instrument $T : \mathscr{C}(X) \otimes \mathscr{B}(\mathscr{H}) \to \mathscr{B}(\mathscr{H})$ with two possible outcomes $(X = \{1, 2\})$ to her particles. Hence, the state becomes $\rho \mapsto p_x^{-1}(T_x \otimes \mathrm{Id})^*(\rho)$, $x = 1, 2$ with probability $p_x = \mathrm{tr}[T_x^*(\rho)]$ (cf. Section 3.2.5 in particular Eq. (3.50) for the definition of T_x). Alice communicates her measuring result x to Bob and if $x = 1$ they keep the particle otherwise $(x = 2)$ they discard it. If the instrument T was correctly chosen Alice and Bob end up with a state $\tilde{\rho}$ with higher maximally entangled fraction. To find an appropriate T firstly note that there are $\psi \in \mathscr{H} \otimes \mathscr{H}$ with $\langle \psi, (\mathrm{Id} \otimes \Theta)\rho\psi \rangle \leqslant 0$ (this follows from Theorem 2.4.3 since ρ is by assumption entangled) and second that we can write each vector $\psi \in \mathscr{H} \otimes \mathscr{H}$ as $(X_\psi \otimes \mathbb{1})\Phi_0$ with the Bell state Φ_0 and an appropriately chosen operator X_ψ (see Section 3.1.1). Now we can define T in terms of the two operations T_1, T_2 (cf. Eq. (3.52)) with

$$T_1(A) = X_\psi^* A X_\psi^{-1}, \quad \mathrm{Id} - T_1 = T_2 \ . \tag{4.24}$$

It is straightforward to check that we end up with

$$\tilde{\rho} = \frac{(T_x \otimes \mathrm{Id})^*(\rho)}{\mathrm{tr}[(T_x \otimes \mathrm{Id})^*(\rho)]} \ , \tag{4.25}$$

such that $\mathscr{F}(\tilde{\rho}) > 1/2$ holds and we can continue with the scheme described in the previous paragraph.

4.3.2. Distillation of isotropic states

Consider now an entangled isotropic state ρ in d dimensions, i.e. we have $\mathscr{H} = \mathbb{C}^d$ and $0 \leqslant \mathrm{tr}(\tilde{F}\rho) \leqslant 1$ (with the operator $\tilde{F}$ of Section 3.1.3). Each such state is distillable via the following scheme [27,85]: First, Alice and Bob apply a filter operation $T : \mathscr{C}(X) \otimes \mathscr{B}(\mathscr{H}) \to \mathscr{B}(\mathscr{H})$ on their respective particle given by $T_1(A) = PAP$, $T_2 = 1 - T_1$ where P is the projection onto a two-dimensional subspace. If both measure the value 1 they get a qubit pair in the state $\tilde{\rho} = (T_1 \otimes T_1)(\rho)$. Otherwise they discard their particles (this requires classical communication). Obviously, the state $\tilde{\rho}$ is entangled (this can be easily checked), hence they can proceed as in the previous subsection.

The scheme just proposed can be used to show that each state ρ which violates the reduction criterion (cf. Section 2.4.3) can be distilled [85]. The basic idea is to project ρ with the twirl $P_{U\bar{U}}$ (which is LOCC as we have seen above; cf. Section 4.3.1) to an isotropic state $P_{U\bar{U}}(\rho)$ and to apply the procedure from the last paragraph afterwards. We only have to guarantee that $P_{U\bar{U}}(\rho)$ is entangled. To this end use a vector $\psi \in \mathscr{H} \otimes \mathscr{H}$ with $\langle \psi, (\mathbb{1} \otimes \mathrm{tr}_1(\rho) - \rho)\psi \rangle < 0$ (which exists by assumption since ρ violates the reduction criterion) and to apply the filter operation given by ψ via Eq. (4.24).

4.3.3. Bound entangled states

It is obvious that separable states are not distillable, because an LOCC operation map separable states to separable states. However, is each entangled state distillable? The answer, maybe somewhat surprising, is no and an entangled state which is not distillable is called *bound entangled* [87] (distillable states are sometimes called *free entangled*, in analogy to thermodynamics). Examples of bound entangled states are all ppt entangled states [87]: This is an easy consequence of the fact that each separable channel (and therefore each LOCC channel as well) maps ppt states to ppt states (this is easy to check), but a maximally entangled state is never ppt. It is not yet known, whether bound entangled npt states exists, however, there are at least some partial results: (1) It is sufficient to solve this question for Werner states, i.e. if we can show that each npt Werner state is distillable it follows that all npt states are distillable [85]. (2) Each npt Gaussian state is distillable [64]. (3) For each $N \in \mathbb{N}$ there is an npt Werner state ρ which is not "N-copy distillable", i.e. $\langle \psi, \rho^{\otimes N} \psi \rangle \geqslant 0$ holds for each pure state ψ with exactly two Schmidt summands [55,58]. This gives some evidence for the existence of bound entangled npt states because ρ is distillable iff it is N-copy distillability for some N [87,55,58].

Since bound entangled states cannot be distilled, they cannot be used for teleportation. Nevertheless bound entanglement can produce a non-classical effect, called "activation of bound entanglement" [92]. To explain the basic idea, assume that Alice and Bob share *one* pair of particles in a distillable state ρ_{f} and many particles in a bound entangled state ρ_{b}. Assume in addition that ρ_{f} cannot be used for teleportation, or, in other words if ρ_{f} is used for teleportation the particle Bob receives is in a state σ' which differs from the state σ Alice has send. This problem cannot be solved by distillation, since Alice and Bob share only one pair of particles in the state ρ_{f}. Nevertheless, they can try to apply an appropriate filter operation on ρ to get with a certain probability a new state which leads to a better quality of the teleportation (or, if the filtering fails, to get nothing at all). It can be shown, however [88], that there are states ρ_{f} such that the error occurring in this process (e.g. measured by the trace norm distance of σ and σ') is always above a certain threshold. This is the point where the bound entangled states ρ_{b} come into play: If Alice and Bob operate with an appropriate protocol on ρ_{f} and many copies of ρ_{b} the distance between σ and σ' can be made arbitrarily small (although the probability to be successful goes to zero). Another example for an activation of bound entanglement is related to distillability of npt states: If Alice and Bob share a certain ppt-entangled state as additional resource each npt state ρ becomes distillable (even if ρ is bound entangled) [60,104]. For a more detailed survey of the role of bound entanglement and further references see [91].

4.4. *Quantum error correction*

If we try to distribute quantum information over large distances or store it for a long time in some sort of "quantum memory" we always have to deal with "decoherence effects", i.e. unavoidable interactions with the environment. This results in a significant information loss, which is particularly bad for the functioning of a quantum computer. Similar problems arise as well in a classical computer, but the methods used there to circumvent the problems cannot be transferred to the quantum regime. E.g. the most simple strategy to protect classical information against noise is redundancy: instead of storing the information once we make three copies and decide during readout by a majority vote which bit to take. It is easy to see that this reduces the probability of an error from order ϵ to ϵ^2. Quantum mechanically however such a procedure is forbidden by the no cloning theorem.

Nevertheless, quantum error correction is possible although we have to do it in a more subtle way than just copying; this was observed for the first time independently in [39,146]. Let us consider first the general scheme and assume that $T : \mathscr{B}(\mathscr{K}) \to \mathscr{B}(\mathscr{K})$ is a noisy quantum channel. To send quantum systems of type $\mathscr{B}(\mathscr{H})$ undisturbed through T we need an *encoding channel* $E : \mathscr{B}(\mathscr{K}) \to \mathscr{B}(\mathscr{H})$ and a decoding channel $D : \mathscr{B}(\mathscr{H}) \to \mathscr{B}(\mathscr{K})$ such that $ETD = \mathrm{Id}$ holds, respectively $D^* T^* E^* = \mathrm{Id}$, in the Schrödinger picture; cf. Fig. 4.4.

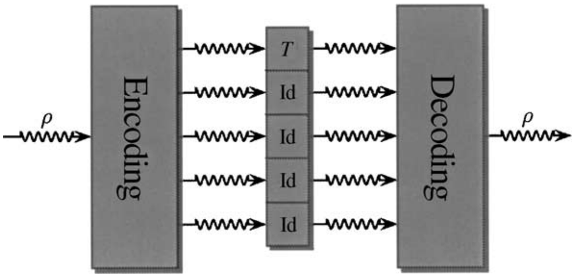


Fig. 4.4. Five-bit quantum code: encoding one qubit into five and correcting one error.

A powerful error correction scheme should not be restricted to one particular type of error, i.e. one particular noisy channel T. Assume instead that $\mathfrak{E} \subset \mathscr{B}(\mathscr{K})$ is a linear subspace of "error operators" and T is any channel given by

$$T_*(\rho) = \sum_j F_j \rho F_j^*, \quad F_j \in \mathfrak{E} . \tag{4.26}$$

An isometry $V : \mathscr{H} \to \mathscr{K}$ is called an *error correcting code* for $\mathfrak{E}$ if for each T of form (4.26) there is a decoding channel $D : \mathscr{B}(\mathscr{H}) \to \mathscr{B}(\mathscr{K})$ with $D_*(T(V\rho V^*)) = \rho$ for all $\rho \in \mathscr{S}(\mathscr{H})$. By the theory of Knill and Laflamme [103] this is equivalent to the factorization condition

$$\langle V\psi, F_j^* F_k V\phi \rangle = \omega(F_j^* F_k)\langle \psi, \phi \rangle , \tag{4.27}$$

where $\omega(F_j^* F_k)$ is a factor which does not depend on the arbitrary vectors $\psi, \phi \in \mathscr{H}$.

The most relevant examples of error correcting codes are those which generalize the classical idea of sending multiple copies in a certain sense. This means we encode a small number N of d-level systems into a big number $M \gg N$ of systems of the same type, which are then transmitted and decoded back into N systems afterwards. During the transmission $K < M$ arbitrary errors are allowed. Hence, we have $\mathscr{H} = \mathscr{H}_1^{\otimes N}$, $\mathscr{K} = \mathscr{H}_1^{\otimes M}$ with $\mathscr{H}_1 = \mathbb{C}^d$ and T is an arbitrary tensor product of K noisy channels S_j, $j = 1, \ldots, K$ and $M - K$ ideal channels Id. The most well-known code for this type of error is the "five-bit code" where one qubit is encoded into five and one error is corrected [16] (cf. Fig. 4.4 for $N = 1, M = 5$ and $K = 1$). To define the corresponding error space $\mathfrak{E}$ consider the finite sets $X = \{1, \ldots, N\}$ and $Y = \{1 + N, \ldots, M + N\}$ and define first for each subset $Z \subset Y$:

$$\begin{aligned}\mathfrak{E}(Z) = \text{span } \{&A_1 \otimes \cdots \otimes A_M \in \mathscr{B}(\mathscr{K}) | \\ &A_j \in \mathscr{B}(\mathscr{H}_1) \text{ arbitrary for } j + N \in Z, \; A_j = \mathbb{1} \text{ otherwise}\} .\end{aligned} \tag{4.28}$$

$\mathfrak{E}$ is now the span of all $\mathfrak{E}(Z)$ with $|Z| \leqslant K$ (i.e. the length of Z is less or equal to K). We say that an error correcting code for this particular $\mathfrak{E}$ *corrects K errors*.

There are several ways to construct error correcting codes (see e.g. [70,38,4]). Most of these methods are somewhat involved however and require knowledge from classical error correction which we want to skip. Therefore, we will only present the scheme proposed in [137], which is quite easy to describe and admits a simple way to check the error correction condition. Let us sketch first the general scheme. We start with an undirected graph Γ with two kinds of vertices: A set of input vertices, labeled by X and a set of output vertices labeled by Y. The links of the graph are given by the adjacency matrix, i.e. an $N + M \times N + M$ matrix Γ with $\Gamma_{jk} = 1$ if node k and j are

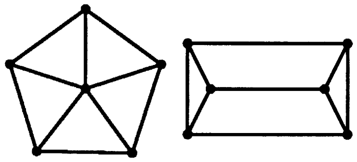
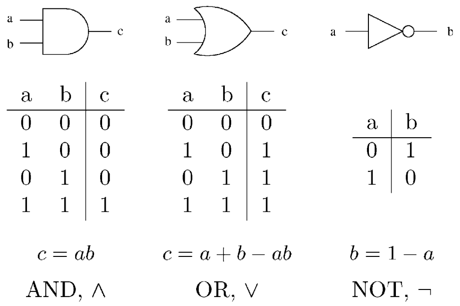


Fig. 4.5. Two graphs belonging to (equivalent) five bit codes. The input node can be chosen in both cases arbitrarily.



a	b	c
0	0	0
1	0	0
0	1	0
1	1	1

$c = ab$

AND, $\wedge$

a	b	c
0	0	0
1	0	1
0	1	1
1	1	1

$c = a + b - ab$

OR, $\vee$

a	b
0	1
1	0

$b = 1 - a$

NOT, $\neg$

Fig. 4.6. Symbols and definition for the three elementary gates AND, OR and NOT.

linked and $\Gamma_{jk} = 0$ otherwise. With respect to Γ we can define now an isometry $V_\Gamma : \mathscr{H}_1^{\otimes N} \to \mathscr{H}_1^{\otimes M}$ by

$$\langle j_{N+1} \dots j_{N+M} | V_\Gamma | j_1 \dots j_N \rangle = \exp\left(\frac{i\pi}{d} \vec{j} \cdot \Gamma \vec{j} \right) \tag{4.29}$$

with $\vec{j} = (j_1, \dots, j_{N+M}) \in \mathbb{Z}_d^{N+M}$ (where $\mathbb{Z}_d$ denotes the cyclic group with d elements). There is an easy condition under which V_Γ is an error correcting code. To write it down we need the following additional terminology: We say that an error correcting code $V : \mathscr{H}_1^{\otimes N} \to \mathscr{H}_1^{\otimes M}$ *detects* the *error configuration* $Z \subset Y$ if

$$\langle V\psi, FV\phi \rangle = \omega(F) \langle \psi, \phi \rangle \quad \forall F \in \mathfrak{E}(Z) \tag{4.30}$$

holds. With Eq. (4.27) it is easy to see that V corrects K errors iff it detects all error configurations of length $2K$ or less. Now we have the following theorem:

Theorem 4.1. *The quantum code V_Γ defined in Eq.* (4.29) *detects the error configuration $Z \subset Y$ if the system of equations*

$$\sum_{l \in X \cup Z} \Gamma_{kl} g_l = 0, \quad k \in Y \setminus E, \quad g_l \in \mathbb{Z}_d \tag{4.31}$$

implies that

$$g_l = 0, \ l \in X \quad \text{and} \quad \sum_{l \in Z} \Gamma_{kl} g_l = 0, \ k \in X \tag{4.32}$$

holds.

We omit the proof, see [137] instead. Two particular examples (which are equivalent!) are given in Fig. 4.5. In both cases we have $N = 1$, $M = 5$ and $K = 1$ i.e. one input node, which can be chosen arbitrarily, five output nodes and the corresponding codes correct one error. For a more detailed survey on quantum error correction, in particular for more examples we refer to [20].

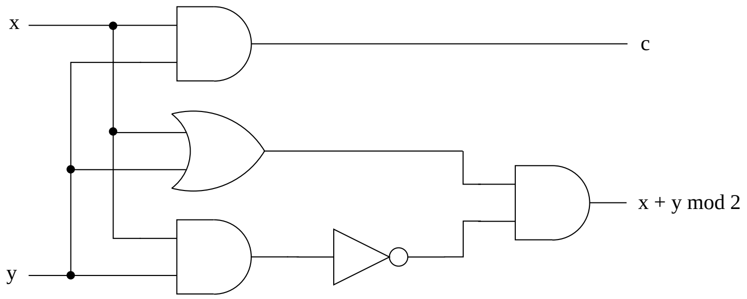


Fig. 4.7. Half-adder circuit as an example for a Boolean network.

4.5. Quantum computing

Quantum computing is without a doubt the most prominent and most far reaching application of quantum information theory, since it promises on the one hand, "exponential speedup" for some problems which are "hard to solve" with a classical computer, and gives completely new insights into classical computing and complexity theory on the other. Unfortunately, an exhaustive discussion would require its own review article. Hence, we are only able to give a short overview (see Part II of [122] for a more complete presentation and for further references).

4.5.1. The network model of classical computing

Let us start with a brief (and very informal) introduction to classical computing (for a more complete review and hints for further reading see [122, Chapter 3]). What we need first is a *mathematical model* for computation. There are, in fact, several different choices and the Turing machine [152] is the most prominent one. More appropriate for our purposes is, however, the so-called *network model*, since it allows an easier generalization to the quantum case. The basic idea is to interpret a classical (deterministic) computation as the evaluation of a map $f : \mathbb{B}^N \rightarrow \mathbb{B}^M$ (where $\mathbb{B} = \{0, 1\}$ denotes the field with two elements) which maps N input bits to M output bits. If $M = 1$ holds f is called a Boolean function and it is for many purposes sufficient to consider this special case—each general f is in fact a Cartesian product of Boolean functions. Particular examples are the three elementary gates AND, OR and NOT defined in Fig. 4.6 and arbitrary algebraic expressions constructed from them: e.g. the XOR gate $(x, y) \mapsto x + y \bmod 2$ which can be written as $(x \vee y) \wedge \neg(x \wedge y)$. It is now a standard result of Boolean algebra that each Boolean function can be represented in this way and there are in general many possibilities to do this. A special case is the *disjunctive normal form* of f; cf [161]. To write such an expression down in form of equations is, however, somewhat confusing. f is therefore expressed most conveniently in graphical form as a *circuit* or *network*, i.e. a graph C with nodes representing elementary gates and edges ("wires") which determine how the gates should be composed; cf. Fig. 4.7 for an example. A *classical computation* can now be defined as a circuit applied to a specified string of input bits.

Variants of this model arise if we replace AND, OR and NOT by another (finite) set G of elementary gates. We only have to guarantee that each function f can be expressed as a composition of elements from G. A typical example for G is the set which contains only the NAND gate $(x, y) \mapsto x \uparrow y = \neg(x \wedge y)$. Since AND, OR and NOT can be rewritten in terms of NAND (e.g. $\neg x = x \uparrow x$) we can calculate each Boolean function by a circuit of NAND gates.

4.5.2. Computational complexity

One of the most relevant questions within classical computing, and the central subject of computational complexity, is whether a given problem is easy to solve or not, where "easy" is defined in terms of the scaling behavior of the resources needed in dependence of the size of the input data. In the following we will give a rough survey over the most basic aspects of this field, while we refer the reader to [124] for a detailed presentation.

To start with, let us specify the basic question in greater detail. First of all the problems we want to analyze are *decision problems* which only give the two possible values "yes" and "no". They are mathematically described by Boolean functions acting on bit strings of arbitrary size. A well-known example is the factoring problem given by the function fac with $\mathrm{fac}(m,l)=1$ if m (more precisely the natural number represented by m) has a divisor less then l and $\mathrm{fac}(m,l)=0$ otherwise. Note that many tasks of classical computation can be reformulated this way, so that we do not get a severe loss of generality. The second crucial point we have to clarify is the question what exactly are the resources we have mentioned above and how we have to quantify them. A natural physical quantity which come into mind immediately is the time needed to perform the computation (space is another candidate, which we do not discuss here, however). Hence, the question we have to discuss is how the computation time t depends on the size L of the input data x (i.e. the length L of the smallest register needed to represent x as a bit string).

However, a precise definition of "computation time" is still model dependent. For a Turing machine we can take simply the number of head movements needed to solve the problem, and in the network model we choose the number of steps needed to execute the whole circuit, if gates which operate on different bits are allowed to work simultaneously.[16] Even with a fixed type of model the functional behavior of t depends on the set of elementary operations we choose, e.g. the set of elementary gates in the network model. It is therefore useful to divide computational problems into *complexity classes* whose definitions do not suffer under model-dependent aspects. The most fundamental one is the class **P** which contains all problems which can be computed in "polynomial time", i.e. t is, as a function of L, bounded from above by a polynomial. The model independence of this class is basically the content of the strong Church Turing hypotheses which states, roughly speaking, that each model of computation can be simulated in polynomial time on a probabilistic Turing machine.

Problems of class **P** are considered "easy", everything else is "hard". However, even if a (decision) problem is hard the situation is not hopeless. E.g. consider the factoring problem fac described above. It is generally believed (although not proved) that this problem is not in class **P**. But if somebody gives us a divisor $p<l$ of m it is easy to check whether p is really a factor, and if the answer is true we have computed $\mathrm{fac}(m,l)$. This example motivates the following definition: A decision problem f is in class **NP** ("non-deterministic polynomial time") if there is a Boolean function f' in class **P** such that $f'(x,y)=1$ for some y implies $f(x)$. In our example fac' is obviously defined by $\mathrm{fac}'(m,l,p)=1 \Leftrightarrow p<l$ and p is a devisor of m. It is obvious that **P** is a subset of **NP** the other inclusion however is rather non-trivial. The conjecture is that $\mathbf{P}\neq\mathbf{NP}$ holds and great parts of

[16] Note that we have glanced over a lot of technical problems at this point. The crucial difficulty is that each circuit C_N allows only the computation of a Boolean function $f_N:\mathbb{B}^N\to\mathbb{B}$ which acts on input data of length N. Since we are interested in answers for arbitrary finite length inputs a sequence C_N, $N\in\mathbb{N}$ of circuits with appropriate uniformity properties is needed; cf. [124] for details.

complexity theory are based on it. Its proof (or disproof), however, represents one of the biggest open questions of theoretical informatics.

To introduce a third complexity class we have to generalize our point of view slightly. Instead of a function $f:\mathbb{B}^N \to \mathbb{B}^M$ we can look at a noisy classical T which sends the input value $x \in \mathbb{B}^N$ to a probability distribution T_{xy}, $y \in \mathbb{B}^M$ on $\mathbb{B}^M$ (i.e. T_{xy} is the transition matrix of the classical channel T; cf. Section 3.2.3). Roughly speaking, we can interpret such a channel as a *probabilistic computation* which can be realized as a circuit consisting of "probabilistic gates". This means there are several different ways to proceed at each step and we use a classical random number generator to decide which of them we have to choose. If we run our device several times on the same input data x we get different results y with probability T_{xy}. The crucial point is now that we can allow some of the outcomes to be wrong as long as there is an easy way (i.e. a class **P** algorithm) to check the validity of the results. Hence, we define **BPP** ("bounded error probabilistic polynomial time") as the class of all decision problems which admit a polynomial time probabilistic algorithm with error probability less than $1/2 - \varepsilon$ (for fixed ε). It is obvious that $\mathbf{P} \subset \mathbf{BPP}$ holds but the relation between **BPP** and **NP** is not known.

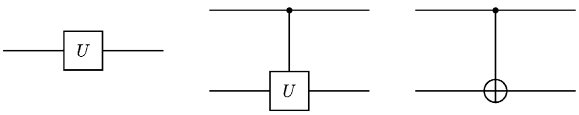
4.5.3. Reversible computing

In the last subsection we have discussed the time needed to perform a certain computation. Other physical quantities which seem to be important are space and energy. Space can be treated in a similar way as time and there are in fact space-related complexity classes (e.g. **PSPACE** which stands for "polynomial space"). Energy, however, is different, because it turns surprisingly out that it is possible to do any calculation *without expending any energy*! One source of energy consumption in a usual computer is the intrinsic irreversibility of the basic operations. E.g. a basic gate like AND maps two input bits to one output bit, which obviously implies that the input cannot be reconstructed from the output. In other words: one bit of information is erased during the operation of the AND gate; hence a small amount of energy is dissipated to the environment. A thermodynamic analysis, known as Landauer's principle, shows that this energy loss is at least $k_B T \ln 2$, where T is the temperature of the environment [106].

If we want to avoid this kind of energy dissipation we are restricted to reversible processes, i.e. it should be possible to reconstruct the input data from the output data. This is called *reversible computation* and it is performed in terms of *reversible gates*, which in turn can be described by invertible functions $f:\mathbb{B}^N \to \mathbb{B}^N$. This does not restrict the class of problems which can be solved however: We can repackage a non-invertible function $f:\mathbb{B}^N \to \mathbb{B}^M$ into an invertible one $f':\mathbb{B}^{N+M} \to \mathbb{B}^{N+M}$ simply by $f'(x,0) = (x, f(x))$ and an appropriate extension to the rest of $\mathbb{B}^{N+M}$. It can be even shown that a reversible computer performs as good as a usual one, i.e. an "irreversible" network can be simulated in polynomial time by a reversible one. This will be of particular importance for quantum computing, because a reversible computer is, as we will see soon, a special case of a quantum computer.

4.5.4. The network model of a quantum computer

Now we are ready to introduce a mathematical model for quantum computation. To this end we will generalize the network model discussed in Section 4.5.1 to the network model of quantum computation.



$|x\rangle \mapsto U|x\rangle$

One qubit gate.

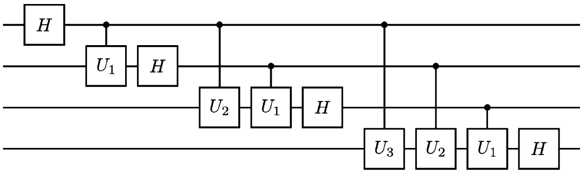
$|0x\rangle \mapsto |0\rangle \otimes |x\rangle$
$|1x\rangle \mapsto |1\rangle \otimes U|x\rangle$

Controlled U gate.

$|0x\rangle \mapsto |0\rangle \otimes |x\rangle$
$|1x\rangle \mapsto |1\rangle \otimes |\neg x\rangle$

CNOT gate.

Fig. 4.8. Universal sets of quantum gates.



$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad U_k = \begin{bmatrix} 1 & 0 \\ 0 & e^{2^{-k}\pi} \end{bmatrix}$$

Fig. 4.9. Quantum circuit for the discrete Fourier transform on a 4-qubit register.

A classical computer operates by a network of gates on a finite number of classical bits. A quantum computer operates on a finite number of qubits in terms of a network of *quantum gates*—this is the rough idea. To be more precise consider the Hilbert space $\mathscr{H}^{\otimes N}$ with $\mathscr{H} = \mathbb{C}^2$ which describes a *quantum register* consisting of N qubits. In $\mathscr{H}$ there is a preferred set $|0\rangle, |1\rangle$ of orthogonal states, describing the two values a classical bit can have. Hence, we can describe each possible value x of a classical register of length N in terms of the *computational basis* $|x\rangle = |x_1\rangle \otimes \cdots \otimes |x_N\rangle$, $x \in \mathbb{B}^N$. A quantum gate is now nothing else but a unitary operator acting on a small number of qubits (preferably 1 or 2) and a quantum network is a graph representing the composition of elementary gates taken from a small set G of unitaries. A *quantum computation* can now be defined as the application of such a network to an input state ψ of the quantum register (cf. Fig. 4.9 for an example). Similar to the classical case the set G should be universal; i.e. each unitary operator on a quantum register of arbitrary length can be represented as a composition of elements from G. Since the group of unitaries on a Hilbert space is continuous, it is not possible to do this with a finite set G. However, we can find at least suitably small sets which have the chance to be realizable technically (e.g. in an ion-trap) somehow in the future. Particular examples are on the one hand the controlled U operations and the set consisting of CNOT and all one-qubit gates on the other (cf. Fig. 4.8; for a proof of universality see Section 4.5 of [122]).

Basically, we could have considered arbitrary quantum operations instead of only unitaries as gates. However in Section 3.2.1, we have seen that we can implement each operation unitarily if we add an ancilla to the systems. Hence, this kind of generalization is already covered by the model. (As long as non-unitarily implemented operations are a desired feature. Decoherence effect due to unavoidable interaction with the environment are a completely different story; we come back to this point at the end of the subsection.) The same holds for measurements at intermediate steps and subsequent conditioned operations. In this case we get basically the same result with a different network where all measurements are postponed to the end. (Often it is however very useful to allow measurements at intermediate steps as we will see in the next subsection.)

Having a mathematical model of quantum computers in mind we are now ready to discuss how it would work in principle.

1. The first step is in most cases preprocessing of the input data on a classical computer. E.g. the Shor algorithm for the factoring problem does not work if the input number m is a pure prime power. However, in this case there is an efficient classical algorithm. Hence, we have to check first whether m is of this particular form and use this classical algorithm where appropriate.
2. In the next step we have to prepare the quantum register based on these preprocessed data. This means in the most simple case to write classical data, i.e. to prepare the state $|x\rangle \in \mathscr{H}^{\otimes N}$ if the (classical) input is $x \in \mathbb{B}^N$. In many cases, however, it might be more intelligent to use a superposition of several $|x\rangle$, e.g. the state
$$\Psi = \frac{1}{\sqrt{2^N}} \sum_{x \in \mathbb{B}^N} |x\rangle \ , \tag{4.33}$$
which represents actually the superposition of all numbers the registers can represent—this is indeed the crucial point of quantum computing and we come back to it below.
3. Now we can apply the quantum circuit C to the input state ψ and after the calculation we get the output state $U\psi$, where U is the unitary represented by C.
4. To read out the data after the calculation we perform a von Neumann measurement in the computational basis, i.e. we measure the observable given by the one-dimensional projectors $|x\rangle\langle x|$, $x \in \mathbb{B}^N$. Hence, we get $x \in \mathbb{B}^N$ with probability $P_N = |\langle\psi|x\rangle|^2$.
5. Finally, we have to postprocess the measured value x on a classical computer to end up with the final result x'. If, however, the output state $U\Psi$ is a proper superposition of basis vectors $|x\rangle$ (and not just one $|x\rangle$) the probability p_x to get this particular x' is less than 1. In other words, we have performed a probabilistic calculation as described in the last paragraph of Section 4.5.2. Hence, we have to check the validity of the results (with a class **P** algorithm on a classical computer) and if they are wrong we have to go back to step 2.

So, why is quantum computing potentially useful? First of all, a quantum computer can perform at least as good as a classical computer. This follows immediately from our discussion of reversible computing in Section 4.5.3 and the fact that any invertible function $f : \mathbb{B}^N \to \mathbb{B}^N$ defines a unitary by $U_f : |x\rangle \mapsto |f(x)\rangle$ (the quantum CNOT gate in Fig. 4.8 arises exactly in this way from the classical CNOT). But, there is on the other hand strong evidence which indicates that a quantum computer can solve problems in polynomial time which a classical computer cannot. The most striking example for this fact is the Shor algorithm, which provides a way to solve the factoring problem (which is most probably not in class **P**) in polynomial time. If we introduce the new complexity class **BQP** of decision problems which can be solved with high probability and in polynomial time with a quantum computer, we can express this conjecture as **BPP** ≠ **BQP**.

The mechanism which gives a quantum computer its potential power is the ability to operate not just on one value $x \in \mathbb{B}^N$, but on whole superpositions of values, as already mentioned in step 2 above. E.g. consider a, not necessarily invertible, map $f : \mathbb{B}^N \to \mathbb{B}^M$ and the unitary operator U_f
$$\mathscr{H}^{\otimes N} \otimes \mathscr{H}^{\otimes M} \ni |x\rangle \otimes |0\rangle \mapsto U_f|x\rangle \otimes |0\rangle = |x\rangle \otimes |f(x)\rangle \in \mathscr{H}^{\otimes N} \otimes \mathscr{H}^{\otimes M} \ . \tag{4.34}$$

If we let act U_f on a register in the state $\Psi \otimes |0\rangle$ from Eq. (4.33) we get the result

$$U_f(\Psi \otimes |0\rangle) = \frac{1}{\sqrt{2^N}} \sum_{x \in \mathbb{B}^N} |x\rangle \otimes |f(x)\rangle \ . \tag{4.35}$$

Hence, a quantum computer can evaluate the function f on all possible arguments $x \in \mathbb{B}^N$ at the same time! To benefit from this feature—usually called *quantum parallelism*—is, however, not as easy as it looks like. If we perform a measurement on $U_f(\Psi \otimes |0\rangle)$ in the computational basis we get the value of f for exactly one argument and the rest of the information originally contained in $U_f(\Psi \otimes |0\rangle)$ is destroyed. In other words it is not possible to read out all pairs $(x, f(x))$ from $U_f(\Psi \otimes |0\rangle)$ and to fill a (classical) lookup table with them. To take advantage from quantum parallelism we have to use a clever algorithm within the quantum computation step (step 3 above). In the next section we will consider a particular example for this.

Before we come to this point, let us give some additional comments which link this section to other parts of quantum information. The first point concerns entanglement. The state $U_f(\Psi \otimes |0\rangle)$ is highly entangled (although Ψ is separable since $\Psi = [2^{-1/2}(|0\rangle + |1\rangle)]^{\otimes N}$), and this fact is essential for the "exponential speedup" of computations we could gain in a quantum computer. In other words, to outperform a classical computer, entanglement is the most crucial resource—this will become more transparent in the next section. The second remark concerns error correction. Up to now we have implicitly assumed that all components of a quantum computer work perfectly without any error. In reality, however, decoherence effects make it impossible to realize unitarily implemented operations, and we have to deal with noisy channels. Fortunately, it is possible within quantum information to correct at least a certain amount of errors, as we have seen in Section 4.4. Hence, unlike an analog computer[17] a quantum computer can be designed fault tolerant, i.e. it can work with imperfectly manufactured components.

4.5.5. Simons problem

We will consider now a particular problem (known as Simons problem; cf. [143]) which shows explicitly how a quantum computer can speed up a problem which is hard to solve with a classical computer. It does not fit, however, exactly into the general scheme sketched in the last subsection, because a quantum "oracle" is involved, i.e. a black box which performs an (a priori unknown) unitary transformation on an input state given to it. The term "oracle" indicates here that we are not interested in the time the black box needs to perform the calculation but only in the number of times we have to access it. Hence, this example does not prove the conjecture **BPP** $\neq$ **BQP** stated above. Other quantum algorithms which we do not have the room here to discuss include: the Deutsch [52] and Deutsch–Josza problem [53], the Grover search algorithm [74,75] and of course Shor's factoring algorithm [139,140].

Hence, let us assume that our black box calculates the unitary U_f from Eq. (4.34) with a map $f : \mathbb{B}^N \to \mathbb{B}^N$ which is two to one and has period a, i.e. $f(x) = f(y)$ iff $y = x + a \bmod 2$. The task is to find a. Classically, this problem is hard, i.e. we have to query the oracle exponentially often. To see this note first that we have to find a pair (x, y) with $f(x) = f(y)$ and the probability to get it with two random queries is 2^{-N} (since there is for each x exactly one $y \neq x$ with $f(x) = f(y)$).

[17] If an analog computer works reliably only with a certain accuracy, we can rewrite the algorithm into a digital one.

If we use the box $2^{N/4}$ times, we get less than $2^{N/2}$ different pairs. Hence, the probability to get the correct solution is $2^{-N/2}$, i.e. arbitrarily small even with exponentially many queries.

Assume now that we let our box act on a quantum register $\mathscr{H}^{\otimes N} \otimes \mathscr{H}^{\otimes N}$ in the state $\Psi \otimes |0\rangle$ with Ψ from Eq. (4.33) to get $U_f(\Psi \otimes |0\rangle)$ from (4.35). Now we measure the second register. The outcome is one of 2^{N-1} possible values (say $f(x_0)$), each of which occurs equiprobable. Hence, after the measurement the first register is the state $2^{-1/2}(|x\rangle + |x+a\rangle)$. Now we let a Hadamard gate H (cf. Fig. 4.9) act on each qubit of the first register and the result is (this follows with a short calculation)

$$\frac{1}{\sqrt{2}} H^{\otimes N}(|x\rangle + |x+a\rangle) = \frac{1}{\sqrt{2^{N-1}}} \sum_{a \cdot y = 0} (-1)^{x \cdot y} |y\rangle , \tag{4.36}$$

where the dot denotes the ($\mathbb{B}$-valued) scalar product in the vector space $\mathbb{B}^N$. Now we perform a measurement on the first register (in computational basis) and we get a $y \in \mathbb{B}^N$ with the property $y \cdot a = 0$. If we repeat this procedure N times and if we get N linear-independent values y_j we can determine a as a solution of the system of equations $y_1 \cdot a = 0, \ldots, y_N \cdot a = 0$. The probability to appear as an outcome of the second measurement is for each y with $y \cdot a = 0$ given by 2^{1-N}. Therefore, the success probability can be made arbitrarily big while the number of times we have to access the box is linear in N.

4.6. Quantum cryptography

Finally, we want to have a short look on quantum cryptography—another more practical application of quantum information, which has the potential to emerge into technology in the not so distant future (see e.g. [95,93,34] for some experimental realizations and [69] for a more detailed overview). Hence, let us assume that Alice has a message $x \in \mathbb{B}^N$ which she wants to send secretly to Bob over a public communication channels. One way to do this is the so-called "one-time pad": Alice generates randomly a second bit-string $y \in \mathbb{B}^N$ of the same length as x sends $x + y$ instead of x. Without knowledge of the key y it is completely impossible to recover the message x from $x + y$. Hence, this is a perfectly secure method to transmit secret data. Unfortunately, it is completely useless without a secure way to transmit the key y to Bob, because Bob needs y to decrypt the message $x + y$ (simply by adding y again). What makes the situation even worse is the fact that the key y can be used only once (therefore the name *one-time* pad). If two messages x_1, x_2 are encrypted with the same key we can use x_1 as a key to decrypt x_2 and vice versa: $(x_1 + y) + (x_2 + y) = x_1 + x_2$, hence both messages are partly compromised.

Due to these problems completely different approaches, namely "public key systems" like DSA and RSA are used today for cryptography. The idea is to use two keys instead of one: a *private key* which is used for decryption and only known to its owner and a *public key* used for encryption, which is publicly available (we do not discuss the algorithms needed for key generation, encryption and decryption here, see [145] and the references therein instead). To use this method, Bob generates a key pair (z, y), keeps his private key (y) at a secure place and sends the public one (z) to Alice over a public channel. Alice encrypts her message with z sends the result to Bob and he can decrypt it with y. The security of this scheme relies on the assumption that the factoring problem is computationally *hard*, i.e. not in class **P**, because to calculate y from z requires the factorization of large integers. Since the latter is tractable on quantum computers via Shor's algorithm, the security of public key systems breaks down if quantum computers become available in the future. Another problem of more fundamental nature is the unproven status of the conjecture that factorization is not solvable in polynomial time. Consequently, security of public key systems is not proven either.

The crucial point is now that quantum information provides a way to distribute a cryptographic key y in a secure way, such that y can be used as a one-time pad afterwards. The basic idea is to use the no cloning theorem to detect possible eavesdropping attempts. To make this more transparent, let us consider a particular example here, namely the probably most prominent protocol proposed by Benett and Brassard in 1984 [10].

1. Assume that Alice wants to transmit bits from the (randomly generated) key $y \in \mathbb{B}^N$ through an ideal quantum channel to Bob. Before they start they settle upon two orthonormal bases $e_0, e_1 \in \mathscr{H}$, respectively $f_0, f_1 \in \mathscr{H}$, which are mutually non-orthogonal, i.e. $|\langle e_j, f_k \rangle| \geqslant \varepsilon > 0$ with ε big enough for each $j, k = 0, 1$. If photons are used as information carrier a typical choice are linearly polarized photons with polarization direction rotated by 45° against each other.
2. To send one bit $j \in \mathbb{B}$ Alice selects now at random one of the two bases, say e_0, e_1 and then she sends a qubit in the state $|e_j\rangle\langle e_j|$ through the channel. Note that neither Bob nor a potential eavesdropper knows which bases she has chosen.
3. When Bob receives the qubit he selects, as Alice before, at random a base and performs the corresponding von Neumann measurement to get one classical bit $k \in \mathbb{B}$, which he records together with the measurement method.
4. Both repeat this procedure until the whole string $y \in \mathbb{B}^N$ is transmitted and then Bob tells Alice (through a classical, public communication channel) bit for bit which base he has used for the measurement (but not the result of the measurement). If he has used the same base as Alice both keep the corresponding bit otherwise they discard it. They end up with a bit-string $y' \in \mathbb{B}^M$ of a reduced length M. If this is not sufficient they have to continue sending random bits until the key is long enough. For large N the rate of successfully transmitted bits per bits sended is obviously $\frac{1}{2}$. Hence, Alice has to send approximately twice as many bits as they need.

To see why this procedure is secure, assume now that the eavesdropper Eve can listen and modify the information sent through the quantum channel and that she can listen on the classical channel but cannot modify it (we come back to this restriction in a minute). Hence, Eve can intercept the qubits sent by Alice and make two copies of it. One she forwards to Bob and the other she keeps for later analysis. Due to the no cloning theorem, however, she has produced errors in both copies and the quality of her own decreases if she tries to make the error in Bob's as small as possible. Even if Eve knows about the two bases e_0, e_1 and f_0, f_1 she does not know which one Alice uses to send a particular qubit[18]. Hence, Eve has to decide randomly which base to choose (as Bob). If e_0, e_1 and f_0, f_1 are chosen optimal, i.e. $|\langle e_j, f_k \rangle|^2 = 0.5$ it is easy to see that the error rate Eve necessarily produces if she randomly measures in one of the bases is 1/4 for large N. To detect this error Alice and Bob simply have to sacrify portions of the generated key and to compare randomly selected bits using their classical channel. If the error rate they detect is too big they can decide to drop the whole key and restart from the beginning.

[18] If Alice and Bob uses only one basis to send the data and Eve knows about it she can produce, of course, ideal copies of the qubits. This is actually the reason why two non-orthogonal bases are necessary.

So let us discuss finally a situation where Eve is able to intercept the quantum *and* the classical channel. This would imply that she can play Bob's part for Alice and Alice's for Bob. As a result she shares a key with Alice and one with Bob. Hence, she can decode all secret data Alice sends to Bob, read it, and encode it finally again to forward it to Bob. To secure against such a "woman in the middle attack", Alice and Bob can use classical authentication protocols which ensure that the correct person is at the other end of the line. This implies that they need a small amount of initial secret material which can be renewed, however, from the new key they have generated through quantum communication.

5. Entanglement measures

In the last section we have seen that entanglement is an essential *resource* for many tasks of quantum information theory, like teleportation or quantum computation. This means that entangled states are needed for the functioning of many processes and that they are consumed during operation. It is therefore necessary to have *measures* which tell us whether the entanglement contained in a number of quantum systems is sufficient to perform a certain task. What makes this subject difficult is the fact that we cannot restrict the discussion to systems in a maximally or at least highly entangled pure state. Due to unavoidable decoherence effects realistic applications have to deal with imperfect systems in mixed states, and exactly in this situation the question for the amount of available entanglement is interesting.

5.1. General properties and definitions

The difficulties arising if we try to quantify entanglement can be divided, roughly speaking, into two parts: Firstly, we have to find a reasonable quantity which describes exactly those properties which we are interested in and secondly we have to calculate it for a given state. In this section we will discuss the first problem and consider several different possibilities to define entanglement measures.

5.1.1. Axiomatics

First of all, we will collect some general properties which a reasonable entanglement measure should have (cf. also [16,154,153,155,89]). To quantify entanglement, means nothing else but to associate a positive real number to each state of (finite dimensional) two-partite systems.

Axiom E0. *An entanglement measure is a function E which assigns to each state ρ of a finite-dimensional bipartite system a positive real number $E(\rho) \in \mathbb{R}^+$.*

Note that we have glanced over some mathematical subtleties here, because E is not just defined on the state space of $\mathscr{B}(\mathscr{H} \otimes \mathscr{K})$ systems for particularly chosen Hilbert spaces $\mathscr{H}$ and $\mathscr{K} - E$ is defined on any state space for arbitrary finite dimensional $\mathscr{H}$ and $\mathscr{K}$. This is expressed mathematically most conveniently by a *family of functions* which behaves naturally under restrictions (i.e. the restriction to a subspace $\mathscr{H}' \otimes \mathscr{K}'$ coincides with the function belonging to $\mathscr{H}' \otimes \mathscr{K}'$). However, we will see soon that we can safely ignore this problem.

The next point concerns the range of E. If ρ is unentangled $E(\rho)$ should be zero of course and it should be maximal on maximally entangled states. But what happens if we allow the dimensions of $\mathscr{H}$ and $\mathscr{K}$ to grow? To get an answer consider first a pair of qubits in a maximally entangled state ρ. It should contain exactly one-bit entanglement, i.e. $E(\rho) = 1$ and N pairs in the state $\rho^{\otimes N}$ should contain N bits. If we interpret $\rho^{\otimes N}$ as a maximally entangled state of a $\mathscr{H} \otimes \mathscr{H}$ system with $\mathscr{H} = \mathbb{C}^N$ we get $E(\rho^{\otimes N}) = \log_2(\dim(\mathscr{H})) = N$, where we have to reshuffle in $\rho^{\otimes N}$ the tensor factors such that $(\mathbb{C}^2 \otimes \mathbb{C}^2)^{\otimes N}$ becomes $(\mathbb{C}^2)^{\otimes N} \otimes (\mathbb{C}^2)^{\otimes N}$ (i.e. "all Alice particles to the left and all Bob particles to the right"; cf. Section 4.3.) This observation motivates the following.

Axiom E1 (Normalization). *E vanishes on separable and takes its maximum on maximally entangled states. More precisely, this means that $E(\sigma) \leqslant E(\rho) = \log_2(d)$ for $\rho, \sigma \in \mathscr{S}(\mathscr{H} \otimes \mathscr{H})$ and ρ maximally entangled.*

One thing an entanglement measure should tell us, is how much quantum information can be *maximally* teleported with a certain amount of entanglement, where this maximum is taken over all possible teleportation schemes and distillation protocols, hence it cannot be increased further by additional LOCC operations on the entangled systems in question. This consideration motivates the following Axiom.

Axiom E2 (LOCC monotonicity). *E cannot increase under LOCC operation, i.e. $E[T(\rho)] \leqslant E(\rho)$ for all states ρ and all LOCC channels T.*

A special case of LOCC operations are, of course, local unitary operations $U \otimes V$. Axiom E2 implies now that $E(U \otimes V\rho U^* \otimes V^*) \leqslant E(\rho)$ and on the other hand $E(U^* \otimes V^* \tilde{\rho} U \otimes V) \leqslant E(\tilde{\rho})$ hence with $\tilde{\rho} = U \otimes V\rho U^* \otimes V$ we get $E(\rho) \leqslant E(U \otimes V\rho V^* \otimes U^*)$ therefore $E(\rho) = E(U \otimes V\rho U^* \otimes V^*)$. We fix this property as a weakened version of Axiom E2.

Axiom E2a (Local unitary invariance). *E is invariant under local unitaries, i.e. $E(U \otimes V\rho U^* \otimes V^*) = E(\rho)$ for all states ρ and all unitaries U, V.*

This axiom shows why we do not have to bother about families of functions as mentioned above. If E is defined on $\mathscr{S}(\mathscr{H} \otimes \mathscr{H})$ it is automatically defined on $\mathscr{S}(\mathscr{H}_1 \otimes \mathscr{H}_2)$ for all Hilbert spaces $\mathscr{H}_k$ with $\dim(\mathscr{H}_k) \leqslant \dim(\mathscr{H})$, because we can embed $\mathscr{H}_1 \otimes \mathscr{H}_2$ under this condition unitarily into $\mathscr{H} \otimes \mathscr{H}$.

Consider now a convex linear combination $\lambda\rho + (1 - \lambda)\sigma$ with $0 \leqslant \lambda \leqslant 1$. Entanglement cannot be "generated" by mixing two states, i.e. $E(\lambda\rho + (1 - \lambda)\sigma) \leqslant \lambda E(\rho) + (1 - \lambda)E(\sigma)$.

Axiom E3 (Convexity). *E is a convex function, i.e. $E(\lambda\rho + (1 - \lambda)\sigma) \leqslant \lambda E(\rho) + (1 - \lambda)E(\sigma)$ for two states ρ, σ and $0 \leqslant \lambda \leqslant 1$.*

The next property concerns the continuity of E, i.e. if we perturb ρ slightly the change of $E(\rho)$ should be small. This can be expressed most conveniently as continuity of E in the trace norm. At this point, however, it is not quite clear, how we have to handle the fact that E is defined for arbitrary Hilbert spaces. The following version is motivated basically by the fact that it is a crucial assumption in Theorems 5.2 and 5.3.

Axiom E4 (Continuity). *Consider a sequence of Hilbert spaces $\mathscr{H}_N$, $N \in \mathbb{N}$ and two sequences of states $\rho_N, \sigma_N \in \mathscr{S}(\mathscr{H}_N \otimes \mathscr{H}_N)$ with $\lim \|\rho_N - \sigma_N\|_1 = 0$. Then we have*

$$\lim_{N\to\infty} \frac{E(\rho_N) - E(\sigma_N)}{1 + \log_2(\dim \mathscr{H}_N)} = 0 \ . \tag{5.1}$$

The last point we have to consider here are additivity properties: Since we are looking at entanglement as a resource, it is natural to assume that we can do with two pairs in the state ρ twice as much as with one ρ, or more precisely $E(\rho \otimes \rho) = 2E(\rho)$ (in $\rho \otimes \rho$ we have to reshuffle tensor factors again; see above).

Axiom E5 (Additivity). *For any pair of two-partite states $\rho, \sigma \in \mathscr{S}(\mathscr{H} \otimes \mathscr{K})$ we have $E(\sigma \otimes \rho) = E(\sigma) + E(\rho)$.*

Unfortunately, this rather natural looking axiom seems to be too strong (it excludes reasonable candidates). It should be however, always true that entanglement cannot increase if we put two pairs together.

Axiom E5a (Subadditivity). *For any pair of states ρ, σ we have $E(\rho \otimes \sigma) \leqslant E(\rho) + E(\sigma)$.*

There are further modifications of additivity available in the literature. Most frequently used is the following, which restricts Axiom E5 to the case $\rho = \sigma$.

Axiom E5b (Weak additivity). *For any state ρ of a bipartite system we have $N^{-1}E(\rho^{\otimes N}) = E(\rho)$.*

Finally, the weakest version of additivity only deals with the behavior of E for large tensor products, i.e. $\rho^{\otimes N}$ for $N \to \infty$.

Axiom E5c (Existence of a regularization). *For each state ρ the limit*

$$E^{\infty}(\rho) = \lim_{N\to\infty} \frac{E(\rho^{\otimes N})}{N} \tag{5.2}$$

exists.

5.1.2. Pure states

Let us consider now a pure state $\rho = |\psi\rangle\langle\psi| \in \mathscr{S}(\mathscr{H} \otimes \mathscr{K})$. If it is entangled its partial trace $\sigma = \mathrm{tr}_{\mathscr{H}} |\psi\rangle\langle\psi| = \mathrm{tr}_{\mathscr{K}} |\psi\rangle\langle\psi|$ is mixed and for a maximally entangled state it is maximally mixed. This suggests to use the von Neumann entropy[19] of ρ, which measures how much a state is mixed, as an entanglement measure for pure states, i.e. we define [9,16]

$$E_{\mathrm{vN}}(\rho) = -\mathrm{tr}[\mathrm{tr}_{\mathscr{H}} \rho \ln(\mathrm{tr}_{\mathscr{H}} \rho)] \ . \tag{5.3}$$

[19] We assume here and in the following that the reader is sufficiently familiar with entropies. If this is not the case we refer to [123].

It is easy to deduce from the properties of the von Neumann entropy that E_{vN} satisfies Axioms E0, E1, E3 and E5b. Somewhat more difficult is only Axiom E2 which follows, however, from a nice theorem of Nielsen [119] which relates LOCC operations (on pure states) to the theory of majorization. To state it here we need first some terminology. Consider two probability distributions $\lambda = (\lambda_1, \ldots, \lambda_M)$ and $\mu = (\mu_1, \ldots, \mu_N)$ both given in decreasing order (i.e. $\lambda_1 \geqslant \cdots \geqslant \lambda_M$ and $\mu_1 \geqslant \cdots \geqslant \mu_N$). We say that λ is *majorized* by μ, in symbols $\lambda \prec \mu$, if

$$\sum_{j=1}^{k} \lambda_j \leqslant \sum_{j=1}^{k} \mu_j \quad \forall k = 1, \ldots, \min M, N \tag{5.4}$$

holds. Now we have the following result (see [119] for a proof).

Theorem 5.1. *A pure state $\psi = \sum_j \lambda_j^{1/2} e_j \otimes e_j' \in \mathscr{H} \otimes \mathscr{K}$ can be transformed into another pure state $\phi = \sum_j \mu_j^{1/2} f_j \otimes f_j' \in \mathscr{H} \otimes \mathscr{K}$ via an LOCC operation, iff the Schmidt coefficients of ψ are majorized by those of ϕ, i.e. $\lambda \prec \mu$.*

The von Neumann entropy of the restriction $\mathrm{tr}_{\mathscr{K}} |\psi\rangle\langle\psi|$ can be immediately calculated from the Schmidt coefficients λ of ψ by $E_{\mathrm{vN}}(|\psi\rangle\langle\psi|) = -\sum_j \lambda_j \ln(\lambda_j)$. Axiom E2 follows therefore from the fact that the entropy $S(\lambda) = -\sum_j \lambda_j \ln(\lambda_j)$ of a probability distribution λ is a *Shur concave* function, i.e. $\lambda \prec \mu$ implies $S(\lambda) \geqslant S(\mu)$; see [121].

Hence, we have seen so far that E_{vN} is one possible candidate for an entanglement measure on pure states. In the following we will see that it is in fact the only candidate which is physically reasonable. There are basically two reasons for this. The first one deals with distillation of entanglement. It was shown by Bennett et al. [9] that each state $\psi \in \mathscr{H} \otimes \mathscr{K}$ of a bipartite system can be prepared out of (a possibly large number of) systems in an arbitrary entangled state ϕ by LOCC operations. To be more precise, we can find a sequence of LOCC operations

$$T_N : \mathscr{B}[(\mathscr{H} \otimes \mathscr{K})^{\otimes M(N)}] \to \mathscr{B}[(\mathscr{H} \otimes \mathscr{K})^{\otimes N}] \tag{5.5}$$

such that

$$\lim_{N \to \infty} \| T_N^*(|\phi\rangle\langle\phi|^{\otimes N}) - |\psi\rangle\langle\psi| \|_1 = 0 \tag{5.6}$$

holds with a non-vanishing *rate* $r = \lim_{N\to\infty} M(N)/N$. This is done either by distillation ($r < 1$ if ψ is higher entangled then ϕ) or by "diluting" entanglement, i.e. creating many less entangled states from few highly entangled ones ($r > 1$). All this can be performed in a *reversible* way: We can start with some maximally entangled qubits, dilute them to get many less entangled states which can be distilled afterwards to get the original states back (again only in an asymptotic sense). The crucial point is that the asymptotic rate r of these processes is given in terms of E_{vN} by $r = E_{\mathrm{vN}}(|\phi\rangle\langle\phi|)/E_{\mathrm{vN}}(|\psi\rangle\langle\psi|)$. Hence, we can say, roughly speaking, that $E_{\mathrm{vN}}(|\psi\rangle\langle\psi|)$ describes exactly the amount of maximally entangled qubits which is contained in $|\psi\rangle\langle\psi|$.

A second somewhat more formal reason is that E_{vN} is the only entanglement measure on the set of pure states which satisfies the axioms formulated above. In other words the following "*uniqueness theorem for entanglement measures*" holds [129,155,57].

Theorem 5.2. *The reduced von Neumann entropy E_{vN} is the only entanglement measure on pure states which satisfies Axioms* E0–E5.

5.1.3. Entanglement measures for mixed states

To find reasonable entanglement measures for mixed states is much more difficult. There are in fact many possibilities (e.g. the maximally entangled fraction introduced in Section 3.1.1 can be regarded as a simple measure) and we want to present therefore only four of the most reasonable candidates. Among those measures which we do not discuss here are negativity quantities ([158] and the references therein) the "best separable approximation" [108], the base norm associated with the set of separable states [157,136] and ppt-distillation rates [133].

The first measure we want to present is oriented along the discussion of pure states: We define, roughly speaking, the asymptotic rate with which maximally entangled qubits can be distilled at most out of a state $\rho \in \mathcal{S}(\mathcal{H} \otimes \mathcal{K})$ as the *entanglement of distillation* $E_{\mathrm{D}}(\rho)$ of ρ; cf. [12]. To be more precise consider all possible distillation protocols for ρ (cf. Section 4.3), i.e. all sequences of LOCC channels

$$T_N : \mathcal{B}(\mathbb{C}^{d_N} \otimes \mathbb{C}^{d_N}) \to \mathcal{B}(\mathcal{H}^{\otimes N} \otimes \mathcal{K}^{\otimes N}) , \tag{5.7}$$

such that

$$\lim_{N\to\infty} \| T_N^*(\rho^{\otimes N}) - |\Omega_N\rangle\langle\Omega_N| \, \|_1 = 0 \tag{5.8}$$

holds with a sequence of maximally entangled states $\Omega_N \in \mathbb{C}^{d_N}$. Now we can define

$$E_{\mathrm{D}}(\rho) = \sup_{(T_N)_{N\in\mathbb{N}}} \limsup_{N\to\infty} \frac{\log_2(d_N)}{N} , \tag{5.9}$$

where the supremum is taken over all possible distillation protocols $(T_N)_{N\in\mathbb{N}}$. It is not very difficult to see that E_{D} satisfies Axioms E0, E1, E2 and E5b. It is not known whether continuity (Axiom E4) and convexity (Axiom E3) holds. It can be shown, however, that E_{D} is not convex (and not additive; Axiom E5) if npt bound entangled states exist (see [141], cf. also Section 4.3.3).

For pure states we have discussed beside distillation the "dilution" of entanglement and we can use, similar to E_{D}, the asymptotic rate with which bipartite systems in a given state ρ can be prepared out of maximally entangled singlets [78]. Hence, consider again a sequence of LOCC channels

$$T_N : \mathcal{B}(\mathcal{H}^{\otimes N} \otimes \mathcal{K}^{\otimes N}) \to \mathcal{B}(\mathbb{C}^{d_N} \otimes \mathbb{C}^{d_N}) \tag{5.10}$$

and a sequence of maximally entangled states $\Omega_N \in \mathbb{C}^{d_N}$, $N \in \mathbb{N}$, but now with the property

$$\lim_{N\to\infty} \| \rho^{\otimes N} - T_N^*(|\Omega_N\rangle\langle\Omega_N|) \, \|_1 = 0 . \tag{5.11}$$

Then we can define the *entanglement cost* $E_{\mathrm{C}}(\rho)$ of ρ as

$$E_{\mathrm{C}}(\rho) = \inf_{(S_N)_{N\in\mathbb{N}}} \liminf_{N\to\infty} \frac{\log_2(d_N)}{N} , \tag{5.12}$$

where the infimum is taken over all dilution protocols S_N, $N \in \mathbb{N}$. It is again easy to see that E_{C} satisfies Axioms E0, E1, E2 and E5b. In contrast to E_{D} however it can be shown that E_{C} is convex (Axiom E3), while it is not known, whether E_{C} is continuous (Axiom E4); cf [78] for proofs.

E_D and E_C are directly based on operational concepts. The remaining two measures we want to discuss here are defined in a more abstract way. The first can be characterized as the minimal convex extension of E_{vN} to mixed states: We define the *entanglement of formation* E_F of ρ as [16]

$$E_F(\rho) = \inf_{\rho=\sum_j p_j|\psi_j\rangle\langle\psi_j|} \sum p_j E_{vN}(|\psi_j\rangle\langle\psi_j|) , \tag{5.13}$$

where the infimum is taken over all decompositions of ρ into a convex sum of pure states. E_F satisfies E0–E4 and E5a (cf. [16] for E2 and [120] for E4 the rest follows directly from the definition). Whether E_F is (weakly) additive (Axiom E5b) is not known. Furthermore, it is conjectured that E_F coincides with E_C. However, proven is only the identity $E_F^\infty = E_C$, where the existence of the regularization E_F^∞ of E_F follows directly from subadditivity.

Another idea to quantify entanglement is to measure the "distance" of the (entangled) ρ from the set of separable states $\mathscr{D}$. It hat turned out [154] that among all possible distance functions the relative entropy is physically most reasonable. Hence, we define the *relative entropy of entanglement* as

$$E_R(\rho) = \inf_{\sigma\in\mathscr{D}} S(\rho|\sigma), \quad S(\rho|\sigma) = [\mathrm{tr}(\rho \log_2 \rho - \rho \log_2 \sigma)] , \tag{5.14}$$

where the infimum is taken over all separable states. It can be shown that E_R satisfies, as E_F the Axioms E0–E4 and E5a, where E1 and E2 are shown in [154] and E4 in [56]; the rest follows directly from the definition. It is shown in [159] that E_R does not satisfy E5b; cf. also Section 5.3. Hence, the regularization E_R^∞ of E_R differs from E_R.

Finally, let us give now some comments on the relation between the measures just introduced. On pure states all measures just discussed, coincide with the reduced von Neumann entropy—this follows from Theorem 5.2 and the properties stated in the last subsection. For mixed states the situation is more difficult. It can be shown however that $E_D \leqslant E_C$ holds and that all "reasonable" entanglement measures lie in between [89].

Theorem 5.3. *For each entanglement measure E satisfying* E0, E1, E2 *and* E5b *and each state* $\rho \in \mathscr{S}(\mathscr{H} \otimes \mathscr{K})$ *we have* $E_D(\rho) \leqslant E(\rho) \leqslant E_C(\rho)$.

Unfortunately, no measure we have discussed in the last subsection satisfies all the assumptions of the theorem. It is possible, however, to get a similar statement for the regularization E^∞ with weaker assumptions on E itself (in particular, without assuming additivity); cf. [57].

5.2. Two qubits

Even more difficult than finding reasonable entanglement measures are explicit calculations. All measures we have discussed above involve optimization processes over spaces which grow exponentially with the dimension of the Hilbert space. A direct numerical calculation for a general state ρ is therefore hopeless. There are, however, some attempts to get either some bounds on entanglement measures or to get explicit calculations for special classes of states. We will concentrate this discussion to some relevant special cases. On the one hand, we will concentrate on E_F and E_R and on the other we will look at two special classes of states where explicit calculations are possible: Two qubit systems in this section and states with symmetry properties in the next one are given.

5.2.1. Pure states

Assume for the rest of this section that $\mathscr{H} = \mathbb{C}^2$ holds and consider first a pure state $\psi \in \mathscr{H} \otimes \mathscr{H}$. To calculate $E_{\mathrm{vN}}(\psi)$ is of course not difficult and it is straightforward to see that (cf. [16] for all material of this and the following subsection)

$$E_{\mathrm{vN}}(\psi) = H[\tfrac{1}{2}(1 + \sqrt{1 - C(\psi)^2})] \tag{5.15}$$

holds, with

$$H(x) = -x \log_2(x) - (1 - x)\log_2(1 - x) \tag{5.16}$$

and the *concurrence* $C(\psi)$ of ψ which is defined by

$$C(\psi) = \left|\sum_{j=0}^{3} \alpha_j^2\right| \quad \text{with } \psi = \sum_{j=0}^{3} \alpha_j \Phi_j\ , \tag{5.17}$$

where Φ_j, $j = 0, \ldots, 3$ denotes the Bell basis (3.3). Since C becomes rather important in the following let us reexpress it as $C(\psi) = |\langle \psi, \Xi\psi\rangle|$, where $\psi \mapsto \Xi\psi$ denotes complex conjugation in the Bell basis. Hence, Ξ is an antiunitary operator and it can be written as the tensor product $\Xi = \xi \otimes \xi$ of the map $\mathscr{H} \ni \phi \mapsto \sigma_2\bar{\phi}$, where $\bar{\phi}$ denotes complex conjugation in the canonical basis and σ_2 is the second Pauli matrix. Hence, local unitaries (i.e. those of the form $U_1 \otimes U_2$) commute with Ξ and it can be shown that this is not only a necessary but also a sufficient condition for a unitary to be local [160].

We see from Eqs. (5.15) and (5.17) that $C(\psi)$ ranges from 0 to 1 and that $E_{\mathrm{vN}}(\psi)$ is a monotone function in $C(\psi)$. The latter can be considered therefore as an entanglement quantity in its own right. For a Bell state we get in particular $C(\Phi_j) = 1$ while a separable state $\phi_1 \otimes \phi_2$ leads to $C(\phi_1 \otimes \phi_2) = 0$; this can be seen easily with the factorization $\Xi = \xi \otimes \xi$.

Assume now that one of the α_j say α_0 satisfies $|\alpha_0|^2 > 1/2$. This implies that $C(\psi)$ cannot be zero since

$$\left|\sum_{j=1}^{3} \alpha_j^2\right| \leqslant 1 - |\alpha_0|^2 \tag{5.18}$$

must hold. Hence, $C(\psi)$ is at least $1 - 2|\alpha_0|^2$ and this implies for E_{vN} *and arbitrary* ψ

$$E_{\mathrm{vN}}(\psi) \geqslant h(|\langle \Phi_0, \psi\rangle|^2) \quad \text{with } h(x) = \begin{cases} H[\frac{1}{2} + \sqrt{x(1-x)}] & x \geqslant \frac{1}{2}\ , \\ 0 & x < \frac{1}{2}\ . \end{cases} \tag{5.19}$$

This inequality remains valid if we replace Φ_0 by any other maximally entangled state $\Phi \in \mathscr{H} \otimes \mathscr{H}$. To see this note that two maximally entangled states $\Phi, \Phi' \in \mathscr{H} \otimes \mathscr{H}$ are related (up to a phase) by a local unitary transformation $U_1 \otimes U_2$ (this follows immediately from their Schmidt decomposition; cf Section 3.1.1). Hence, if we replace the Bell basis in Eq. (5.17) by $\Phi_j' = U_1 \otimes U_2 \Phi_j$, $j = 0, \ldots, 3$ we get for the corresponding C' the equation $C'(\psi) = \langle U_1^* \otimes U_2^* \psi, \Xi U_1^* \otimes U_2^* \psi\rangle = C(\psi)$ since Ξ commutes with local unitaries. We can even replace $|\langle \Phi_0, \psi\rangle|^2$ with the supremum over all maximally entangled states and therefore get

$$E_{\mathrm{vN}}(\psi) \geqslant h[\mathscr{F}(|\psi\rangle\langle\psi|)]\ , \tag{5.20}$$

where $\mathscr{F}(|\psi\rangle\langle\psi|)$ is the maximally entangled fraction of $|\psi\rangle\langle\psi|$ which we have introduced in Section 3.1.1.

To see that even equality holds in Eq. (5.20) note first that it is sufficient to consider the case $\psi = a|00\rangle + b|11\rangle$ with $a, b \geqslant 0$, $a^2 + b^2 = 1$, since each pure state ψ can be brought into this form (this follows again from the Schmidt decomposition) by a local unitary transformation which on the other hand does not change E_{vN}. The maximally entangled state which maximizes $|\langle\psi, \Phi\rangle|^2$ is in this case Φ_0 and we get $\mathscr{F}(|\psi\rangle\langle\psi|) = (a+b)^2/2 = 1/2 + ab$. Straightforward calculations now show that $h[\mathscr{F}(|\psi\rangle\langle\psi|)] = h(1/2 + ab) = E_{\mathrm{vN}}(\psi)$ holds as stated.

5.2.2. EOF for Bell diagonal states

It is easy to extend inequality (5.20) to mixed states if we use the convexity of E_{F} and the fact that E_{F} coincides with E_{vN} on pure states. Hence, (5.20) becomes

$$E_{\mathrm{F}}(\rho) \geqslant h[\mathscr{F}(\rho)] . \tag{5.21}$$

For general two-qubit states this bound is not achieved however. This can be seen with the example $\rho = 1/2(|\phi_1\rangle\langle\phi_1| + |00\rangle\langle 00|)$, which we have already considered in the last paragraph of Section 3.1.1. It is easy to see that $\mathscr{F}(\rho) = \frac{1}{2}$ holds hence $h[\mathscr{F}(\rho)] = 0$ but ρ is entangled. Nevertheless, we can show that equality holds in Eq. (5.21) if we restrict it to the Bell diagonal states $\rho = \sum_{j=0}^{3} \lambda j |\Phi_j\rangle\langle\Phi_j|$. To prove this statement we have to find a convex decomposition $\rho = \sum_j \mu_j |\Psi_j\rangle\langle\Psi_j|$ of such a ρ into pure states $|\Psi_j\rangle\langle\Psi_j|$ such that $h[\mathscr{F}(\rho)] = \sum_j \mu_j E_{\mathrm{vN}}(|\Psi_j\rangle\langle\Psi_j|$ holds. Since $E_{\mathrm{F}}(\rho)$ cannot be smaller than $h[\mathscr{F}(\rho)]$ due to inequality (5.21) this decomposition must be optimal and equality is proven.

To find such Ψ_j assume first that the biggest eigenvalue of ρ is greater than 1/2, and let, without loss of generality, λ_1 be this eigenvalue. A good choice for the Ψ_j are then the eight pure states

$$\sqrt{\lambda_0}\Phi_0 + \mathrm{i}\left(\sum_{j=1}^{3}(\pm\sqrt{\lambda_j})\Phi_j\right) . \tag{5.22}$$

The reduced von Neumann entropy of all these states equals $h(\lambda_1)$, hence $\sum_j \mu_j E_{\mathrm{vN}}(|\Psi_j\rangle\langle\Psi_j|) = h(\lambda_1)$ and therefore $E_{\mathrm{F}}(\rho) = h(\lambda_1)$. Since the maximally entangled fraction of ρ is obviously λ_1 we see that (5.21) holds with equality.

Assume now that the highest eigenvalue is less than 1/2. Then we can find phase factors $\exp(\mathrm{i}\phi_j)$ such that $\sum_{j=0}^{3} \exp(\mathrm{i}\phi_j)\lambda_j = 0$ holds and ρ can be expressed as a convex linear combination of the states

$$\mathrm{e}^{\mathrm{i}\phi_0/2}\sqrt{\lambda_0}\Phi_0 + \mathrm{i}\left(\sum_{j=1}^{3}(\pm\mathrm{e}^{\mathrm{i}\phi_j/2}\sqrt{\lambda_j})\Phi_j\right) . \tag{5.23}$$

The concurrence C of all these states is 0 hence their entanglement is 0 by Eq. (5.15), which in turn implies $E_{\mathrm{F}}(\rho) = 0$. Again, we see that equality is achieved in (5.21) since the maximally entangled fraction of ρ is less than 1/2. Summarizing this discussion we have shown (cf. Fig. 5.1)

Proposition 5.4. *A Bell diagonal state ρ is entangled iff its highest eigenvalue λ is greater than* 1/2. *In this case the entanglement of formation of ρ is given by*

$$E_{\mathrm{F}}(\rho) = H[\tfrac{1}{2} + \sqrt{\lambda(1-\lambda)}] . \tag{5.24}$$

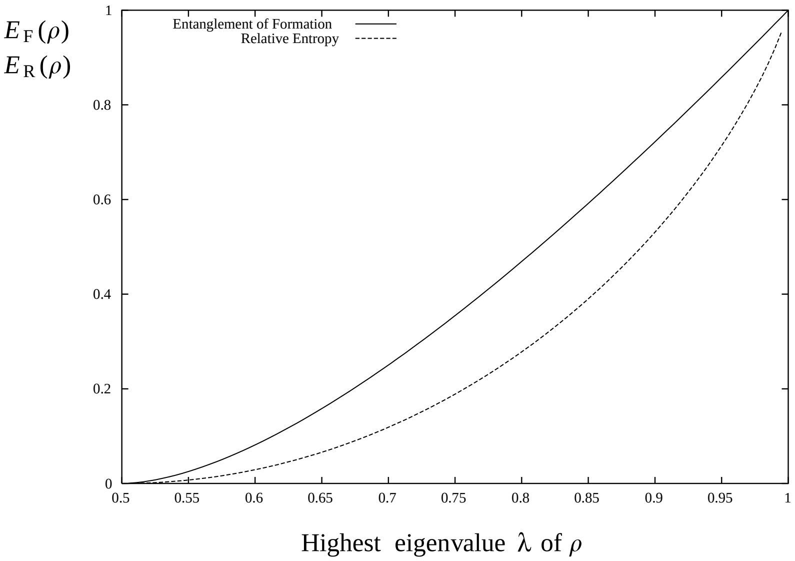


Fig. 5.1. Entanglement of formation and relative entropy of entanglement for the Bell diagonal states, plotted as a function of the highest eigenvalue λ of ρ.

5.2.3. Wootters formula

If we have a general two-qubit state ρ there is a formula of Wootters [172] which allows an easy calculation of E_F. It is based on a generalization of the concurrence C to mixed states. To motivate it rewrite $C^2(\psi) = |\langle\psi, \Xi\psi\rangle|$ as

$$C^2(\psi) = \operatorname{tr}(|\psi\rangle\langle\psi||\Xi\psi\rangle\langle\Xi\psi|) = \operatorname{tr}(\rho\Xi\rho\Xi) = \operatorname{tr}(R^2) \tag{5.25}$$

with

$$R = \sqrt{\sqrt{\rho}\Xi\rho\Xi\sqrt{\rho}} \; . \tag{5.26}$$

Here we have set $\rho = |\psi\rangle\langle\psi|$. The definition of the Hermitian matrix R however makes sense for arbitrary ρ as well. If we write λ_j, $j = 1, \dots, 4$ for the eigenvalues of R and λ_1 is without loss of generality, the biggest one we can define the *concurrence* of an arbitrary two-qubit state ρ as [172]

$$C(\rho) = \max(0, 2\lambda_1 - \operatorname{tr}(R)) = \max(0, \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4) \; . \tag{5.27}$$

It is easy to see that $C(|\psi\rangle\langle\psi|)$ coincides with $C(\psi)$ from (5.17). The crucial point is now that Eq. (5.15) holds for $E_F(\rho)$ if we insert $C(\rho)$ instead of $C(\psi)$:

Theorem 5.5 (Wootters formula). *The entanglement of formation of a two-qubit system in a state ρ is given by*

$$E_F(\rho) = H[\tfrac{1}{2}(1 + \sqrt{1 - C(\rho)^2})] \; , \tag{5.28}$$

where the concurrence of ρ is given in Eq. (5.27) *and H denotes the binary entropy from* (5.16).

To prove this theorem we firstly have to find a convex decomposition $\rho = \sum_j \mu_j |\Psi_j\rangle\langle\Psi_j|$ of ρ into pure states Ψ_j such that the average reduced von Neumann entropy $\sum_j \mu_j E_{\mathrm{vN}}(\Psi_j)$ coincides with the right-hand side of Eq. (5.28). Secondly, we have to show that we have really found the minimal decomposition. Since this is much more involved than the simple case discussed in Section 5.2.2 we omit the proof and refer to [172] instead. Note however that Eq. (5.28) really coincides with the special cases we have derived for the pure and the Bell diagonal states. Finally, let us add the remark that there is no analog of Wootters' formula for higher dimensional Hilbert spaces. It can be shown [160] that the essential properties of the Bell basis Φ_j, $j = 0, \ldots, 3$ which would be necessary for such a generalization are available only in 2 × 2 dimensions.

5.2.4. Relative entropy for Bell diagonal states

To calculate the relative entropy of entanglement E_{R} for two-qubit systems is more difficult. However, there is at least an easy formula for the Bell diagonal states which we will give in the following [154]:

Proposition 5.6. *The relative entropy of entanglement for a Bell diagonal state ρ with highest eigenvalue λ is given by* (*cf. Fig. 5.1*)

$$E_{\mathrm{R}}(\rho) = \begin{cases} 1 - H(\lambda), & \lambda > \frac{1}{2}\ , \\ 0, & \lambda \leqslant \frac{1}{2}\ . \end{cases} \tag{5.29}$$

Proof. For a Bell diagonal state $\rho = \sum_{j=0}^{3} \lambda_j |\Phi_j\rangle\langle\Phi_j|$ we have to calculate

$$E_{\mathrm{R}}(\rho) = \inf_{\sigma \in \mathscr{D}} [\mathrm{tr}(\rho \log_2 \rho - \rho \log_2 \sigma)] \tag{5.30}$$

$$= \mathrm{tr}(\rho \log_2 \rho) + \inf_{\sigma \in \mathscr{D}} \left[-\sum_{j=0}^{3} \lambda_j \langle \Phi_j, \log_2(\sigma) \Phi_j \rangle \right] . \tag{5.31}$$

Since log is a concave function we have $-\log_2\langle\Phi_j, \sigma\Phi_j\rangle \leqslant \langle\Phi_j, -\log_2(\sigma)\Phi_j\rangle$ and therefore

$$E_{\mathrm{R}}(\rho) \geqslant \mathrm{tr}(\rho \log_2 \rho) + \inf_{\sigma \in \mathscr{D}} \left[-\sum_{j=0}^{3} \lambda_j \log_2 \langle \Phi_j, \sigma \Phi_j \rangle \right] . \tag{5.32}$$

Hence, only the diagonal elements of σ in the Bell basis enter the minimization on the right-hand side of this inequality and this implies that we can restrict the infimum to the set of separable Bell diagonal state. Since a Bell diagonal state is separable iff all its eigenvalues are less than 1/2 (Proposition 5.2.1) we get

$$E_{\mathrm{R}}(\rho) \geqslant \mathrm{tr}(\rho \log_2 \rho) + \inf_{p_j \in [0,1/2]} \left[-\sum_{j=0}^{3} \lambda_j \log_2 p_j \right] \quad \text{with} \sum_{j=0}^{3} p_j = 1\ . \tag{5.33}$$

This is an optimization problem (with constraints) over only four real parameters and easy to solve. If the highest eigenvalue of ρ is greater than 1/2 we get $p_1 = 1/2$ and $p_j = \lambda_j/(2 - 2\lambda)$, where we have chosen without loss of generality $\lambda = \lambda_1$. We get a lower bound on $E_{\mathrm{R}}(\rho)$ which is achieved if we insert the corresponding σ in Eq. (5.31). Hence, we have proven the statement for $\lambda > 1/2$. which completes the proof, since we have already seen that $\lambda \leqslant 1/2$ implies that ρ is separable (Proposition 5.4). □

5.3. Entanglement measures under symmetry

The problems occurring if we try to calculate quantities like E_{R} or E_{F} for general density matrices arise from the fact that we have to solve optimization problems over very high dimensional spaces. One possible strategy to get explicit results is therefore parameter reduction by symmetry arguments. This can be done if the state in question admits some invariance properties like Werner, isotropic or OO-invariant states; cf. Section 3.1. In the following, we will give some particular examples for such calculations, while a detailed discussion of the general idea (together with much more examples and further references) can be found in [159].

5.3.1. Entanglement of formation

Consider a compact group of unitaries $G \subset \mathscr{B}(\mathscr{H} \otimes \mathscr{H})$ (where $\mathscr{H}$ is again arbitrary finite dimensional), the set of G-invariant states, i.e. all ρ with $[V, \rho]{=}0$ for all $V \in G$ and the corresponding twirl operation $P_G\sigma{=}\int_G V\sigma V^*\,\mathrm{d}V$. Particular examples we are looking at are: (1) Werner states where G consists of all unitaries $U \otimes U$, (2) isotropic states where each $V \in G$ has the form $V = U \otimes \bar{U}$ and finally (3) OO-invariant states where G consists of unitaries $U \otimes U$ with real matrix elements ($U = \bar{U}$) and the twirl is given in Eq. (3.24).

One way to calculate E_{F} for a G-invariant state ρ consists now of the following steps: (1) Determine the set M_ρ of pure states Φ such that $P_G|\Phi\rangle\langle\Phi| = \rho$ holds. (2) Calculate the function

$$P_G\mathscr{S} \ni \rho \mapsto \epsilon_G(\rho) = \inf\{E_{\mathrm{vN}}(\sigma) \,|\, \sigma \in M_\rho\} \in \mathbb{R}\;, \tag{5.34}$$

where we have denoted the set of G-invariant states with $P_G\mathscr{S}$. (3) Determine $E_{\mathrm{F}}(\rho)$ then in terms of the *convex hull* of ϵ, i.e.

$$E_{\mathrm{F}}(\rho) = \inf\left\{\sum_j \lambda_j \epsilon(\sigma_j) | \sigma_j \in P_G\mathscr{S},\ 0 \leqslant \lambda_j \leqslant 1,\ \rho = \sum_j \lambda_j\sigma_j,\ \sum_j \lambda_j = 1\right\}\;. \tag{5.35}$$

The equality in the last equation is of course a non-trivial statement which has to be proved. We skip this point, however, and refer the reader to [159]. The advantage of this scheme relies on the fact that spaces of G invariant states are in general very low dimensional (if G is not too small). Hence, the optimization problem contained in step 3 has a much bigger chance to be tractable than the one we have to solve for the original definition of E_{F}. There is of course no guarantee that any of this three steps can be carried out in a concrete situation. For the three examples mentioned above, however, there are results available, which we will present in the following.

5.3.2. Werner states

Let us start with Werner states [159]. In this case ρ is uniquely determined by its flip expectation value $\mathrm{tr}(\rho F)$ (cf. Section 3.1.2). To determine $\Phi \in \mathscr{H} \otimes \mathscr{H}$ such that $P_{UU}|\Phi\rangle\langle\Phi| = \rho$ holds, we have to solve therefore the equation

$$\langle\Phi, F\Phi\rangle = \sum_{jk} \Phi_{jk}\overline{\Phi_{kj}} = \mathrm{tr}(F\rho)\;, \tag{5.36}$$

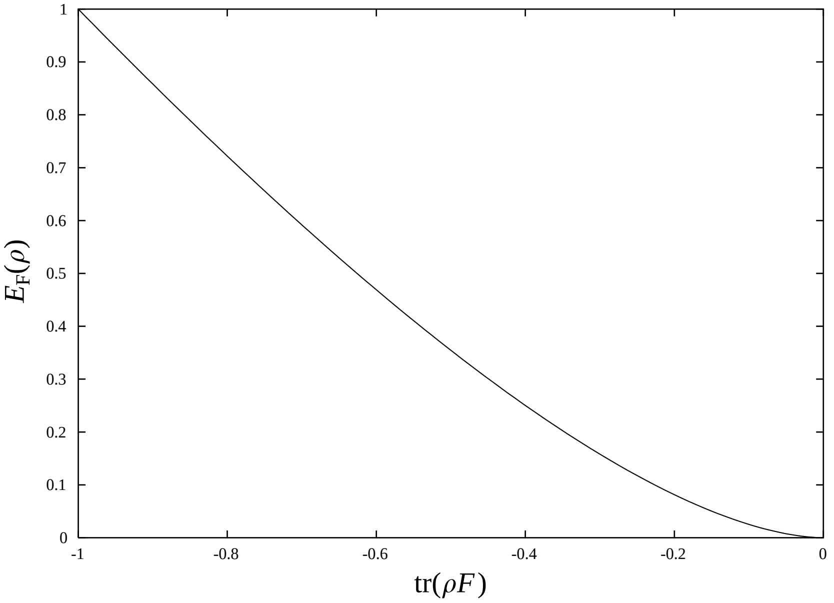


Fig. 5.2. Entanglement of formation for Werner states plotted as function of the flip expectation.

where Φ_{jk} denote components of Φ in the canonical basis. On the other hand, the reduced density matrix $\rho = \mathrm{tr}_1|\Phi\rangle\langle\Phi|$ has the matrix elements $\rho_{jk} = \sum_l \Phi_{jl}\Phi_{kl}$. By exploiting $U \otimes U$ invariance we can assume without loss of generality that ρ is diagonal. Hence, to get the function ε_{UU} we have to minimize

$$E_{\mathrm{vN}}(|\Phi\rangle\langle\Phi|) = \sum_j S\left[\sum_k |\Phi_{jk}|^2\right] \tag{5.37}$$

under constraint (5.36), where $S(x) = -x\log_2(x)$ denotes the von Neumann entropy. We skip these calculations here (see [159] instead) and state the results only. For $\mathrm{tr}(F\rho) \geqslant 0$ we get $\varepsilon(\rho) = 0$ (as expected since ρ is separable in this case) and with H from (5.16)

$$\epsilon_{UU}(\rho) = H[\tfrac{1}{2}(1 - \sqrt{1 - \mathrm{tr}(F\rho)^2})] \tag{5.38}$$

for $\mathrm{tr}(F\rho) < 0$. The minima are taken for Φ where all Φ_{jk} except one diagonal element are zero in the case $\mathrm{tr}(F\rho) \geqslant 0$ and for Φ with only two (non-diagonal) coefficients Φ_{jk}, Φ_{kj}, $j \neq k$ non-zero if $\mathrm{tr}(\rho F) < 0$. The function ε is convex and coincides therefore with its convex hull such that we get

Proposition 5.7. *For any Werner state ρ the entanglement of formation is given by* (*cf.* Fig. 5.2)

$$E_F(\rho) = \begin{cases} H[\frac{1}{2}(1 - \sqrt{1 - \mathrm{tr}(F\rho)^2})]\,, & \mathrm{tr}(F\rho) < 0\,, \\ 0, & \mathrm{tr}(F\rho) \geqslant 0\,. \end{cases} \tag{5.39}$$

5.3.3. Isotropic states

Let us now consider isotropic, i.e. $U \otimes \bar{U}$ invariant states. They are determined by the expectation value $\mathrm{tr}(\rho\tilde{F})$ with $\tilde{F}$ from Eq. (3.14). Hence, we have to look first for pure states Φ with $\langle \Phi, \tilde{F}\Phi\rangle = \mathrm{tr}(\rho\tilde{F})$ (since this determines, as for Werner states above, those Φ with $P_{U\bar{U}}(|\Phi\rangle\langle\Phi|) = \rho$). To this end assume that Φ has the Schmidt decomposition $\Phi = \sum_j \lambda_j f_j \otimes f'_j = U_1 \otimes U_2 \sum_j \lambda_j e_j \otimes e_j$ with appropriate unitary matrices U_1, U_2 and the canonical basis e_j, $j = 1,\dots,d$. Exploiting the $U \otimes \bar{U}$ invariance of ρ we get

$$\mathrm{tr}(\rho\tilde{F}) = \left\langle (\mathbb{1} \otimes V)\sum_j \lambda_j e_j \otimes e_j, \tilde{F}(\mathbb{1} \otimes V)\sum_k \lambda_k e_k \otimes e_k \right\rangle \tag{5.40}$$

$$= \sum_{j,k,l,m} \lambda_j \lambda_k \langle e_j \otimes Ve_j, e_l \otimes e_l\rangle \langle e_m \otimes e_m, e_k \otimes Ve_k\rangle \tag{5.41}$$

$$= \left| \sum_j \lambda_j \langle e_j, Ve_j\rangle \right|^2 \tag{5.42}$$

with $V = U_1^{\mathrm{T}} U_2$ and after inserting the definition of $\tilde{F}$. Following our general scheme, we have to minimize $E_{\mathrm{vN}}(|\Phi\rangle\langle\Phi|)$ under the constraint given in Eq. (5.42). This is explicitly done in [150]. We will only state the result here, which leads to the function

$$\epsilon_{U\bar{U}}(\rho) = \begin{cases} H(\gamma) + (1-\gamma)\log_2(d-1), & \mathrm{tr}(\rho\tilde{F}) \geqslant \dfrac{1}{d}\,, \\ 0, & \mathrm{tr}(\rho\tilde{F}) < 0 \end{cases} \tag{5.43}$$

with

$$\gamma = \frac{1}{d^2}\left(\sqrt{\mathrm{tr}(\rho\tilde{F})} + \sqrt{[d-1][d - \mathrm{tr}(\rho\tilde{F})]}\right)^2 . \tag{5.44}$$

For $d \geqslant 3$ this function is not convex (cf. Fig. 5.3), hence we get

Proposition 5.8. *For any isotropic state the entanglement of formation is given as the convex hull*

$$E_{\mathrm{F}}(\rho) = \inf\left\{ \sum_j \lambda_j \epsilon_{U\bar{U}}(\sigma_j) \,\middle|\, \rho = \sum_j \lambda_j \sigma_j,\ P_{U\bar{U}}\sigma = \sigma \right\} \tag{5.45}$$

of the function $\varepsilon_{U\bar{U}}$ *in Eq.* (5.43).

5.3.4. OO-invariant states

The results derived for isotropic and Werner states can be extended now to a large part of the set of OO-invariant states without solving new minimization problems. This is possible, because the definition of E_{F} in Eq. (5.13) allows under some conditions an easy extension to a suitable set of non-symmetric states. If more precisely a non-trivial, minimizing decomposition $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$ of ρ is known, all states ρ' which are a convex linear combination of the same $|\psi_j\rangle\langle\psi_j|$ but arbitrary p'_j have the same E_{F} as ρ (see [159] for proof of the statement). For the general scheme we have presented in Section 5.3.1 this implies the following: If we know the pure states $\sigma \in M_\rho$ which solve the minimization problem for $\epsilon(\rho)$ in Eq. (5.34) we get a minimizing decomposition of ρ in terms

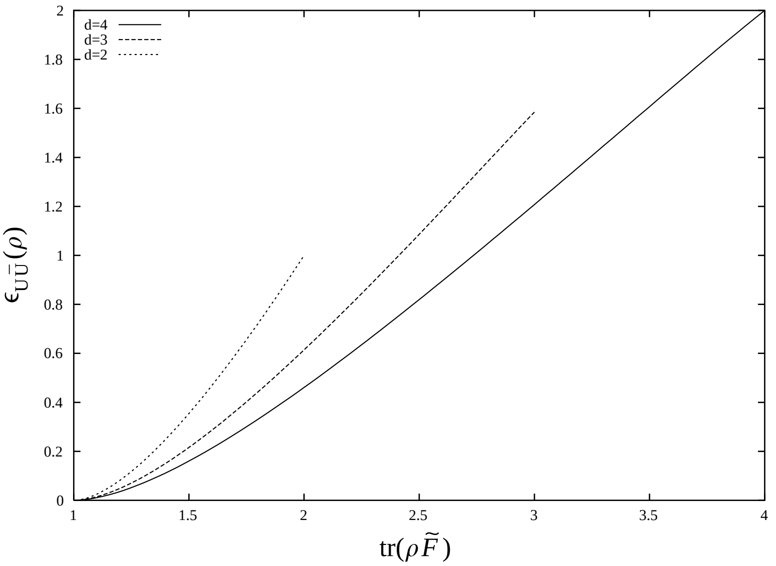


Fig. 5.3. ε-function for isotopic states plotted as a function of the flip expectation. For $d > 2$ it is not convex near the right endpoint.

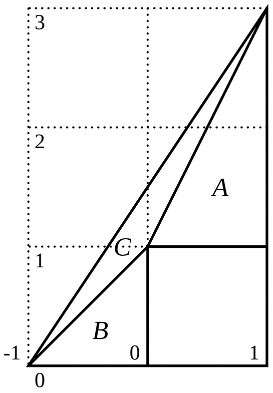


Fig. 5.4. State space of OO-invariant states.

of $U \in G$ translated copies of σ. This follows from the fact that ρ is by definition of M_ρ the twirl of σ. Hence any convex linear combination of pure states $U\sigma U^*$ with $U \in G$ has the same E_F as ρ.

A detailed analysis of the corresponding optimization problems in the case of Werner and isotropic states (which we have omitted here; see [159,150] instead) leads therefore to the following results about OO-invariant states: The space of OO-invariant states decomposes into four regions: The separable square and three triangles A, B, C; cf. Fig. 5.4. For all states ρ in triangle A we can calculate $E_F(\rho)$ as for Werner states in Proposition 5.7 and in triangle B we have to apply the result for isotropic states from Proposition 5.8. This implies in particular that E_F depends in A only on $\mathrm{tr}(\rho F)$ and in B only on $\mathrm{tr}(\rho\tilde{F})$ and the dimension.

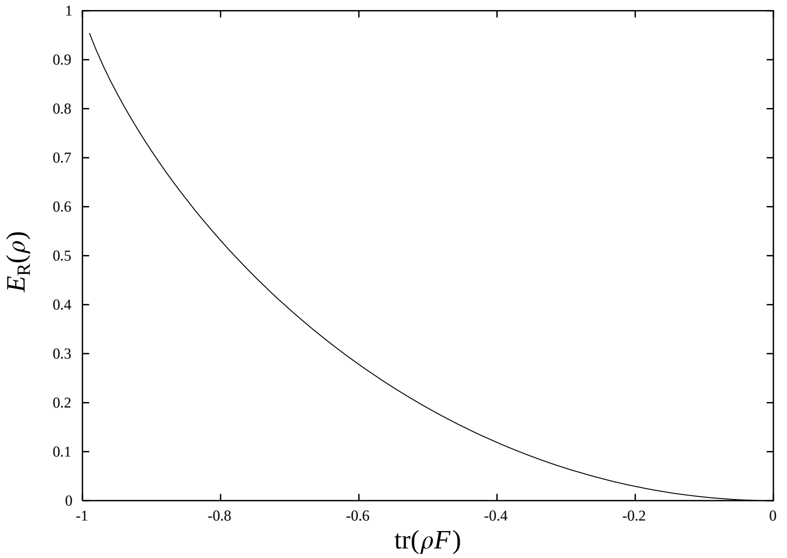


Fig. 5.5. Relative entropy of entanglement for Werner states, plotted as a function of the flip expectation.

5.3.5. Relative entropy of entanglement

To calculate $E_R(\rho)$ for a symmetric state ρ is even easier as the treatment of $E_F(\rho)$, because we can restrict the minimization in the definition of $E_R(\rho)$ in Eq. (5.14) to G-invariant separable states, provided G is a group of local unitaries. To see this assume that $\sigma \in \mathscr{D}$ minimizes $S(\rho|\sigma)$ for a G-invariant state ρ. Then we get $S(\rho|U\sigma U^*) = S(\rho|\sigma)$ for all $U \in G$ since the relative entropy S is invariant under unitary transformations of both arguments and due to its convexity we even get $S(\rho|P_G\sigma) \leqslant S(\rho|\sigma)$. Hence $P_G\sigma$ minimizes $S(\rho|\cdot)$ as well, and since $P_G\sigma \in \mathscr{D}$ holds for a group G of local unitaries, we get $E_R(\sigma, \rho) = S(\rho|P_G\sigma)$ as stated.

The sets of Werner and isotropic states are just intervals and the corresponding separable states form subintervals over which we have to perform the optimization. Due to the convexity of the relative entropy in both arguments, however, it is clear that the minimum is attained exactly at the boundary between entangled and separable states. For Werner states this is the state σ_0 with $\mathrm{tr}(F\sigma_0) = 0$, i.e. it gives equal weight to both minimal projections. To get $E_R(\rho)$ for a Werner state ρ we have to calculate therefore only the relative entropy with respect to this state. Since all Werner states can be simultaneously diagonalized this is easily done and we get (cf. Fig. 5.5)

$$E_R(\rho) = 1 - H\left(\frac{1 + \mathrm{tr}(F\rho)}{2}\right) . \qquad (5.46)$$

Similarly, the boundary point σ_1 for isotropic states is given by $\mathrm{tr}(\tilde{F}\sigma_1) = 1$ which leads to (cf. Fig. 5.6)

$$E_R(\rho) = \log_2 d - \left(1 - \frac{\mathrm{tr}(\tilde{F}\rho)}{d}\right)\log_2(d-1) - S\left(\frac{\mathrm{tr}(\tilde{F}\rho)}{d}, \frac{1 - \mathrm{tr}(\tilde{F}\rho)}{d}\right) \qquad (5.47)$$

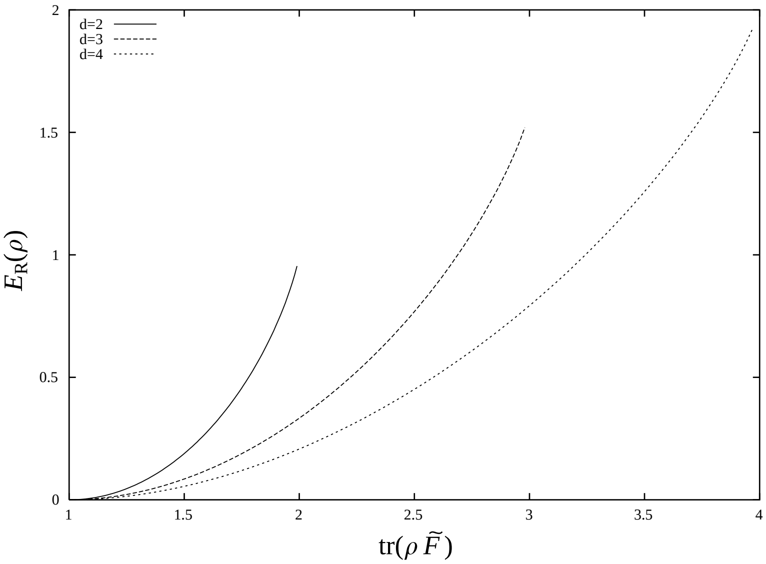


Fig. 5.6. Relative entropy of entanglement for isotropic states and $d = 2, 3, 4$, plotted as a function of $\mathrm{tr}(\rho\tilde{F})$.

for each entangled isotropic state ρ, and 0 if ρ is separable. ($S(p_1, p_2)$ denotes here the entropy of the probability vector (p_1, p_2).)

Let us now consider OO-invariant states. As for EOF we divide the state space into the separable square and the three triangles A, B, C; cf. Fig. 5.4. The state at the coordinates $(1, d)$ is a maximally entangled state and all separable states on the line connecting $(0, 1)$ with $(1, 1)$ minimize the relative entropy for this state. Hence consider a particular state σ on this line. The convexity property of the relative entropy immediately shows that σ is a minimizer for all states on the line connecting σ with the state at $(1, d)$. In this way, it is easy to calculate $E_R(\rho)$ for all ρ in A. In a similar way we can treat the triangle B: We just have to draw a line from ρ to the state at $(-1, 0)$ and find the minimizer for ρ at the intersection with the separable border between $(0, 0)$ and $(0, 1)$. For all states in the triangle C the relative entropy is minimized by the separable state at $(0, 1)$.

An application of the scheme just reviewed is a proof that E_R is not additive, i.e. it does not satisfy Axiom E5b. To see this consider the state $\rho = \mathrm{tr}(P_-)^{-1}P_-$ where P_- denotes the projector on the antisymmetric subspace. It is a Werner state with flip expectation -1 (i.e. it corresponds to the point $(-1, 0)$ in Fig. 5.4). According to our discussion above $S(\rho|\cdot)$ is minimized in this case by the separable state σ_0 and we get $E_R(\rho) = 1$ independently of the dimension d. The tensor product $\rho^{\otimes 2}$ can be regarded as a state in $\mathcal{S}(\mathcal{H}^{\otimes 2} \otimes \mathcal{H}^{\otimes 2})$ with $U \otimes U \otimes V \otimes V$ symmetry, where U, V are unitaries on $\mathcal{H}$. Note that the corresponding state space of $UUVV$ invariant states can be parameterized by the expectation of the three operators $F \otimes \mathbb{1}$, $\mathbb{1} \otimes F$ and $F \otimes F$ (cf. [159]) and we can apply the machinery just described to get the minimizer $\tilde{\sigma}$ of $S(\rho|\cdot)$. If $d > 2$ holds it turns out that

$$\tilde{\sigma} = \frac{d+1}{2d\,\mathrm{tr}(P_+)^2} P_+ \otimes P_+ + \frac{d-1}{2d\,\mathrm{tr}(P_-)^2} P_- \otimes P_- \tag{5.48}$$

holds (where $P_\pm$ denote the projections onto the symmetric and antisymmetric subspaces of $\mathscr{H}\otimes\mathscr{H}$) and not $\tilde{\sigma}=\sigma_0\otimes\sigma_0$ as one would expect. As a consequence we get the inequality

$$E_{\mathrm{R}}(\rho^{\otimes 2}) = 2-\log_2\left(\frac{2d-1}{d}\right) < 2 = S(\rho^{\otimes 2}|\sigma_0^{\otimes 2}) = 2E_{\mathrm{R}}(\rho) . \tag{5.49}$$

$d=2$ is a special case, where $\sigma_0^{\otimes 2}$ and $\tilde{\sigma}$ (and all their convex linear combination) give the same value 2. Hence for $d>2$ the relative entropy of entanglement is, as stated, not additive.

6. Channel capacity

In Section 4.4 we have seen that it is possible to send (quantum) information undisturbed through a noisy quantum channel, if we encode one qubit into a (possibly long and highly entangled) string of qubits. This process is wasteful, since we have to use many instances of the channel to send just one qubit of quantum information. It is therefore natural to ask, which resources we need at least if we are using the best possible error correction scheme. More precisely the question is: With which maximal *rate*, i.e. information sent per channel usage, we can transmit quantum information undisturbed through a noisy channel? This question naturally leads to the concept of channel capacities which we will review in this section.

6.1. The general case

We are mainly interested in classical and quantum capacities. The basic ideas behind both situations are however quite similar. In this section we will consider therefore a general definition of capacity which applies to arbitrary channels and both kinds of information. (See also [168] as a general reference for this section.)

6.1.1. The definition

Hence consider two observable algebras $\mathscr{A}_1$, $\mathscr{A}_2$ and an arbitrary channel $T:\mathscr{A}_1\to\mathscr{A}_2$. To send systems described by a third observable algebra $\mathscr{B}$ undisturbed through T we need an *encoding channel* $E:\mathscr{A}_2\to\mathscr{B}$ and a *decoding channel* $D:\mathscr{B}\to\mathscr{A}_1$ such that ETD equals the ideal channel $\mathscr{B}\to\mathscr{B}$, i.e. the identity on $\mathscr{B}$. Note that the algebra $\mathscr{B}$ describing the systems to send, and the input, respectively output, algebra of T need not to be of the same type, e.g. $\mathscr{B}$ can be classical while $\mathscr{A}_1,\mathscr{A}_2$ are quantum (or vice versa).

In general (i.e. for arbitrary T and $\mathscr{B}$) it is of course impossible to find such a pair E and D. In this case we are interested at least in encodings and decodings which make the error produced during the transmission as small as possible. To make this statement precise we need a measure for this error and there are in fact many good choices for such a quantity (all of them leading to equivalent results, cf. Section 6.3.1). We will use in the following the "cb-norm difference" $\|ETD-\mathrm{Id}\|_{\mathrm{cb}}$, where Id is the identity (i.e. ideal) channel on $\mathscr{B}$ and $\|\cdot\|_{\mathrm{cb}}$ denotes the norm of *complete boundedness* ("cb-norm" for short)

$$\|T\|_{\mathrm{cb}} = \sup_{n\in\mathbb{N}}\|T\otimes\mathrm{Id}_n\|, \quad \mathrm{Id}_n:\mathscr{B}(\mathbb{C}^n)\to\mathscr{B}(\mathbb{C}^n) . \tag{6.1}$$

The cb-norm improves the sometimes annoying property of the usual operator norm that quantities like $\|T \otimes \mathrm{Id}_{\mathscr{B}(\mathbb{C}^d)}\|$ may increase with the dimension d. On infinite-dimensional observable algebras $\|T\|_{\mathrm{cb}}$ can be infinite although each term in the supremum is finite. A particular example for a map with such a behavior is the transposition on an infinite-dimensional Hilbert space. A map with finite cb-norm is therefore called completely bounded. In a finite-dimensional setup each linear map is completely bounded. For the transposition Θ on $\mathbb{C}^d$ we have in particular $\|\Theta\|_{\mathrm{cb}} = d$. The cb-norm has some nice features which we will use frequently; this includes its multiplicativity $\|T_1 \otimes T_2\|_{\mathrm{cb}} = \|T_1\|_{\mathrm{cb}}\|T_2\|_{\mathrm{cb}}$ and the fact that $\|T\|_{\mathrm{cb}} = 1$ holds for each (unital) channel. Another useful relation is $\|T\|_{\mathrm{cb}} = \|T \otimes \mathrm{Id}_{\mathscr{B}(\mathscr{H})}\|$, which holds if T is a map $\mathscr{B}(\mathscr{H}) \to \mathscr{B}(\mathscr{H})$. For more properties of the cb-norm let us refer to [125].

Now we can define the quantity

$$\Delta(T, \mathscr{B}) = \inf_{E,D} \|ETD - \mathrm{Id}_{\mathscr{B}}\|_{\mathrm{cb}} , \tag{6.2}$$

where the infimum is taken over all channels $E : \mathscr{A}_2 \to \mathscr{B}$ and $D : \mathscr{B} \to \mathscr{A}_1$ and $\mathrm{Id}_{\mathscr{B}}$ is again the ideal $\mathscr{B}$-channel. Δ describes, as indicated above, the smallest possible error we have to take into account if we try to transmit *one* $\mathscr{B}$ system through *one* copy of the channel T using any encoding E and decoding D. In Section 4.4, however, we have seen that we can reduce the error if we take M copies of the channel instead of just one. More generally we are interested in the transmission of "codewords of length" N, i.e. $\mathscr{B}^{\otimes N}$ systems using M copies of the channel T. Encodings and decodings are in this case channels of the form $E : \mathscr{A}_2^{\otimes M} \to \mathscr{B}^{\otimes N}$ respectively $D : \mathscr{B}^{\otimes N} \to \mathscr{A}_1^{\otimes M}$. If we increase the number M of channels the error $\Delta(T^{\otimes M}, \mathscr{B}^{\otimes N(M)})$ decreases provided the rate with which N grows as a function of M is not too large. A more precise formulation of this idea leads to the following definition.

Definition 6.1. Let T be a channel and $\mathscr{B}$ an observable algebra. A number $c \geqslant 0$ is called *achievable rate* for T with respect to $\mathscr{B}$, if for any pair of sequences M_j, N_j, $j \in \mathbb{N}$ with $M_j \to \infty$ and $\limsup_{j\to\infty} N_j/M_j < c$ we have

$$\lim_{j\to\infty} \Delta(T^{\otimes M_j}, \mathscr{B}^{\otimes N_j}) = 0 . \tag{6.3}$$

The supremum of all achievable rates is called the *capacity* of T with respect to $\mathscr{B}$ and denoted by $C(T, \mathscr{B})$.

Note that by definition $c = 0$ is an achievable rate hence $C(T, \mathscr{B}) \geqslant 0$. If on the other hand each $c > 0$ is achievable we write $C(T, \mathscr{B}) = \infty$. At a first look it seems cumbersome to check all pairs of sequences with given upper ratio when testing c. Due to some monotonicity properties of Δ, however, it can be shown that it is sufficient to check only one sequence provided the M_j satisfy the additional condition $M_j/(M_{j+1}) \to 1$.

6.1.2. Simple calculations

We see that there are in fact many different capacities of a given channel depending on the type of information we want to transmit. However, there are only two different cases we are interested in: $\mathscr{B}$ can be either classical or quantum. We will discuss both special cases in greater detail in the next two sections. Before we do this, however, we will have a short look on some simple calculations which can be done in the general case. To this end it is convenient to introduce the notations

$$\mathscr{M}_d = \mathscr{B}(\mathbb{C}^d) \quad \text{and} \quad \mathscr{C}_d = \mathscr{C}(\{1,\dots,d\}) \tag{6.4}$$

as shorthand notations for $\mathscr{B}(\mathbb{C}^d)$ and $\mathscr{C}(\{1,\dots,d\})$ since some notations become otherwise a little bit clumsy. First of all let us have a look on capacities of ideal channels. If $\mathrm{Id}_{\mathscr{M}_f}$ and $\mathrm{Id}_{\mathscr{C}_f}$ denote the identity channels on the quantum algebra $\mathscr{M}_f$, respectively the classical algebra $\mathscr{C}_f$, we get

$$C(\mathrm{Id}_{\mathscr{C}_f}, \mathscr{M}_d) = 0, \quad C(\mathrm{Id}_{\mathscr{C}_f}, \mathscr{C}_d) = C(\mathrm{Id}_{\mathscr{M}_f}, \mathscr{M}_d) = C(\mathrm{Id}_{M_f}, \mathscr{C}_d) = \frac{\log_2 f}{\log_2 d} \ . \tag{6.5}$$

The first equation is the channel capacity version of the no-teleportation theorem: It is impossible to transfer quantum information through a classical channel. The other equations follow simply by counting dimensions.

For the next relation it is convenient to associate to a pair of channels T, S the quantity $C(T,S)$ which arises if we replace in Definition 6.1 and Eq. (6.2) the ideal channel $\mathrm{Id}_{\mathscr{B}}$ by an arbitrary channel S. Hence $C(T,S)$ is a slight generalization of the channel capacity which describes with which asymptotic rate the channel S can be approximated by T (and appropriate encodings and decodings). These generalized capacities satisfy the *two-step coding inequality*, i.e. for the three channels T_1, T_2, T_3 we have

$$C(T_3, T_1) \geqslant C(T_2, T_1)C(T_3, T_2). \tag{6.6}$$

To prove it consider the relations

$$\begin{aligned} &\|T_1^{\otimes N} - E_1E_2T_3^{\otimes K}D_2D_1\|_{\mathrm{cb}} \\ &\quad = \|T_1^{\otimes N} - E_1T_2^{\otimes M}D_1 + E_1T_2^{\otimes M}D_1 - E_1E_2T_3^{\otimes K}D_2D_1\|_{\mathrm{cb}} &(6.7)\\ &\quad \leqslant \|T_1^{\otimes N} - E_1T_2^{\otimes M}D_1\|_{\mathrm{cb}} + \|E_1\|_{\mathrm{cb}}\|T_2^{\otimes M} - E_2T_3^{\otimes K}D_2\|_{\mathrm{cb}}\|D_1\|_{\mathrm{cb}} &(6.8)\\ &\quad \leqslant \|T_1^{\otimes N} - E_1T_2^{\otimes M}D_1\|_{\mathrm{cb}} + \|T_2^{\otimes M} - E_2T_3^{\otimes K}D_2\|_{\mathrm{cb}} \ , &(6.9) \end{aligned}$$

where we have used for the last inequality the fact that the cb-norm of a channel is one. If c_1 is an achievable rate of T_1 with respect to T_2 such that $\limsup_{j\to\infty} M_j/N_j < c_1$ and c_2 is an achievable rate of T_2 with respect to T_3 such that $\limsup_{j\to\infty} N_j/K_j < c_2$ we see that

$$\limsup_{j\to\infty} \frac{M_j}{K_j} = \limsup_{j\to\infty} \frac{M_j}{N_j}\frac{N_j}{K_j} \leqslant \limsup_{j\to\infty} \frac{M_j}{N_j} \limsup_{k\to\infty} \frac{N_k}{K_k} \ . \tag{6.10}$$

If we choose the sequences M_j, N_j and K_j clever enough (cf. the remark following Definition 6.1) this implies that c_1c_2 is an achievable rate for T_1 with respect to T_3 and this proves Eq. (6.6).

As a first application of (6.6), we can relate all capacities $C(T,\mathscr{M}_d)$ (and $C(T,\mathscr{C}_d)$) for different d to one another. If we choose $T_3 = T$, $T_1 = \mathrm{Id}_{\mathscr{M}_d}$ and $T_2 = \mathrm{Id}_{\mathscr{M}_f}$ we get with (6.5) $C(T,\mathscr{M}_d) \leqslant (\log_2 f/\log_2 d)C(T,\mathscr{M}_f)$, and exchanging d with f shows that even equality holds.

A similar relation can be shown for $C(T, \mathscr{C}_d)$. Hence, the dimension of the observable algebra $\mathscr{B}$ describing the type of information to be transmitted, enters only via a multiplicative constant, i.e. it is only a choice of units and we define the *classical capacity* $C_c(T)$ and the *quantum capacity* $C_q(T)$ of a channel T as

$$C_c(T) = C(T, \mathscr{C}_2), \quad C_q(T) = C(T, \mathscr{M}_2) . \tag{6.11}$$

A second application of Eq. (6.6) is a relation between the classical and the quantum capacity of a channel. Setting $T_3 = T$, $T_1 = \mathrm{Id}_{\mathscr{C}_2}$ and $T_2 = \mathrm{Id}_{\mathscr{M}_2}$ we get again with (6.5),

$$C_q(T) \leqslant C_c(T) . \tag{6.12}$$

Note that it is now not possible to interchange the roles of $\mathscr{C}_2$ and $\mathscr{M}_2$. Hence equality does not hold here.

Another useful relation concerns concatenated channels: We transmit information of type $\mathscr{B}$ first through a channel T_1 and then through a second channel T_2. It is reasonable to assume that the capacity of the composition $T_2 T_1$ cannot be bigger than capacity of the channel with the smallest bandwidth. This conjecture is indeed true and known as the "*Bottleneck inequality*":

$$C(T_2 T_1, \mathscr{B}) \leqslant \min\{C(T_1, \mathscr{B}), C(T_2, \mathscr{B})\} . \tag{6.13}$$

To see this consider an encoding and a decoding channel E, respectively D, for $(T_2 T_1)^{\otimes M}$, i.e. in the definition of $C(T_2 T_1, \mathscr{B})$ we look at

$$\|\mathrm{Id}_{\mathscr{B}}^{\otimes N} - E(T_2 T_1)^{\otimes M} D\|_{\mathrm{cb}} = \|\mathrm{Id}_{\mathscr{B}}^{\otimes N} - (E T_2^{\otimes M}) T_1^{\otimes M} D\|_{\mathrm{cb}} . \tag{6.14}$$

This implies that $E T_2^{\otimes M}$ and D are an encoding and a decoding channel for T_1. Something similar holds for D and $T_1^{\otimes M} D$ with respect to T_2. Hence each achievable rate for $T_2 T_1$ is also an achievable rate for T_2 and T_1, and this proves Eq. (6.13).

Finally, we want to consider two channels T_1, T_2 in parallel, i.e. we consider the tensor product $T_1 \otimes T_2$. If E_j, D_j, $j = 1, 2$ are encoding, respectively decoding, channels for $T_1^{\otimes M}$ and $T_2^{\otimes M}$ such that $\|\mathrm{Id}_{\mathscr{B}}^{\otimes N_j} - E_j T_j^{\otimes M} D_j\|_{\mathrm{cb}} \leqslant \epsilon$ holds, we get

$$\|\mathrm{Id} - \mathrm{Id} \otimes (E_2 T^{\otimes M} D_2) + \mathrm{Id} \otimes (E_2 T^{\otimes M} D_2) - E_1 \otimes E_2 (T_1 \otimes T_2)^{\otimes M} D_1 \otimes D_2\|_{\mathrm{cb}} \tag{6.15}$$

$$\leqslant \|\mathrm{Id} \otimes (\mathrm{Id} - E_2 T^{\otimes M} D_2)\|_{\mathrm{cb}} + \|(\mathrm{Id} - E_1 T_1^{\otimes M} D_1) \otimes E_2 T^{\otimes M} D_2\|_{\mathrm{cb}} \tag{6.16}$$

$$\leqslant \|\mathrm{Id} - E_2 T^{\otimes M} D_2\|_{\mathrm{cb}} + \|\mathrm{Id} - E_1 T_1^{\otimes M} D_1\|_{\mathrm{cb}} \leqslant 2\epsilon . \tag{6.17}$$

Hence $c_1 + c_2$ is achievable for $T_1 \otimes T_2$ if c_j is achievable for T_j. This implies the inequality

$$C(T_1 \otimes T_2, \mathscr{B}) \geqslant C(T_1, \mathscr{B}) + C(T_2, \mathscr{B}) . \tag{6.18}$$

When all channels are ideal, or when all systems involved are classical even equality holds, i.e. channel capacities are *additive* in this case. However, if quantum channels are considered, it is one of the big open problems of the field, to decide under which conditions additivity holds.

6.2. The classical capacity

In this section we will discuss the classical capacity $C_c(T)$ of a channel T. There are in fact three different cases to consider: T can be either classical or quantum and in the quantum case we can use either ordinary encodings and decodings or a dense coding scheme (cf. Section 4.1.3).

6.2.1. Classical channels

Let us consider first a classical to classical channel $T:\mathscr{C}(Y) \to \mathscr{C}(X)$. This is basically the situation of classical information theory and we will only have a short look here—mainly to show how this (well known) situation fits into the general scheme described in the last section.[20]

First of all we have to calculate the error quantity $\Delta(T, \mathscr{C}_2)$ defined in Eq. (6.20). As stated in Section 3.2.3 T is completely determined by its transition probabilities T_{xy}, $(x, y) \in X \times Y$ describing the probability to receive $x \in X$ when $y \in Y$ was sent. Since the cb-norm for a classical algebra coincides with the ordinary norm we get (we have set $X = Y$ for this calculation)

$$\|\mathrm{Id} - T\|_{\mathrm{cb}} = \|\mathrm{Id} - T\| = \sup_{x,f} \left| \sum_y (\delta_{xy} - T_{xy}) f_y \right| \tag{6.19}$$

$$= 2 \sup_x (1 - T_{xx}) , \tag{6.20}$$

where the supremum in the first equation is taken over all $f \in \mathscr{C}(X)$ with $\|f\| = \sup_y |f_y| \leqslant 1$. We see that the quantity in Eq. (6.20) is exactly twice the *maximal error probability*, i.e. the maximal probability of sending x and getting anything different. Inserting this quantity for Δ in Definition 6.1 applied to a classical channel T and the "bit-algebra" $\mathscr{B} = \mathscr{C}_2$, we get exactly the Shannons classical definition of the capacity of a discrete memoryless channel [138].

Hence we can apply the Shannons *noisy channel coding theorem* to calculate $C_c(T)$ for a classical channel. To state it we have to introduce first some terminology. Consider therefore a state $p \in \mathscr{C}^*(X)$ of the classical input algebra $\mathscr{C}(X)$ and its image $q = T^*(p) \in \mathscr{C}^*(Y)$ under the channel. p and q are probability distributions on X, respectively Y, and p_x can be interpreted as the probability that the "letter" $x \in X$ was send. Similarly $q_y = \sum_x T_{xy} p_x$ is the probability that $y \in Y$ was received and $P_{xy} = T_{xy} p_x$ is the probability that $x \in X$ was sent and $y \in Y$ was received. The family of all P_{xy} can be interpreted as a probability distribution P on $X \times Y$ and the T_{xy} can be regarded as conditional probability of P under the condition x. Now we can introduce the *mutual information*

$$I(p, T) = S(p) + S(q) - S(P) = \sum_{(x,y) \in X \times Y} P_{xy} \log_2 \left(\frac{P_{xy}}{p_x q_y} \right) , \tag{6.21}$$

where $S(p)$, $S(q)$ and $S(P)$ denote the entropies of p, q and P. The mutual information describes, roughly speaking, the information that p and q contain about each other. E.g. if p and q are completely uncorrelated (i.e. $P_{xy} = p_x q_y$) we get $I(p, T) = 0$. If T is on the other hand an ideal bit-channel and p equally distributed we have $I(p, T) = 1$. Now we can state the Shannons Theorem which expresses the classical capacity of T in terms of mutual informations [138]:

[20] Please note that this implies in particular that we do not give a complete review of the foundations of classical information theory here; cf. [101,62,49] instead.

Theorem 6.2 (Shannon). *The classical capacity of $C_c(T)$ of a classical communication channel $T : \mathscr{C}(Y) \to \mathscr{C}(X)$ is given by*

$$C_c(T) = \sup_p I(p, T) , \tag{6.22}$$

where the supremum is taken over all states $p \in \mathscr{C}^*(X)$.

6.2.2. Quantum channels

If we transmit classical data through a quantum channel $T : \mathscr{B}(\mathscr{H}) \to \mathscr{B}(\mathscr{H})$ the encoding $E : \mathscr{B}(\mathscr{H}) \to \mathscr{C}_2$ is a parameter-dependent preparation and the decoding $D : \mathscr{C}_2 \to \mathscr{B}(\mathscr{H})$ is an observable. Hence, the composition ETD is a channel $\mathscr{C}_2 \to \mathscr{C}_2$, i.e. a purely classical channel and we can calculate its capacity in terms of the Shannons Theorem (Theorem 6.2). This observation leads to the definition of the "*one-shot*" classical capacity of T:

$$C_{c,1}(T) = \sup_{E,D} C_c(ETD) , \tag{6.23}$$

where the supremum is taken over all encodings and decodings of classical bits. The term "one-shot" in this definition arises from the fact that we need apparently only one invocation of the channel T. However, many uses of the channel are hidden in the definition of the classical capacity on the right-hand side. Hence, $C_{c,1}(T)$ can be defined alternatively in the same way as $C_c(T)$ except that no entanglement is allowed during encoding and decoding, or more precisely in Definition 6.1 we consider only encodings $E : \mathscr{B}(\mathscr{K})^{\otimes M} \to \mathscr{C}_2^{\otimes N}$ which prepare separable states and only decodings $D : \mathscr{C}_2^{\otimes N} \to \mathscr{B}(\mathscr{H})^{\otimes M}$ which lead to separable observables. It is not yet known, whether entangled codings can help to increase the transmission rate. Therefore, we only know that

$$C_{c,1}(T) \leqslant C_c(T) = \sup_{M \in \mathbb{N}} \frac{1}{M} C_{c,1}(T^{\otimes M}) \tag{6.24}$$

holds. One reason why $C_{c,1}(T)$ is an interesting quantity relies on the fact that we have, due to the following theorem by Holevo [80], a computable expression for it.

Theorem 6.3. *The one-shot classical capacity $C_{c,1}(T)$ of a quantum channel $T : \mathscr{B}(\mathscr{H}) \to \mathscr{B}(\mathscr{H})$ is given by*

$$C_{c,1}(T) = \sup_{p_j, \rho_j} \left[S\left(\sum_j p_j T^*[\rho_j] \right) - \sum_j p_j S(T^*[\rho_j]) \right] , \tag{6.25}$$

where the supremum is taken over all probability distributions p_j and collections of density operators ρ_j.

6.2.3. Entanglement assisted capacity

Another classical capacity of a quantum channel arises, if we use dense coding schemes instead of simple encodings and decodings to transmit the data through the channel T. In other words we can define the *entanglement enhanced classical capacity* $C_e(T)$ in the same way as $C_c(T)$ but by replacing the encoding and decoding channels in Definition 6.1 and Eq. (6.2) by dense coding protocols. Note that this implies that the sender Alice and the receiver Bob share an (arbitrary) amount of (maximally) entangled states prior to the transmission.

For this quantity a coding theorem was recently proven by Bennett and others [18] which we want to state in the following. To this end assume that we are transmitting systems in the state $\rho \in \mathscr{B}^*(\mathscr{H})$ through the channel and that ρ has the purification $\Psi \in \mathscr{H} \otimes \mathscr{H}$, i.e. $\rho = \mathrm{tr}_1 |\Psi\rangle\langle\Psi| = \mathrm{tr}_2 |\Psi\rangle\langle\Psi|$. Then we can define the *entropy exchange*

$$S(\rho, T) = S[(T \otimes \mathrm{Id})(|\Psi\rangle\langle\Psi|)] \ . \tag{6.26}$$

The density operator $(T \otimes \mathrm{Id})(|\Psi\rangle\langle\Psi|)$ has the output state $T^*(\rho)$ and the input state ρ as its partial traces. It can be regarded therefore as the quantum analog of the input/output probability distribution T_{xy} defined in Section 6.2.1. Another way to look at $S(\rho, T)$ is in terms of an ancilla representation of T: If $T^*(\rho) = \mathrm{tr}_{\mathscr{K}}(U\rho \otimes \rho_{\mathscr{K}} U^*)$ with a unitary $U : \mathscr{H} \otimes \mathscr{K}$ and a pure environment state $\rho_{\mathscr{K}}$ it can be shown [7] that $S(\rho, T) = S[T^*_{\mathscr{K}}\rho]$ where $T_{\mathscr{K}}$ is the channel describing the information transfer into the environment, i.e. $T^*_{\mathscr{K}}(\rho) = \mathrm{tr}_{\mathscr{H}}(U\rho \otimes \rho_{\mathscr{K}} U^*)$, in other words $S(\rho, T)$ is the final entropy of the environment. Now we can define

$$I(\rho, T) = S(\rho) + S(T^*\rho) - S(\rho, T) \ , \tag{6.27}$$

which is the quantum analog of the mutual information given in Eq. (6.21). It has a number of nice properties, in particular positivity, concavity with respect to the input state and additivity [2] and its maximum with respect to ρ coincides actually with $C_{\mathrm{e}}(T)$ [18].

Theorem 6.4. *The entanglement assisted capacity $C_{\mathrm{e}}(T)$ of a quantum channel $T : \mathscr{B}(\mathscr{H}) \to \mathscr{B}(\mathscr{H})$ is given by*

$$C_{\mathrm{e}}(T) = \sup_{\rho} I(\rho, T) \ , \tag{6.28}$$

where the supremum is taken over all input states $\rho \in \mathscr{B}^*(\mathscr{H})$.

Due to the nice additivity properties of the quantum mutual information $I(\rho, T)$ the capacity $C_{\mathrm{e}}(T)$ is known to be additive as well. This implies that it coincides with the corresponding “one-shot” capacity, and this is an essential simplification compared to the classical capacity $C_{\mathrm{c}}(T)$.

6.2.4. Examples

Although the expressions in Theorems 6.3 and 6.4 are much easier than the original definitions they still involve some optimization problems over possibly large parameter spaces. Nevertheless, there are special cases which allow explicit calculations. As a first example we will consider the “quantum erasure channel” which transmits with probability $1 - \vartheta$ the d-dimensional input state intact while it is replaced with probability ϑ by an “erasure symbol”, i.e. a $(d + 1)$th pure state ψ_{e} which is orthogonal to all others [72]. In the Schrödinger picture this is

$$\mathscr{B}^*(\mathbb{C}^d) \ni \rho \mapsto T^*(\rho) = (1 - \vartheta)\rho + \vartheta\,\mathrm{tr}(\rho)|\psi_{\mathrm{e}}\rangle\langle\psi_{\mathrm{e}}| \in \mathscr{B}^*(\mathbb{C}^{d+1}) \ . \tag{6.29}$$

This example is very unusual, because all capacities discussed up to now (including the quantum capacity as we will see in Section 6.3.2) can be calculated explicitly: We get $C_{\mathrm{c},1}(T) = C_{\mathrm{c}}(T) = (1 - \vartheta)\log_2(d)$ for the classical and $C_{\mathrm{e}}(T) = 2C_{\mathrm{c}}(T)$ for the entanglement enhanced classical capacity [15,17]. Hence the gain by entanglement assistance is exactly a factor two; cf. Fig. 6.1.

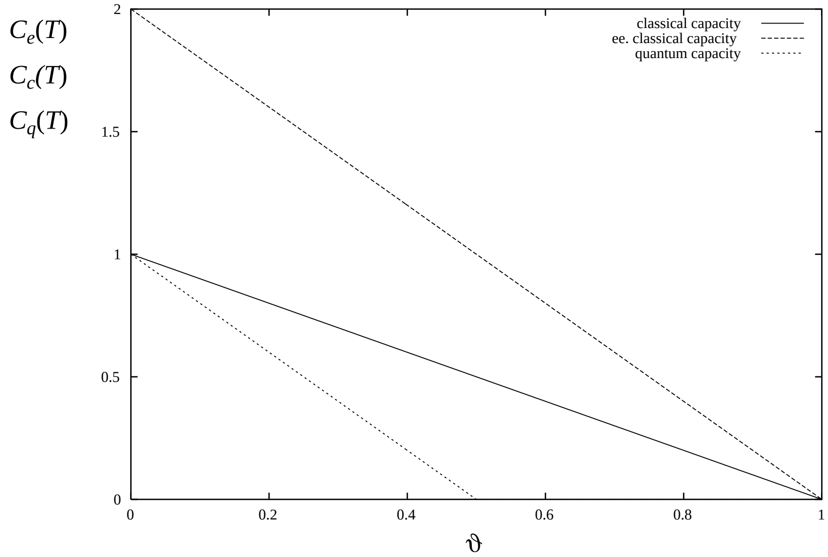


Fig. 6.1. Capacities of the quantum erasure channel plotted as a function of the error probability.

Our next example is the depolarizing channel

$$\mathscr{B}^*(\mathbb{C}^d) \ni \rho \mapsto T^*(\rho) = (1-\vartheta)\rho + \vartheta \operatorname{tr}(\rho)\frac{\mathbb{1}}{d} \in \mathscr{B}^*(\mathbb{C}^d) , \tag{6.30}$$

already discussed in Section 3.2. It is more interesting and more difficult to study. It is in particular not known whether C_c and $C_{c,1}$ coincide in this case (i.e. the value of C_c is not known. Therefore we can compare $C_e(T)$ only with $C_{c,1}$. Using the unitary covariance of T (cf. Section 3.2.2) we see first that $I(U\rho U^*, T) = I(\rho, T)$ holds for all unitaries U (to calculate $S(U\rho U^*, T)$ note that $U \otimes U\Psi$ is a purification of $U\rho U^*$ if Ψ is a purification of ρ). Due to the concavity of $I(\rho, T)$ in the first argument we can average over all unitaries and see that the maximum in Eq. (6.28) is achieved on the maximally mixed state. Straightforward calculation therefore shows that

$$C_e(T) = \log_2(d^2) + \left(1 - \vartheta\frac{d^2-1}{d^2}\right)\log_2\left(1 - \vartheta\frac{d^2-1}{d^2}\right) + \vartheta\frac{d^2-1}{d^2}\log_2\frac{\vartheta}{d^2} \tag{6.31}$$

holds, while we have

$$C_{c,1}(T) = \log_2(d) + \left(1 - \vartheta\frac{d-1}{d}\right)\log_2\left(1 - \vartheta\frac{d-1}{d}\right) + \vartheta\frac{d-1}{d}\log_2\frac{\vartheta}{d} , \tag{6.32}$$

where the maximum in Eq. (6.25) is achieved for an ensemble of equiprobable pure states taken from an orthonormal basis in $\mathscr{H}$ [82]. This is plausible since the first term under the sup in Eq. (6.25) becomes maximal and the second becomes minimal: $\sum_j p_j T^*\rho_j$ is maximally mixed in this case and its entropy is therefore maximal. The entropies of the $T^*\rho_j$ are on the other hand minimal if the ρ_j are pure. In Fig. 6.2 we have plotted both capacities as a function of the noise parameter ϑ and in Fig. 6.3 we have plotted the quotient $C_e(T)/C_{c,1}(T)$ which gives an upper bound on the gain we get from entanglement assistance.

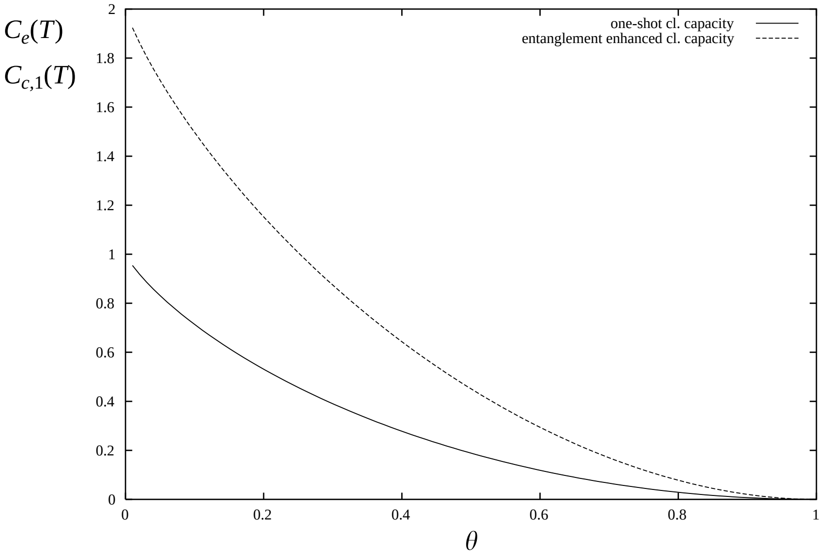


Fig. 6.2. Entanglement enhanced and one-shot classical capacity of a depolarizing qubit channel.

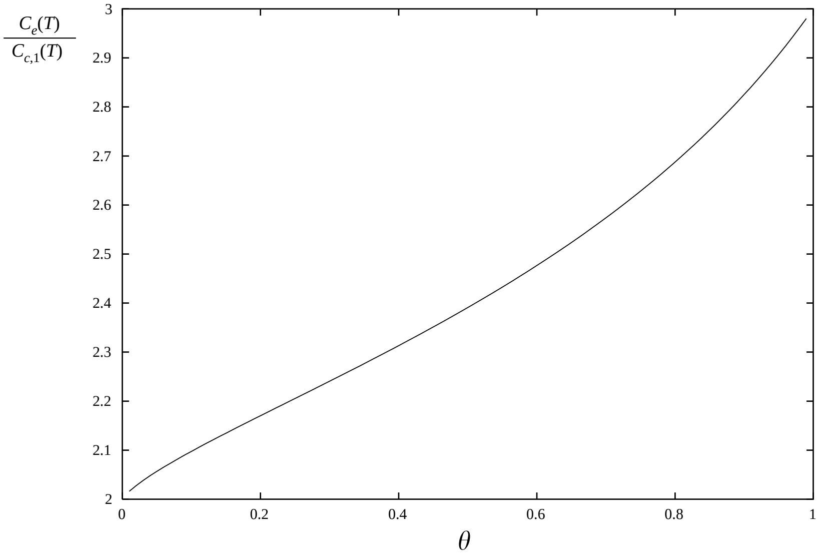


Fig. 6.3. Gain of using entanglement assisted versus unassisted classical capacity for a depolarizing qubit channel.

As a third example we want to consider Gaussian channels defined in Section 3.3.4. Hence consider the Hilbert space $\mathscr{H} = \mathrm{L}^2(\mathbb{R})$ describing a one-dimensional harmonic oscillator (or one mode of the electromagnetic field) and the amplification/attenuation channel T defined in Eq. (3.74). The results we want to state concern a slight modification of the original definitions of $C_{\mathrm{c},1}(T)$ and $C_{\mathrm{e}}(T)$: We will consider capacities for channels with *constraint input*. This means that only a restricted class of states ρ on the input Hilbert space of the channel are allowed for encoding. In our case this means

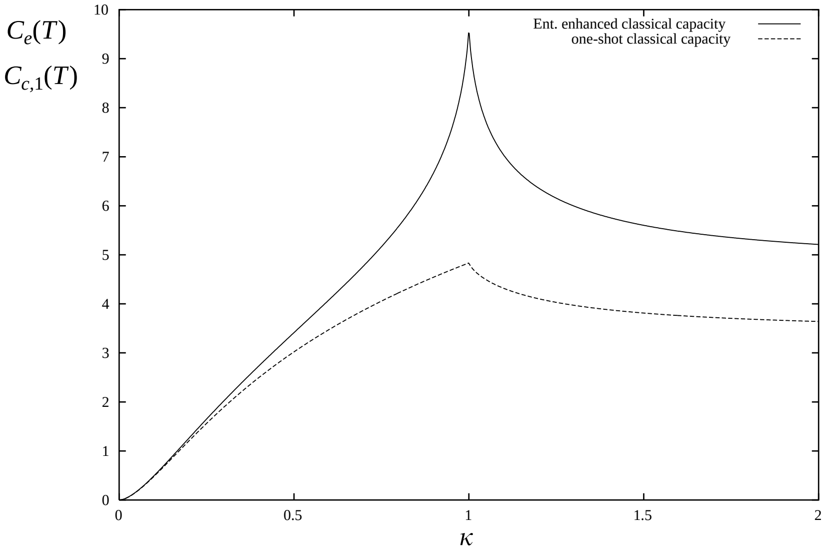


Fig. 6.4. One-shot and entanglement enhanced classical capacity of a Gaussian amplification/attenuation channel with $N_c = 0$ and input noise $N = 10$.

that we will consider the constraint $\mathrm{tr}(\rho aa^*) \leqslant N$ for a positive real number $N > 0$ and with the usual creation and annihilation operators a^*, a. This can be rewritten as an energy constraint for a quadratic Hamiltonian; hence this is a physically realistic restriction.

For the entanglement enhanced capacity it can be shown now that the maximum in Eq. (6.28) is taken on Gaussian states. To get $C_e(T)$ it is sufficient therefore to calculate the quantum mutual information $I(T, \rho)$ for the Gaussian state ρ_N from Eq. (3.64). The details can be found in [84,18], we will only state the results here. With the abbreviation

$$g(x) = (x+1)\log_2(x+1) - x\log_2 x \ , \tag{6.33}$$

we get $S(\rho_N) = g(N)$ and $S(T[\rho_N]) = g(N')$ with $N' = k^2 N + \max\{0, k^2 - 1\} + N_c$ (cf. Eq. (3.75)) for the entropies of input and output states and

$$S(\rho, T) = g\left(\frac{D + N' - N - 1}{2}\right) + g\left(\frac{D - N' + N - 1}{2}\right) \tag{6.34}$$

with

$$D = \sqrt{(N + N' + 1)^2 - 4k^2 N(N+1)} \tag{6.35}$$

for the entropy exchange. The sum of all three terms gives $C_e(T)$ which we have plotted in Fig. 6.4 as a function of k.

To calculate the one-shot capacity $C_{c,1}(T)$ the optimization in Eq. (6.25) has to be calculated over probability distributions p_j and collections of density operators ρ_j such that $\sum_j p_j \,\mathrm{tr}(aa^*\rho_j) \leqslant N$ holds. It is conjectured but not yet proven [84] that the maximum is achieved on coherent states

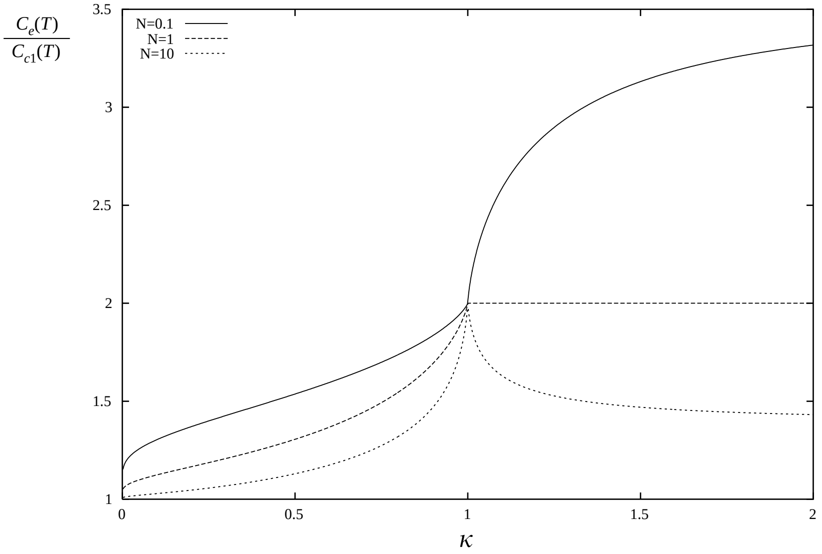


Fig. 6.5. Gain of using entanglement assisted versus unassisted classical capacity for a Gaussian amplification/attenuation channel with $N_c = 0$ and input noise $N = 0.1, 1, 10$.

with Gaussian probability distribution $p(x) = (\pi N)^{-1}\exp(-|x|^2/N)$. If this is true we get

$$C_{c,1}(T) = g(N') - g(N_0') \quad \text{with } N_0' = \max\{0, k^2 - 1\} + N_c \; . \tag{6.36}$$

The result is plotted as a function of k in Fig. 6.4 and the ratio $G = C_e/C_1$ in Fig. 6.5. G gives an upper bound on the *gain* of using entanglement assisted versus unassisted classical capacity.

6.3. The quantum capacity

The quantum capacity of a quantum channel $T : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ is more difficult to treat than the classical capacities discussed in the last section. There is, in particular, no coding theorem available which would allow explicit calculations. Nevertheless, there are partial results available, which we will review in the following.

6.3.1. Alternative definitions

Let us start with two alternative definitions of $C_q(T)$. The first one proposed by Bennett [16] differs only in the error quantity which should go to zero. Instead of the cb-norm the *minimal fidelity* is used. For a channel $T : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ and a subspace $\mathcal{H}' \subset \mathcal{H}$ it is defined as

$$\mathcal{F}_p(\mathcal{H}', T) = \inf_{\psi \in \mathcal{H}'} \langle \psi, T[|\psi\rangle\langle\psi|]\psi\rangle \tag{6.37}$$

and if $\mathcal{H}' = \mathcal{H}$ holds we simply write $\mathcal{F}_p(T)$. Hence a number c is an achievable rate if

$$\lim_{j\to\infty} \mathcal{F}_p(E_j T^{\otimes M_j} D_j) = 1 \tag{6.38}$$

holds for sequences

$$E_j : \mathscr{B}(\mathscr{H})^{\otimes M_j} \to \mathscr{M}_2^{\otimes N_j}, \quad \mathscr{D}_j : \mathscr{M}_2^{\otimes N_j} \to \mathscr{B}(\mathscr{H})^{\otimes M_j}, \quad j \in \mathbb{N} \tag{6.39}$$

of encodings and decodings and sequences of integers M_j, N_j, $j \in \mathbb{N}$ satisfying the same constraints as in Definition 6.1 (in particular $\lim_{j\to\infty} N_j/M_j < c$). The equivalence to our version of $C_q(T)$ follows now from the estimates [168]

$$\|T - \mathrm{Id}\| \leqslant \|T - \mathrm{Id}\|_{\mathrm{cb}} \leqslant 4\sqrt{\|T - \mathrm{Id}\|}\ , \tag{6.40}$$

$$\|T - \mathrm{Id}\| \leqslant 4\sqrt{1 - \mathscr{F}_p(T)} \leqslant 4\sqrt{\|T - \mathrm{Id}\|}\ . \tag{6.41}$$

A second version of $C_q(T)$ is given in [7]. To state it let us define first a *quantum source* as a sequence ρ_N; $N \in \mathbb{N}$ of density operators $\rho_N \in \mathscr{B}^*(\mathscr{K}^{\otimes N})$ (with an appropriate Hilbert space $\mathscr{K}$) and the *entropy rate* of this source as $\limsup_{N\to\infty} S(\rho_N)/N$. In addition we need the *entanglement fidelity* of a state ρ (with respect to a channel T)

$$\mathscr{F}_e(\rho, T) = \langle \Psi, (T \otimes \mathrm{Id})[|\Psi\rangle\langle\Psi|]\Psi\rangle\ , \tag{6.42}$$

where Ψ is the purification of ρ. Now we define $c \geqslant 0$ to be achievable if there is a quantum source ρ_N, $N \in \mathbb{N}$ with entropy rate c such that

$$\lim_{n\to\infty} \mathscr{F}_e(\rho_N, E'_N T^{\otimes N} D'_N) = 1 \tag{6.43}$$

holds with encodings and decodings

$$E'_N : \mathscr{B}(\mathscr{H})^{\otimes N} \to \mathscr{B}(\mathscr{K}^{\otimes N}), \quad \mathscr{D}'_N : \mathscr{B}(\mathscr{K}^{\otimes N}) \to \mathscr{B}(\mathscr{H})^{\otimes N}, \quad j \in \mathbb{N}\ . \tag{6.44}$$

Note that these E'_N, D'_N play a slightly different role than the E_j, D_j in Eq. (6.39) (and in Definition 6.1), because the number of tensor factors of the input and the output algebra is always identical, while in Eq. (6.39) the quotients of these numbers lead to the achievable rate. To relate both definitions we have to derive an appropriately chosen family of subspaces $\mathscr{H}'_N \subset \mathscr{K}^{\otimes N}$ from the ρ_N such that the minimal fidelities $\mathscr{F}_p(\mathscr{H}'_N, E'_N T^{\otimes N} D'_N)$ of *these subspaces* go to 1 as $N \to \infty$. If we identify the $\mathscr{H}'_N$ with tensor products of $\mathbb{C}^2$ and the E_j, D_j of Eq. (6.39) with restrictions of E'_N, D'_N to these tensor products we recover Eq. (6.38). A precise implementation of this rough idea can be found in [6] and it shows that both definitions just discussed are indeed equivalent.

6.3.2. Upper bounds and achievable rates

Although there is no coding theorem for the quantum capacity $C_q(T)$, there is a fairly good candidate which is related to the *coherent information*

$$J(\rho, T) = S(T^*\rho) - S(\rho, T)\ . \tag{6.45}$$

Here $S(T^*\rho)$ is the entropy of the output state and $S(\rho, T)$ is the entropy exchange defined in Eq. (6.26). It is argued [7] that $J(\rho, T)$ plays a role in quantum information theory which is analogous to that of the (classical) mutual information (6.21) in classical information theory. $J(\rho, T)$ has some nasty properties, however: it can be negative [41] and it is known to be not additive [54]. To relate it to $C_{\text{q}}(T)$ it is therefore not sufficient to consider a one-shot capacity as in the Shannons Theorem (Theorem 6.2). Instead, we have to define

$$C_s(T) = \sup_N \frac{1}{N} C_{s,1}(T^{\otimes N}) \quad \text{with } C_{s,1}(T) = \sup_\rho J(\rho, T) \; . \tag{6.46}$$

In [7,8] it is shown that $C_s(T)$ is an upper bound on $C_{\text{q}}(T)$. Equality, however, is conjectured but not yet proven, although there are good heuristic arguments [110,90].

A second interesting quantity which provides an upper bound on the quantum capacity uses the transposition operation Θ on the output systems. More precisely it is shown in [84] that

$$C_{\text{q}}(T) \leqslant C_\theta(T) = \log_2 \|T\Theta\|_{\text{cb}} \tag{6.47}$$

holds for any channel. In contrast to many other calculations in this field it is particular easy to derive this relation from properties of the cb-norm. Hence we are able to give a proof here. We start with the fact that $\|\Theta\|_{\text{cb}} = d$ if d is the dimension of the Hilbert space on which Θ operates. Assume that $N_j/M_j \to c \leqslant C_{\text{q}}(T)$ and j large enough such that $\|\text{Id}_2^{N_j} - E_j T^{\otimes M_j} D_j\| \leqslant \epsilon$ with appropriate encodings and decodings E_j, D_j. We get

$$2^{N_j} = \|\text{Id}_2^{N_j}\Theta\|_{\text{cb}} \leqslant \|\Theta(\text{Id}_2^{N_j} - E_j T^{\otimes M_j} D_j)\|_{\text{cb}} + \|\Theta E_j T^{\otimes M_j} D_j\|_{\text{cb}} \tag{6.48}$$

$$\leqslant 2^{N_j}\|\text{Id}_2^{N_j} - E_j T^{\otimes M_j} D_j\|_{\text{cb}} + \|\Theta E_j \Theta(\Theta T)^{\otimes M_j} D_j\|_{\text{cb}} \tag{6.49}$$

$$\leqslant 2^{N_j}\epsilon + \|\Theta T\|_{\text{cb}}^{M_j} \; , \tag{6.50}$$

where we have used for the last equation the fact that D_j and $\Theta E_j \Theta$ are channels and that the cb-norm is multiplicative. Taking logarithms on both sides we get

$$\frac{N_j}{M_j} + \frac{\log_2(1-\epsilon)}{M_j} \leqslant \log_2 \|\Theta T\|_{\text{cb}} \; . \tag{6.51}$$

In the limit $j \to \infty$ this implies $c \leqslant \log_2 \|\Theta T\|$ and therefore $C_{\text{q}}(T) \leqslant \log_2 \|\Theta T\|_{\text{cb}} = C_\theta(T)$ as stated.

Since $C_\theta(T)$ is an upper bound on $C_{\text{q}}(T)$ it is particularly useful to check whether the quantum capacity for a particular channel is zero. If, e.g., T is classical we have $\Theta T = T$ since the transposition coincides on a classical algebra $\mathscr{C}_d$ with the identity (elements of $\mathscr{C}_d$ are just diagonal matrices). This implies $C_\theta(T) = \log_2 \|\Theta T\|_{\text{cb}} = \log_2 \|T\|_{\text{cb}} = 0$, because the cb-norm of a channel is 1. We see therefore that the quantum capacity of a classical channel is 0—this is just another proof of the no-teleportation theorem. A slightly more general result concerns channels $T = RS$ which are the composition of a preparation $R : \mathscr{M}_d \to \mathscr{C}_f$ and a subsequent measurement $S : \mathscr{C}_f \to \mathscr{M}_d$. It is easy to see that $\Theta T = \Theta RS$ is a channel, because $\Theta R \Theta$ is a channel and Θ is the identity on $\mathscr{C}_f$, hence $\Theta R \Theta = \Theta R$ and $\Theta R \Theta S = \Theta RS = \Theta T$. Again we get $C_\theta(T) = 0$.

Let us consider now some examples. The most simple case is again the quantum erasure channel from Eq. (6.29). As for the classical capacities its quantum capacity can be explicitly calculated [15] and we have $C_{\text{q}}(T) = \max(0, (1 - 2\vartheta)\log_2(d))$; cf. Fig. 6.1.

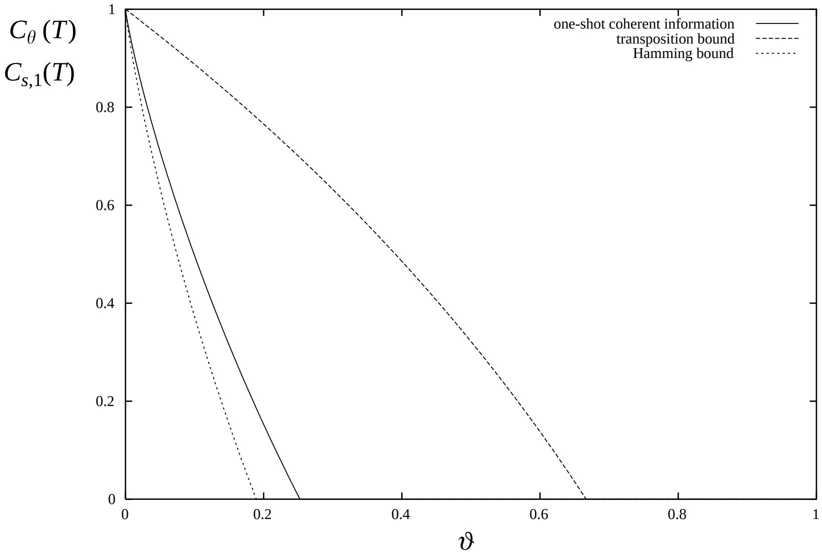


Fig. 6.6. $C_\theta(T)$, $C_s(T)$ and the Hamming bound of a depolarizing qubit channel plotted as function of the noise parameter ϑ.

For the depolarizing channel (6.30) precise calculations of $C_q(T)$ are not available. Hence let us consider first the coherent information. $J(T,\rho)$ inherits from T its unitary covariance, i.e. we have $J(U\rho U^*,T)=J(\rho,T)$. In contrast to the mutual information, however, it does not have nice concavity properties, which makes the optimization over all input states more difficult to solve. Nevertheless, the calculation of $J(\rho,T)$ is straightforward and we get in the qubit case (if ϑ is the noise parameter of T and λ is the highest eigenvalue of ρ):

$$J(\rho,T)=S\left(\lambda(1-\vartheta)+\frac{\vartheta}{2}\right)-S\left(\frac{1-\vartheta/2+A}{2}\right)-S\left(\frac{1-\vartheta/2-A}{2}\right)$$
$$-S\left(\frac{\lambda\vartheta}{2}\right)-S\left(\frac{(1-\lambda)\vartheta}{2}\right), \tag{6.52}$$

where $S(x)=-x\log_2(x)$ denotes again the entropy function and

$$A=\sqrt{(2\lambda-1)^2(1-\vartheta/2)^2+4\lambda(1-\lambda)(1-\vartheta)^2}\,. \tag{6.53}$$

Optimization over λ can be performed at least numerically (the maximum is attained at the left boundary ($\lambda=1/2$) if J is positive there, and the right boundary otherwise). The result is plotted together with $C_\theta(T)$ in Fig. 6.6 as a function of θ. The quantity $C_\theta(T)$ is much easier to compute and we get

$$C_\theta(T)=\max\left\{0,\,\log_2\left(2-\frac{3}{2}\,\theta\right)\right\}\,. \tag{6.54}$$

To get a lower bound on $C_q(T)$ we have to show that a certain rate $r\leqslant C_q(T)$ can be achieved with an appropriate sequence

$$E_M:\mathscr{M}_d^{\otimes M}\to\mathscr{M}_2^{\otimes N(M)},\quad M,N(M)\in\mathbb{N} \tag{6.55}$$

of error correcting codes and corresponding decodings D_M. I.e. we need

$$\lim_{j\to\infty} N(M)/M = r \quad \text{and} \quad \lim_{j\to\infty} \|E_M T^{\otimes M} D_M - \mathrm{Id}\|_{\mathrm{cb}} = 0 \;. \tag{6.56}$$

To find such a sequence note first that we can look at the depolarizing channel as a device which produces an error with probability ϑ and leaves the quantum information intact otherwise. If more and more copies of T are used in parallel, i.e. if M goes to infinity, the number of errors approaches therefore ϑM. In other words, the probability to have more than ϑM errors vanishes asymptotically. To see this consider

$$T^{\otimes M} = ((\vartheta - 1)\mathrm{Id} + \vartheta d^{-1}\,\mathrm{tr}(\cdot)\mathbb{1})^{\otimes M} = \sum_{K=1}^{M} (1-\vartheta)^K \vartheta^{N-K} T_K^{(M)} \;, \tag{6.57}$$

where $T_K^{(M)}$ denotes the sum of all M-fold tensor products with $d^{-1}\,\mathrm{tr}(\cdot)\mathbb{1}$ on N places and Id on the $N-K$ remaining—i.e. $T_K^{(M)}$ is a channel which produces exactly K errors on M transmitted systems. Now we have

$$\left\| T^{\otimes M} - \sum_{K\leqslant \vartheta M} (1-\vartheta)^K \vartheta^{N-K} T_K^{(M)} \right\|_{\mathrm{cb}} \tag{6.58}$$

$$= \left\| \sum_{K>\vartheta M} (1-\vartheta)^K \vartheta^{N-K} T_K^{(M)} \right\|_{\mathrm{cb}} \tag{6.59}$$

$$\leqslant \sum_{K>\vartheta M}^{M} (1-\vartheta)^K \vartheta^{N-K} \|T_K^{(M)}\|_{\mathrm{cb}} \tag{6.60}$$

$$\leqslant \sum_{K>\vartheta M}^{M} \binom{M}{K} (1-\vartheta)^K \vartheta^{N-K} = R \;. \tag{6.61}$$

The quantity R is the tail a of binomial series and vanishes therefore in the limit $M \to \infty$ (cf. e.g. [131, Appendix B]). This shows that for $M \to \infty$ only terms $T_K^{(M)}$ with $K \leqslant \vartheta M$ are relevant in Eq. (6.57)—in other words at most ϑM errors occur asymptotically, as stated. This implies that we need a sequence of codes E_M which encode $N(M)$ qubits and correct ϑM errors on M places. One way to get such a sequence is "random coding"—the classical version of this method is well known from the proof of Shannons theorem. The idea is, basically, to generate error correcting codes of a certain type randomly. E.g. we can generate a sequence of random graphs with $N(M)$ input and M output vertices (cf. Section 4.4). If we can show that the corresponding codes correct (asymptotically) ϑM errors, the corresponding rate $r = \lim_{M\to\infty} N(M)/M$ is achievable. For the depolarizing channel[21] such an analysis, using randomly generated stabilizer codes shows [16,71]

$$C_{\mathrm{q}}(T) \leqslant 1 - H(\vartheta) - \vartheta \log_2 3 \;, \tag{6.62}$$

[21] With a more thorough discussion similar results can be obtained for a much more general class of channels, e.g. all T in a neighborhood of the identity channel; cf. [114].

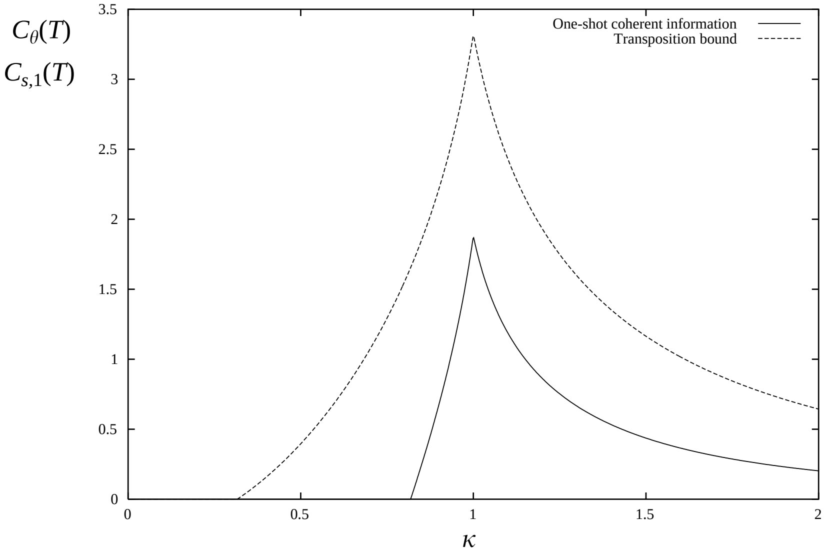


Fig. 6.7. $C_\theta(T)$ and $C_s(T)$ of a Gaussian amplification/attenuation channel as a function of amplification parameter k.

where H is the binary entropy from Eq. (5.16). This bound can be further improved using a more clever coding strategy; cf. [54].

As a third example let us consider again the Gaussian channel studied already in Section 6.2.4. For $C_\theta(T)$ we have (the corresponding calculation is not trivial and uses properties of Gaussian channels which we have not discussed; cf. [84].)

$$C_\theta(T) = \max\{0, \log_2(k^2+1) - \log_2(|k^2-1| + 2N_c)\} \tag{6.63}$$

and we see that $C_\theta(T)$ and therefore $C_q(T)$ become zero if N_c is large enough (i.e. $N_c \geqslant \max\{1, k^2\}$). The coherent information for the Gaussian state ρ_N from Eq. (3.64) has the form

$$J(\rho_N, T) = g(N') - g\left(\frac{D + N' - N - 1}{2}\right) - g\left(\frac{D - N' + N - 1}{2}\right) \tag{6.64}$$

with N', D and g as in Section 6.2.4. It increases with N and we can calculate therefore the maximum over all Gaussian states (which might differ from $C_S(T)$) as

$$C_G(T) = \lim_{N\to\infty} J(\rho_N, T) = \log_2 k^2 - \log_2 |k^2 - 1| - g\left(\frac{N_c}{k^2-1}\right) . \tag{6.65}$$

We have plotted both quantities in Fig. 6.7 as a function of k.

Finally let us have a short look on the special case $k = 1$, i.e. T describes in this case only the influence of classical Gaussian noise on the transmitted qubits. If we set $k = 1$ in Eq. (6.64) and take the limit $N \to \infty$ we get $C_G(T) = -\log_2(N_c e)$ and $C_\theta(T)$ becomes $C_\theta(T) = \max\{0, -\log_2(N_c)\}$; both quantities are plotted in Fig. 6.8. This special case is interesting because the one-shot coherent

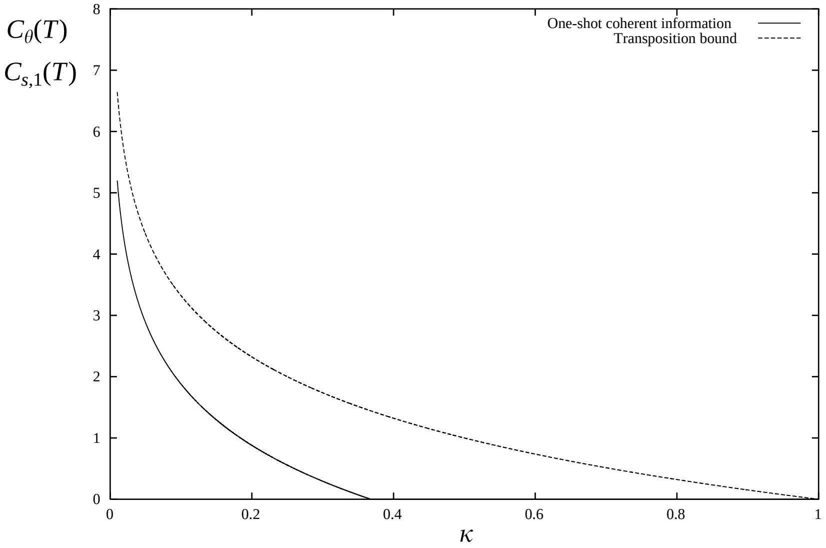


Fig. 6.8. $C_\theta(T)$ and $C_s(T)$ of a Gaussian amplification/attenuation channel as a function of the noise parameter N_c (and with $k=1$).

information $C_G(T)$ is achievable, provided the noise parameter N_c satisfies certain conditions[22] [77]. Hence there is strong evidence that the quantum capacity lies between the two lines in Fig. 6.8.

6.3.3. Relations to entanglement measures

The duality lemma proved in Section 2.3.3 provides an interesting way to derive bounds on channel capacities and capacity-like quantities from entanglement measures (and vice versa) [16,90]: To derive a state of a bipartite system from a channel T we can take a maximally entangled state $\Psi \in \mathscr{H} \otimes \mathscr{H}$, send one particle through T and get a less entangled pair in the state $\rho_T = (\mathrm{Id} \otimes T^*)|\Psi\rangle\langle\Psi|$. If on the other hand an entangled state $\rho \in \mathscr{S}(\mathscr{H} \otimes \mathscr{H})$ is given, we can use it as a resource for teleportation and get a channel T_ρ. The two maps $\rho \mapsto T_\rho$ and $T \mapsto \rho_T$ are, however, not inverse to one another. This can be seen easily from the duality lemma (Theorem 2.10): For each state $\rho \in \mathscr{S}(\mathscr{H} \otimes \mathscr{H})$ there is a channel T and a pure state $\Phi \in \mathscr{H} \otimes \mathscr{H}$ such that $\rho = (\mathrm{Id} \otimes T^*)|\Phi\rangle\langle\Phi|$ holds; but Φ is in general not maximally entangled (and uniquely determined by ρ). Nevertheless, there are special cases in which the state derived from T_ρ coincides with ρ: A particular class of examples is given by teleportation channels derived from a Bell-diagonal state.

On ρ_T we can evaluate an entanglement measure $E(\rho_T)$ and get in this way a quantity which is related to the capacity of T. A particularly interesting candidate for E is the "one-way LOCC" distillation rate $E_{D,\to}$. It is defined in the same way as the entanglement of distillation E_D, except that only one-way LOCC operation are allowed in Eq. (5.8). According to [16] $E_{D,\to}$ is related to C_q by the inequalities $E_{D,\to}(\rho) \geqslant C_q(T_\rho)$ and $E_{D,\to}(T_\rho) \leqslant C_q(T)$. Hence if $\rho_{T_\rho} = \rho$ we can calculate $E_{D,\to}(\rho)$ in terms of $C_q(T_\rho)$ and vice versa.

[22] It is only shown that $\log_2(\lfloor 1/(N_c e) \rfloor)$ can be achieved, where $\lfloor x \rfloor$ denotes the biggest integer less than x. It is very likely however that this is only a restriction of the methods used in the proof and not of the result.

A second interesting example is the transposition bound $C_\theta(T)$ introduced in the last subsection. It is related to the *logarithmic negativity* [158]

$$E_\theta(\rho_T) = \log_2 \|(\mathrm{Id} \otimes \Theta)\rho_T\|_1 \ , \tag{6.66}$$

which measures the degree with which the partial transpose of ρ fails to be positive. E_θ can be regarded as entanglement measure although it has some drawbacks: it is not LOCC monotone (Axiom E2), it is not convex (Axiom E3) and most severe: It does not coincides with the reduced von Neumann entropy on pure states, which we have considered as "*the*" entanglement measure for pure states. On the other hand, it is easy to calculate and it gives bounds on distillation rates and teleportation capacities [158]. In addition E_θ can be used together with the relation between depolarizing channels and isotropic states to derive Eq. (6.54) in a very simple way.

7. Multiple inputs

We have seen in Section 4 that many tasks of quantum information which are impossible with one-shot operations can be approximated by channels which operate on a large number of equally prepared inputs. Typical examples are approximate cloning, undoing noise and distillation of entanglement. There are basically two questions which are interesting for a quantitative analysis: First, we can search for the optimal solutions for a fixed number N of input systems and second we can ask for the asymptotic behavior in the limit $N \to \infty$. In the latter case the asymptotic rate, i.e. the number of outputs (of a certain quality) per input system is of particular interest.

7.1. The general scheme

Both types of questions just mentioned can be treated (up to certain degree) independently from the (impossible) task we are dealing with. In the following we will study the corresponding general scheme. Hence consider a channel $T : \mathscr{B}(\mathscr{H}^{\otimes M}) \to \mathscr{B}(\mathscr{H}^{\otimes N})$ which operates on N input systems and produces M outputs of the same type. Our aim is to optimize a "*figure of merit*" $\mathscr{F}(T)$ which measures the deviation of $T^*(\rho^{\otimes N})$ from the target functional we want to approximate. The particular type of device we are considering is mainly fixed by the choice of $\mathscr{F}(T)$ and we will discuss in the following the most relevant examples. (Note that we have considered them already on a qualitative level in Section 4; cf. in particular Sections 4.2 and 4.3.)

7.1.1. Figures of merit

Let us start with pure state cloning [68,31,32,35,166,98], i.e. for each (unknown) pure input state $\sigma = |\psi\rangle\langle\psi|$, $\psi \in \mathscr{H}$ the M clones $T^*(\sigma^{\otimes N})$ produced by the channel T should approximate M copies of the input in the common state $\sigma^{\otimes M}$ as good as possible. There are in fact two different possibilities to measure the distance of $T^*(\sigma^{\otimes N})$ to $\sigma^{\otimes M}$. We can either check the quality of each clone separately or we can test in addition the correlations between output systems. With the notation

$$\sigma^{(j)} = \mathbb{1}^{\otimes(j-1)} \otimes \sigma \otimes \mathbb{1}^{\otimes(M-j)} \in \mathscr{B}(\mathscr{H}^{\otimes M}) \tag{7.1}$$

a figure of merit for the first case is given by

$$\mathscr{F}_{\mathrm{c},1}(T) = \inf_{j=1,\dots,N} \inf_{\sigma \text{ pure}} \operatorname{tr}(\sigma^{(j)} T^*(\sigma^{\otimes N})) . \tag{7.2}$$

It measures the worst one-particle fidelity of the output state $T^*(\sigma^{\otimes N})$. If we are interested in correlations too, we have to choose

$$\mathscr{F}_{\mathrm{c,all}}(T) = \inf_{\sigma \text{ pure}} \operatorname{tr}(\sigma^{\otimes M} T^*(\sigma^{\otimes N})) , \tag{7.3}$$

which is again a "worst case" fidelity, but now of the full output with respect to M uncorrelated copies of the input σ.

Instead of fidelities we can consider other error quantities like trace-norm distances or relative entropies. In general, however, we do not get significantly different results from such alternative choices; hence, we can safely ignore them. Real variants arise if we consider instead of the infima over all pure states quantities which prefer a (possibly discrete or even finite) class of states. Such a choice leads to "state-dependent cloning", because the corresponding optimal devices perform better as "universal" ones (i.e. those described by the figures of merit above) on *some states* but much worse on the rest. We ignore state-dependent cloning in this work, because the universal case is physically more relevant and technically more challenging. Other cases which we do not discuss either include "asymmetric cloning", which arises if we trade in Eq. (7.2) the quality of one particular output system against the rest (see [40]), and cloning of mixed states. The latter is much more difficult than the pure state case and even for classical systems, where it is related to the so-called "bootstrap" technique [59], non-trivial.

Closely related to cloning is purification, i.e. undoing noise. This means we are considering N systems originally prepared in the same (unknown) pure state σ but which have passed a depolarizing channel

$$R^*\sigma = \vartheta\sigma + (1-\vartheta)\mathbb{1}/d \tag{7.4}$$

afterwards. The task is now to find a device T acting on N of the decohered systems such that $T^*(R^*\sigma)$ is as close as possible to the original pure state. We have the same basic choices for a figure of merit as in the cloning problem. Hence, we define

$$\mathscr{F}_{R,1}(T) = \inf_{j=1,\dots,N} \inf_{\sigma \text{ pure}} \operatorname{tr}(\sigma^{(j)} T^*[(R^*\sigma)^{\otimes N}]) \tag{7.5}$$

and

$$\mathscr{F}_{R,\mathrm{all}}(T) = \inf_{\sigma \text{ pure}} \operatorname{tr}(\sigma^{\otimes M} T^*[(R^*\sigma)^{\otimes N}]) . \tag{7.6}$$

These quantities can be regarded as generalizations of $\mathscr{F}_{\mathrm{c},1}$ and $\mathscr{F}_{\mathrm{c,all}}$ which we recover if R^* is the identity.

Another task we can consider is the approximation of a map Θ which is positive but not completely positive, like the transposition. Positivity and normalization imply that Θ^* maps states to states but Θ cannot be realized by a physical device. An explicit example is the universal not gate (UNOT) which maps each pure qubit state σ to its orthocomplement $\sigma^{\perp}$ [36]. It is given the anti-unitary operator

$$\psi = \alpha|0\rangle + \beta|1\rangle \mapsto \Theta\psi = \bar{\alpha}|0\rangle - \bar{\beta}|1\rangle . \tag{7.7}$$

Since $\Theta\sigma$ is a state if σ is, we can ask again for a channel T such that $T^*(\sigma^{\otimes N})$ approximates $(\Theta\sigma)^{\otimes M}$. As in the two previous examples we have the choice to allow arbitrary correlations in the output or not and we get the following figures of merit:

$$\mathscr{F}_{\theta,1}(T) = \inf_{j=1,\dots,N} \inf_{\sigma\,\text{pure}} \text{tr}((\Theta\sigma)^{(j)} T^*(\sigma^{\otimes N})) \tag{7.8}$$

and

$$\mathscr{F}_{\theta,\text{all}}(T) = \inf_{\sigma\,\text{pure}} \text{tr}((\Theta\sigma)^{\otimes M} T^*(\sigma^{\otimes N})) \ . \tag{7.9}$$

Note that we can plug in for Θ basically any functional which maps states to states. In addition we can combine Eqs. (7.5) and (7.6) on the one hand with (7.8) and (7.9) on the other. As result we would get a measure for devices which undo an operation R and approximate an impossible machine Θ at the same time.

7.1.2. Covariant operations

All the functionals just defined give rise to optimization problems which we will study in greater detail in the next sections. This means we are interested in two things: First of all the maximal value of $\mathscr{F}_{\#,\natural}$ (with $\# = c, R, \theta$ and $\natural = 1, \text{all}$) given by

$$\mathscr{F}_{\#,\natural}(N, M) = \inf_{T} \mathscr{F}_{\#,\natural}(T) \ , \tag{7.10}$$

where the supremum is taken over all channels $T : \mathscr{B}(\mathscr{H}^{\otimes M}) \to \mathscr{B}(\mathscr{H}^{\otimes N})$, and second the particular channel $\hat{T}$ where the optimum is attained. At a first look a complete solution of these problems seems to be impossible, due to the large dimension of the space of all T, which scales exponentially in M and N. Fortunately, all $\mathscr{F}_{\#,\natural}(T)$ admit a large symmetry group which allows in many cases the explicit calculation of the optimal values $\mathscr{F}_{\#,\natural}(N, M)$ and the determination of optimizers $\hat{T}$ with a certain covariance behavior. Note that this is an immediate consequence of our decision to restrict the discussion to "universal" procedures, which do not prefer any particular input state.

Let us consider permutations of the input systems first: If $p \in S_N$ is a permutation on N places and V_p the corresponding unitary on $\mathscr{H}^{\otimes N}$ (cf. Eq. (3.7)) we get obviously $T^*(V_p \rho^{\otimes N} V_p^*) = T^*(\rho^{\otimes N})$, hence

$$\mathscr{F}_{\#,\natural}[\alpha_p(T)] = \mathscr{F}_{\#,\natural}(T) \quad \forall p \in S_N \quad \text{with } [\alpha_p(T)](A) = V_p^* T(A) V_p \ . \tag{7.11}$$

In other words: $\mathscr{F}_{\#,\natural}(T)$ is invariant under permutations of the input systems. Similarly, we can show that $\mathscr{F}_{\#,\natural}(T)$ is invariant under permutations of the output systems:

$$\mathscr{F}_{\#,\natural}[\beta_p(T)] = \mathscr{F}(T) \quad \forall p \in S_M \quad \text{with } [\beta_p(T)](A) = T(V_p^* A V_p) \ . \tag{7.12}$$

To see this consider e.g. for $\# = c$ and $\natural = \text{all}$

$$\text{tr}[\sigma^{\otimes M} V_p T^*(\rho^{\otimes N}) V_p^*] = \text{tr}[V_p \sigma^{\otimes M} V_p^* T^*(\rho^{\otimes N})] = \text{tr}[\sigma^{\otimes M} T^*(\rho^{\otimes N})] \ . \tag{7.13}$$

For the other cases similar calculations apply.

Finally, none of the $\mathscr{F}_{\#,\natural}(T)$ singles out a preferred direction in the one-particle Hilbert space $\mathscr{H}$. This implies that we can rotate T by local unitaries of the form $U^{\otimes N}$, respectively $U^{\otimes M}$, without changing $\mathscr{F}_{\#,\natural}(T)$. More precisely we have

$$\mathscr{F}_{\#,\natural}[\gamma_U(T)] = \mathscr{F}_{\#,\natural}(T) \quad \forall U \in U(d) \tag{7.14}$$

with

$$[\gamma_U(T)](A) = U^{*\otimes N} T(U^{\otimes M} A U^{*\otimes M}) U^{\otimes N} \; . \tag{7.15}$$

The validity of Eq. (7.14) can be proven in the same way as (7.11) and (7.12). The details are therefore left to the reader.

Now we can average over the groups S_N, S_M and $U(d)$. Instead of the operation T we consider

$$\bar{T} = \frac{1}{N!M!} \sum_{p\in S_N} \sum_{q\in S_M} \int_G \alpha_p \beta_q \gamma_U(T)\, \mathrm{d}U \; , \tag{7.16}$$

where $\mathrm{d}U$ denotes the normalized, left invariant Haar measure on $U(d)$. We see immediately that $\bar{T}$ has the following symmetry properties:

$$\alpha_p(\bar{T}) = \bar{T}, \quad \beta_q(\bar{T}) = \bar{T}, \quad \gamma_U(\bar{T}) = \bar{T} \quad \forall p \in S_N \quad \forall q \in S_M \quad \forall U \in U(d) \tag{7.17}$$

and we will call each operation T *fully symmetric*, if it satisfies this equation. The concavity of $\mathscr{F}_{\#,\natural}$ implies immediately that it cannot decrease if we replace T by $\bar{T}$:

$$\mathscr{F}_{\#,\natural}(T) = \mathscr{F}_{\#,\natural}\left(\frac{1}{N!M!} \sum_{p\in S_N} \sum_{q\in S_M} \int_G \alpha_p \beta_q \gamma_U(T)\, \mathrm{d}U\right) \tag{7.18}$$

$$\geqslant \frac{1}{N!M!} \sum_{p\in S_N} \sum_{q\in S_M} \int_G \mathscr{F}_{\#,\natural}[\alpha_p \beta_q \gamma_U(T)]\, \mathrm{d}U = \mathscr{F}_{\#,\natural}(T) \; . \tag{7.19}$$

To calculate the optimal value $\mathscr{F}_{\#,\natural}(N,M)$ it is therefore completely sufficient to search a maximizer for $\mathscr{F}_{\#,\natural}(T)$ only among fully symmetric T and to evaluate $\mathscr{F}_{\#,\natural}(T)$ for this particular operation. This simplifies the problem significantly because the size of the parameter space is extremely reduced. Of course, we do not know from this argument whether the optimum is attained on non-symmetric operations, however this information is in general less important (and for some problems like optimal cloning a uniqueness result is available).

7.1.3. Group representations

To get an idea how this parameter reduction can be exploited practically, let us reconsider Theorem 3.1: The two representations $U \mapsto U^{\otimes N}$ and $p \mapsto V_p$ of $U(d)$, respectively S_N, on $\mathscr{H}^{\otimes N}$ are "commutants" of each other, i.e., any operator on $\mathscr{H}^{\otimes N}$ commuting with all $U^{\otimes N}$ is a linear combination of the V_p, and conversely. This knowledge can be used to decompose the representation $U^{\otimes N}$ (and V_p as well) into irreducible components. To reduce the group theoretic overhead, we will discuss this procedure first for qubits only and come back to the general case afterwards.

Hence assume that $\mathscr{H} = \mathbb{C}^2$ holds. Then $\mathscr{H}^{\otimes N}$ is the Hilbert space of N (distinguishable) spin-1/2 particles and it can be decomposed into terms of eigenspaces of total angular momentum. More precisely consider

$$L_k = \frac{1}{2} \sum_j \sigma_k^{(j)}, \quad k = 1,2,3 \tag{7.20}$$

the k-component of total angular momentum (i.e. σ_k is the kth Pauli matrix and $\sigma^{(j)} \in \mathscr{B}(\mathscr{H}^{\otimes N})$ is defined according to Eq. (7.1)) and $\vec{L}^2 = \sum_k L_k^2$. The eigenvalue expansion of $\vec{L}^2$ is well known to be

$$\vec{L} = \sum_j s(s+1)P_s \quad \text{with } s = \begin{cases} 0, 1, \dots, N/2, & N \text{ even}, \\ 1/2, 3/2, \dots, N/2, & N \text{ odd}, \end{cases} \tag{7.21}$$

where the P_s denote the projections to the eigenspaces of $\vec{L}^2$. It is easy to see that both representations $U \mapsto U^{\otimes N}$ and $p \mapsto V_p$ commute with $\vec{L}$. Hence the eigenspaces $P_s\mathscr{H}^{\otimes N}$ of $\vec{L}^2$ are invariant subspaces of $U^{\otimes N}$ and V_p and this implies that the restriction of $U^{\otimes N}$ and V_p to them are representations of SU(2), respectively S_N. Since $\vec{L}^2$ is constant on $P_s\mathscr{H}^{\otimes N}$ the SU(2) representation we get in this way must be (naturally isomorphic to) a multiple of the irreducible spin-s representation π_s. It is defined by

$$\pi_s \left[\exp\left(\frac{\mathrm{i}}{2} \sigma_k \right) \right] = \exp\left(\mathrm{i} L_k^{(s)} \right) \quad \text{with } L_k^{(s)} = \frac{1}{2} \sum_{j=1}^{2s} \sigma_k^{(j)} , \tag{7.22}$$

on the representation space

$$\mathscr{H}_s = \mathscr{H}_+^{\otimes 2s} \tag{7.23}$$

(the Bose-subspace of $\mathscr{H}^{\otimes 2s}$). Hence we get

$$P_s\mathscr{H}^{\otimes N} \cong \mathscr{H}_s \otimes \mathscr{K}_{N,s}, \quad U^{\otimes N}\psi = (\pi_s(U) \otimes \mathbb{1})\psi \quad \forall \psi \in P_s\mathscr{H}^{\otimes N}. \tag{7.24}$$

Since V_p and $U^{\otimes N}$ commute the Hilbert space $\mathscr{K}_{N,s}$ carries a representation $\hat{\pi}_{N,s}(p)$ of S_N which is irreducible as well. Note that $\mathscr{K}_{N,s}$ depends in contrast to $\mathscr{H}_s$ on the number N of tensor factors and its dimension is (see [100] or [142] for general d)

$$\dim \mathscr{K}_{N,s} = \frac{2s+1}{N/2+s+1} \begin{pmatrix} N \\ N/2 - s \end{pmatrix}. \tag{7.25}$$

Summarizing the discussion we get

$$\mathscr{H}^{\otimes N} \cong \bigoplus_s \mathscr{H}_s \otimes \mathscr{K}_{N,s}, \quad U^{\otimes N} \cong \bigoplus_s \pi_s(U) \otimes \mathbb{1}, \quad V_p \cong \bigoplus_s \mathbb{1} \otimes \hat{\pi}(p) . \tag{7.26}$$

Let us consider now a fully symmetric operation T. Permutation invariance ($\alpha_p(T) = T$ and $\beta_p(T) = T$) implies together with Eq. (7.26) that

$$T(A_j \otimes B_j) = \bigoplus_s \left[\frac{\mathrm{tr}\,(B_j)}{\dim \mathscr{K}_{N,j}} T_{sj}(A_j) \otimes \mathbb{1} \right] \quad \text{with } T_{sj} : \mathscr{B}(\mathscr{H}_j) \to \mathscr{B}(\mathscr{H}_s) \tag{7.27}$$

holds if $A_j \otimes B_j \in \mathscr{B}(\mathscr{H}_j \otimes \mathscr{K}_{N,j})$. The operations T_{sj} are unital and have, according to $\gamma_U(T) = T$ the following covariance properties:

$$\pi_s(U)T(A_j)\pi_s(U^*) = T[\pi_j(U)A_j\pi_j(U^*)] \quad \forall U \in \mathrm{SU}(2) \ . \tag{7.28}$$

The classification of all fully symmetric channels T is reduced therefore to the study of all these T_{sj}.

We can apply now the covariant version of Stinespring's theorem (Theorem 3.3) to find that

$$T_{sj}(A_j) = V^*(A_j \otimes \mathbb{1})V, \quad V : \mathscr{H}_s \to \mathscr{H}_j \otimes \tilde{\mathscr{H}}, \quad V\pi_s(U) = \pi_j(U) \otimes \tilde{\pi}(U)V \ , \tag{7.29}$$

where $\tilde{\pi}$ is a representation of SU(2) on $\tilde{\mathscr{H}}$. If $\tilde{\pi}$ is irreducible with total angular momentum l the "intertwining operator" V is well known: Its components in a particularly chosen basis coincide with certain Clebsh–Gordon coefficients. Hence, the corresponding operation is uniquely determined (up to unitary equivalence) and we write

$$T_{sjl}(A_j) = [V_l(A_j \otimes \mathbb{1})V_l], \quad V_l\pi_s(U) = \pi_j(U) \otimes \pi_l(U)V_l \ , \tag{7.30}$$

where l can range from $|j - s|$ to $j + s$. Since in a general representation $\tilde{\pi}$ can be decomposed into irreducible components we see that each covariant T_{sj} is a convex linear combination of the T_{sjl} and we get with Eq. (7.27)

$$T(A_j \otimes B_j) = \bigoplus_s \left[\sum_l c_{jl}[T_{sjl}(A_j) \otimes (\mathrm{tr}\,(B_j)\mathbb{1})] \right] \ , \tag{7.31}$$

where the c_{jl} are constrained by $c_{jl} > 0$ and $\sum_j c_{jl} = (\dim \mathscr{K}_{N,j})^{-1}$. In this way we have parameterized the set of fully symmetric operations completely in terms of group theoretical data and we can rewrite $\mathscr{F}_{\#,\natural}(T)$ accordingly. This leads to an optimization problem for a quantity depending only on s, j and l, which is at least in some cases solvable.

To generalize the scheme just presented to the case $\mathscr{H} = \mathbb{C}^d$ with arbitrary d we only have to find a replacement for the decomposition in Eq. (7.26). This, however, is well known from group theory:

$$\mathscr{H}^{\otimes N} \cong \bigoplus_Y \mathscr{H}_Y \otimes \mathscr{K}_Y, \quad U^{\otimes N} \cong \bigoplus_Y \pi_Y(U) \otimes \mathbb{1}, \quad V_p \cong \bigoplus_Y \mathbb{1} \otimes \hat{\pi}_Y(p) \ , \tag{7.32}$$

where $\pi_Y : U(d) \to \mathscr{B}(\mathscr{H}_Y)$ and $\hat{\pi}_Y : S_N \to \mathscr{B}(\mathscr{K}_Y)$ are irreducible representations. The summation index Y runs over all Young frames with d rows and N boxes, i.e. by the arrangements of N boxes into d rows of lengths $Y_1 \geqslant Y_2 \geqslant \cdots \geqslant Y_d \geqslant 0$ with $\sum_k Y_k = N$. The relation to total angular momentum s used as the parameter for $d = 2$ is given by $Y_1 - Y_2 = 2s$, which determines Y together with $Y_1 + Y_2 = N$ completely. The rest of the arguments applies without significant changes, this is in particular the case for Eq. (7.31) which holds for general d if we replace s, j and l by Young frames. However, the representation theory of $U(d)$ becomes much more difficult. The generalization of results available for qubits ($d = 2$) to $d > 2$ is therefore not straightforward.

Finally, let us give a short comment on Gaussian states here. Obviously, the methods just described do not apply in this case. However, we can consider instead of $U^{\otimes N}$-covariance, covariance with respect to phase-space translations. Following this idea some results concerning optimal cloning of Gaussian states are obtained (see [43] and the references therein), but the corresponding general theory is not as far developed as in the finite-dimensional case.

7.1.4. Distillation of entanglement

Finally, let us have another look at distillation of entanglement. The basic idea is quite the same as for optimal cloning: Use multiple inputs to approximate a task which is impossible with one-shot operations. From a more technical point of view, however, it does not fit into the general scheme proposed up to now. Nevertheless, some of the arguments can be adopted in an easy way. First of all we have to replace the "one-particle" Hilbert space $\mathscr{H}$ with a twofold tensor product $\mathscr{H}_A \otimes \mathscr{H}_B$ and the channels we have to look at are LOCC operations

$$T : \mathscr{B}(\mathscr{H}_A^{\otimes M} \otimes \mathscr{H}_B^{\otimes M}) \to \mathscr{B}(\mathscr{H}_A^{\otimes N} \otimes \mathscr{H}_B^{\otimes N}) ; \tag{7.33}$$

cf. Section 4.3. Our aim is to determine T such that $T^*(\rho^{\otimes N})$ is for each distillable (mixed) state $\rho \in \mathscr{B}^*(\mathscr{H}_A \otimes \mathscr{H}_B)$, close to the M-fold tensor product $|\Psi\rangle\langle\Psi|^{\otimes M}$ of a maximally entangled state $\Psi \in \mathscr{H}_A \otimes \mathscr{H}_B$. A figure of merit with a similar structure as the $\mathscr{F}_{\#,\mathrm{all}}$ studied above can be derived directly from the definition of the entanglement measure E_D in Section 5.1.3: We define (replacing the trace-norm distance with a fidelity)

$$\mathscr{F}_\mathrm{D}(T) = \inf_\rho \inf_\Psi \langle \Psi^{\otimes M}, T^*(\rho^{\otimes N})\Psi^{\otimes M}\rangle , \tag{7.34}$$

where the infima are taken over all maximally entangled states Ψ and all distillable states ρ. Alternatively, we can look at state-dependent measures, which seem to be particularly important if we try to calculate $E_\mathrm{D}(\rho)$ for some state ρ. In this case we simply get

$$\mathscr{F}_{\mathrm{D},\rho}(T) = \inf_\Psi \langle \Psi^{\otimes M}, T^*(\rho^{\otimes N})\Psi^{\otimes M}\rangle . \tag{7.35}$$

To translate the group theoretical analysis of the last two subsections is somewhat more difficult. As in the case of $\mathscr{F}_{\#,\natural}$ we can restrict the search for optimizers to permutation invariant operations, i.e. $\alpha_p(T) = T$ and $\beta_p(T) = T$ in the terminology of Section 7.1.2. Unitary covariance

$$U^{\otimes N} T(A) U^{*\otimes N} = T(U^{\otimes M} A U^{*\otimes M}) , \tag{7.36}$$

however, cannot be assumed *for all* unitaries U of $\mathscr{H}_A \otimes \mathscr{H}_B$, but only for local ones ($U = U_A \otimes U_B$) in the case of $\mathscr{F}_\mathrm{D}$ or only for local U which leave ρ invariant for $\mathscr{F}_{\mathrm{D},\rho}$. This makes the analog of the decomposition scheme from Section 7.1.3 more difficult and such a study is (up to my knowledge) not yet done. A related subproblem arises if we consider $\mathscr{F}_{\mathrm{D},\rho}$ from Eq. (7.35) for a state ρ with special symmetry properties; e.g. an OO-invariant state. The corresponding optimization might be simpler and a solution would be relevant for the calculation of E_D.

7.2. Optimal devices

Now we can consider the optimization problems associated to the figures of merit discussed in the last section. This means that we are searching for those devices which approximate the impossible tasks in question in the best possible way. As pointed out at the beginning of this Section this can be done for finite N and in the limit $N \to \infty$. The latter is postponed to the next section.

7.2.1. Optimal cloning

The quality of an optimal, pure state cloner is defined by the figures of merit $\mathscr{F}_{\mathrm{c},\#}$ in Eqs. (7.2) and (7.3) and the group theoretic ideas sketched in Section 7.1.3 allow the complete solution of this problem. We will demonstrate some of the basic ideas in the qubit case first and state the final result afterwards in full generality.

The solvability of this problem relies in part on the special structure of the figures of merit $\mathscr{F}_{c,\#}$, which allows further simplifications of the general scheme sketched in Section 7.1.3. If we consider e.g. $\mathscr{F}_{c,1}(T)$ (the other case works similarly) we get

$$\mathscr{F}_{c,1}(T) = \inf_{j=1,\dots,N} \inf_{\sigma \text{ pure}} \operatorname{tr}(\sigma^{(j)} T^*(\sigma^{\otimes N})) \tag{7.37}$$

$$= \inf_{j=1,\dots,N} \inf_{\sigma \text{ pure}} \operatorname{tr}(T(\sigma^{(j)})\sigma^{\otimes N})) \tag{7.38}$$

$$= \inf_{j=1,\dots,N} \inf_{\psi} \langle \psi^{\otimes N}, T(\sigma^{(j)})\psi^{\otimes N} \rangle \ . \tag{7.39}$$

Hence $\mathscr{F}_{c,\#}$ only depends on the $\mathscr{B}(\mathscr{H}_+^{\otimes N})$ component (where $\mathscr{H}_+^{\otimes N}$ denotes again the Bose-subspace of $\mathscr{H}^{\otimes N}$) of T and we can assume without loss of generality that T is of the form

$$T : \mathscr{B}(\mathscr{H}^{\otimes M}) \to \mathscr{B}(\mathscr{H}_+^{\otimes N}) \ . \tag{7.40}$$

The restriction of $U^{\otimes N}$ to $\mathscr{H}_+^{\otimes N}$ is an irreducible representation (for any d) and in the qubit case ($d = 2$) we have $U^{\otimes N}\psi = \pi_s(U)\psi$ with $s = N/2$ for all $\psi \in \mathscr{H}_+^{\otimes N}$. The decomposition of T from Eq. (7.27) contains therefore only those summands with $s = N/2$. This simplifies the optimization problem significantly, since the number of variables needed to parametrize all relevant cloning maps according to Eq. (7.31) is reduced from 3 to 2. A more detailed (and non-trivial) analysis shows that the maximum for $\mathscr{F}_{c,1}$ and $\mathscr{F}_{c,\text{all}}$ is attained if all terms in (7.31) except the one with $s=N/2$, $j=N/2$ and $l=(M-N)/2$ vanish. The precise result is stated in the following theorem ([68,31,32] for qubits and [166,98] for general d).

Theorem 7.1. *For each $\mathscr{H} = \mathbb{C}^d$ both figures of merit $\mathscr{F}_{c,1}$ and $\mathscr{F}_{c,\text{all}}$ are maximized by the cloner*

$$\hat{T}^*(\rho) = \frac{\mathrm{d}[N]}{\mathrm{d}[M]} S_M(\rho \otimes \mathbb{1})S_M \ , \tag{7.41}$$

where d[N], d[M] *denote the dimensions of the symmetric tensor products $\mathscr{H}_+^{\otimes N}$, respectively $\mathscr{H}_+^{\otimes M}$, and S_M is the projection from $\mathscr{H}^{\otimes M}$ to $\mathscr{H}_+^{\otimes M}$. This implies for the optimal fidelities*

$$\mathscr{F}_{c,1}(N,M) = \frac{d-1}{d}\frac{N}{N+d}\frac{M+d}{M} \tag{7.42}$$

and

$$\mathscr{F}_{c,\text{all}}(N,M) = \frac{\mathrm{d}[N]}{\mathrm{d}[M]} \ . \tag{7.43}$$

$\hat{T}$ is the unique *solution for both optimization problems, i.e. there is no other operation T of form* (7.40) *which maximizes $\mathscr{F}_{c,1}$ or $\mathscr{F}_{c,\text{all}}$.*

There are two aspects of this result which deserve special attention. One is the relation to state estimation which is postponed to Section 7.2.3. The second concerns the role of correlations: It does not matter whether we are looking for the quality of each single clone ($\mathscr{F}_{c,1}$) only, or whether correlations are taken into account ($\mathscr{F}_{c,\text{all}}$). In both cases we get the same optimal solution. This is a special feature of pure states, however. Although there are no concrete results for quantum systems, it can be checked quite easily in the classical case that considering correlations changes the optimal cloner for arbitrary mixed states drastically.

7.2.2. Purification

To find an optimal purification device, i.e. maximizing $\mathscr{F}_{R,\#}$, is more difficult than the cloning problem, because the simplification from Eq. (7.40) does not apply. Hence we have to consider all the summands in the direct sum decomposition of T from Eq. (7.31) and solutions are available only for qubits. Therefore we will assume for the rest of this subsection that $\mathscr{H} = \mathbb{C}^2$ holds. The SU(2) symmetry of the problem allows us to assume without loss of generality that the pure initial state ψ coincides with one of the basis vectors. Hence we get for the (noisy) input states of the purifier

$$\rho(\beta) = \frac{1}{2\cosh(\beta)} \exp\left(2\beta\frac{\sigma_3}{2}\right) = \frac{1}{e^{\beta} + e^{-\beta}} \begin{pmatrix} e^{\beta} & 0 \\ 0 & e^{-\beta} \end{pmatrix} \tag{7.44}$$

$$= \tanh(\beta)|\psi\rangle\langle\psi| + (1 - \tanh(\beta))\tfrac{1}{2}\mathbb{1}, \quad \psi = |0\rangle , \tag{7.45}$$

The parameterization of ρ in terms of the "pseudo-temperature" β is chosen here, because it simplifies some calculations significantly (as we will see soon). The relation to the form of $\rho = R^*\sigma$ initially given in Eq. (7.4) is obviously $\vartheta = \tanh(\beta)$.

To state the main result of this subsection we have to decompose the product state $\rho(\beta)^{\otimes N}$ into spin-s components. This can be done in terms of Eq. (7.26). $\rho(\beta)$ is not unitary of course. However, we can apply (7.26) by analytic continuation, i.e. we treat $\rho(\beta)$ in the same way as we would $\exp(i\beta\sigma_3)$. It is then straightforward to get

$$\rho(\beta)^{\otimes N} = \bigoplus_s w_N(s)\rho_s(\beta) \otimes \frac{\mathbb{1}}{\dim \mathscr{K}_{N,s}} \tag{7.46}$$

with

$$w_N(s) = \frac{\sinh((2s+1)\beta)}{\sinh(\beta)(2\cosh(\beta))^N} \dim \mathscr{K}_{N,s} \tag{7.47}$$

and

$$\rho_s(\beta) = \frac{\sinh(\beta)}{\sinh((2s+1)\beta)} \exp(2\beta L_3^{(s)}) ,$$

where $L_3^{(s)}$ is the three-component of angular momentum in the spin-s representation and the dimension of $\mathscr{K}_{N,s}$ is given in Eq. (7.25). By (7.23) the representation space of π_s coincides with the symmetric tensor product $\mathscr{H}_+^{2s}$. Hence we can interpret $\rho_s(\beta)$ as a state of $2s$ (indistinguishable) particles. In other words the decomposition of $\rho(\beta)^{\otimes N}$ leads in a natural way to a family of operations

$$Q_s : \mathscr{B}(\mathscr{H}_+^{\otimes 2s}) \to \mathscr{B}(\mathscr{H}^{\otimes N}) \quad \text{with } Q_s^*[\rho(\beta)^{\otimes N}] = \rho_s(\beta) . \tag{7.48}$$

We can think of the family Q_s, of operations as an instrument Q which measures the number of output systems and transforms $\rho(\beta)^{\otimes N}$ to the appropriate $\rho_s(\beta)$. The crucial point is now that the purity of $\rho_s(\beta)$, measured in terms of fidelities with respect to ψ increases provided $s > 1/2$ holds.

Hence, we can think of Q as a purifier which arises naturally by reduction to irreducible spin components [46]. Unfortunately, Q does not produce a fixed number of output systems. The most obvious way to construct a device which produces always the same number M of outputs is to run the optimal $2s \to M$ cloner $\hat{T}_{2s\to M}$ if $2s < M$ or to drop $2s - M$ particles if $M \leqslant 2s$ holds. More precisely we can define $\hat{Q}: \mathscr{B}(\mathscr{H}^{\otimes M}) \to \mathscr{B}(\mathscr{H}^{\otimes N})$ by

$$\hat{Q}^*[\rho(\beta)^{\otimes N}] = \sum_s w_N(s)\hat{T}^*_{2s\to M}[\rho_s(\beta)] \tag{7.49}$$

with

$$\hat{T}^*_{2s\to M}(\rho) = \begin{cases} \dfrac{d[2s]}{d[M]} S_M(\rho \otimes \mathbb{1})S_M, & \text{for } M > 2s, \\ \mathrm{tr}_{2s-M}\,\rho & \text{for } M \leqslant 2s. \end{cases} \tag{7.50}$$

tr_{2s-M} denotes here the partial trace over the $2s-M$ first tensor factors. Applying the general scheme of Section 7.1.3 shows that this is the best way to get exactly M purified qubits [100]:

Theorem 7.2. *The operation* $\hat{Q}$ *defined in Eq.* (7.49) *maximizes* $\mathscr{F}_{R,1}$ *and* $\mathscr{F}_{R,\mathrm{all}}$. *It is called therefore the* optimal purifier. *The maximal values for* $\mathscr{F}_{R,1}$ *and* $\mathscr{F}_{R,\mathrm{all}}$ *are given by*

$$\mathscr{F}_{R,1}(N,M) = \sum_s w_N(s) f_1(M,\beta,s), \quad \mathscr{F}_{R,\mathrm{all}}(N,M) = \sum_s w_N(s) f_{\mathrm{all}}(M,\beta,s) \tag{7.51}$$

with

$$\begin{aligned} &2f_1(M,\beta,s) - 1 \\ &\quad = \begin{cases} \dfrac{2s+1}{2s}\coth((2s+1)\beta) - \dfrac{1}{2s}\coth\beta & \mathit{for}\ 2s > M, \\ \dfrac{1}{2s+2}\,\dfrac{M+2}{M}((2s+1)\coth((2s+1)\beta) - \coth\beta) & \mathit{for}\ 2s \leqslant M, \end{cases} \end{aligned} \tag{7.52}$$

and

$$f_{\mathrm{all}}(M,\beta,s) = \begin{cases} \dfrac{2s+1}{M+1}\,\dfrac{1-e^{-2\beta}}{1-e^{-(4s+2)\beta}} & M \leqslant 2s \\ \dfrac{1-e^{-2\beta}}{1-e^{-(4s+2)\beta}}\begin{pmatrix} 2s \\ M \end{pmatrix}^{-1} \sum_K \begin{pmatrix} K \\ M \end{pmatrix} e^{2\beta(K-s)} & M > 2s. \end{cases} \tag{7.53}$$

The expression for the optimal fidelities given here look rather complicated and are not very illuminating. We have plotted there both quantities as a function of ϑ (Fig. 7.1) of N (Fig. 7.2) and M (Fig. 7.3). While the first two plots looks quite similar the functional behavior in dependence of M seems to be very different. The study of the asymptotic behavior in the next section will give a precise analysis of this observation.

7.2.3. Estimating pure states

We have already seen in Section 4.2 that the cloning problem and state estimation are closely related, because we can construct an approximate cloner T from an estimator E simply by running

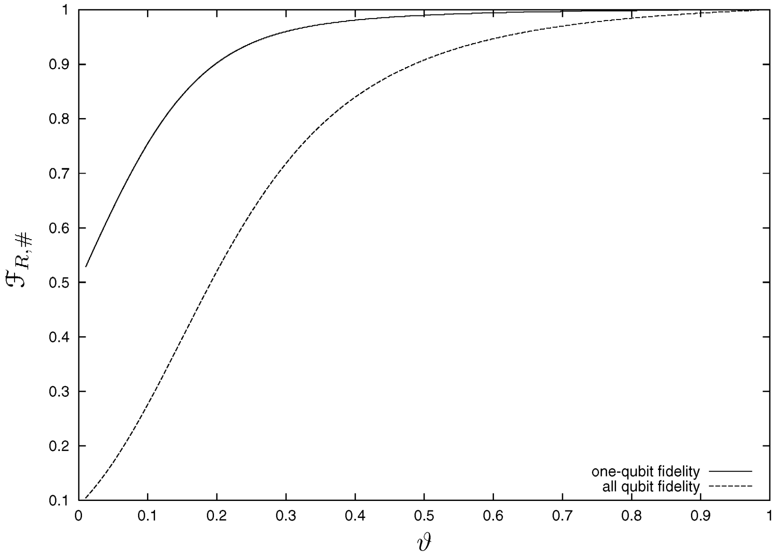


Fig. 7.1. One- and all-qubit fidelities of the optimal purifier for $N = 100$ and $M = 10$. Plotted as a function of the noise parameter ϑ.

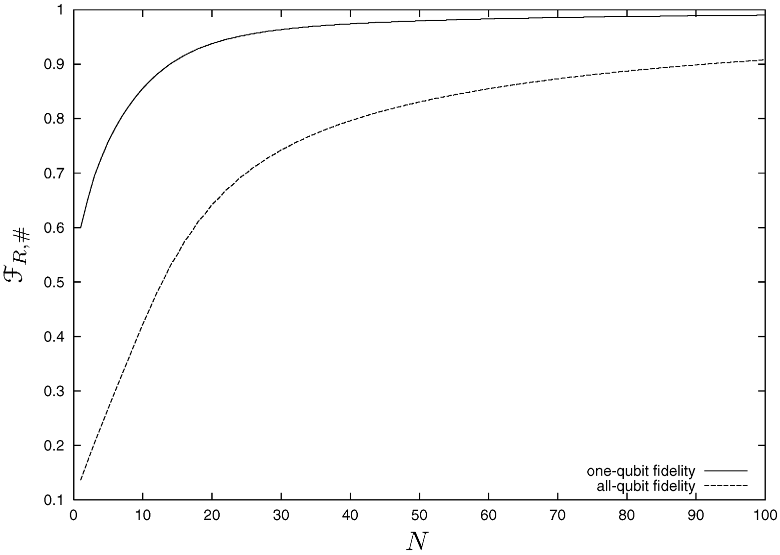


Fig. 7.2. One- and all-qubit fidelities of the optimal purifier for $\vartheta = 0.5$ and $M = 10$. Plotted as a function of N.

E on the N input states, and preparing M systems according to the attained classical information. In this section we want to go the other way round and show that the optimal cloner derived in Theorem 7.1 leads immediately to an optimal pure state estimator; cf. [33].

To this end let us assume that E has the form (cf. Section 4.2)

$$\mathscr{C}(X) \ni f \mapsto E(f) = \sum_{\sigma \in X} f(\sigma) E_\sigma \in \mathscr{B}(\mathscr{H}^{\otimes N}) , \tag{7.54}$$

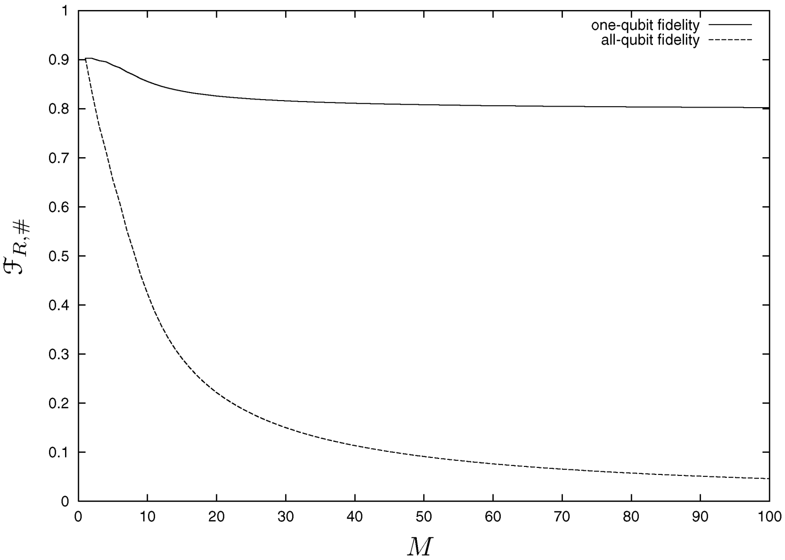


Fig. 7.3. One- and all-qubit fidelities of the optimal purifier for $\vartheta = 0.5$ and $N = 10$. Plotted as a function of M.

where $X \subset \mathscr{B}^*(\mathscr{H})$ is a finite set[23] of pure states. The quality of E can be measured in analogy to Section 7.1.1 by a fidelity-like quantity

$$\mathscr{F}_s(E) = \inf_{\psi\in\mathscr{H}} \langle\psi, \rho_\psi\psi\rangle = \inf_{\psi\in\mathscr{H}} \sum_{\sigma\in X} \langle\psi^{\otimes N}, E_\sigma\psi^{\otimes N}\rangle\langle\psi, \sigma\psi\rangle , \tag{7.55}$$

where $\rho_\psi = \sum_\sigma \langle\psi^{\otimes N}, E_\sigma\psi^{\otimes n}\rangle\sigma$ is the (density matrix valued) expectation value of E and the infimum is taken over all pure states ψ. Hence $\mathscr{F}_s(E)$ measures the worst fidelity of ρ_ψ with respect to the input state ψ. If we construct now a cloner T_E from E by

$$T_E^*(|\psi\rangle\langle\psi|^{\otimes N}) = \sum_\sigma \langle\psi^{\otimes N}, E_\sigma\psi^{\otimes n}\rangle\sigma^{\otimes M} \tag{7.56}$$

its one-particle fidelity $\mathscr{F}_{c,1}(T_E)$ coincides obviously with $\mathscr{F}_s(E)$. Since we can produce in this way arbitrary many clones of the same quality we see that $\mathscr{F}_s(E)$ is smaller than $\mathscr{F}_{c,1}(N, M)$ for all M and therefore

$$\mathscr{F}_s(E) \leqslant \mathscr{F}_{c,1}(N, \infty) = \lim_{M\to\infty} \mathscr{F}_{c,1}(N, M) = \frac{d-1}{d}\,\frac{N}{N+d} , \tag{7.57}$$

where we can look at $\mathscr{F}_{c,1}(N, \infty)$ as the optimal quality of a cloner which produces arbitrary many outputs from N input systems.

To see that this bound can be saturated consider an asymptotically exact family

$$\mathscr{C}(X_M) \ni f \mapsto E^M(f) = \sum_{\sigma\in X} f(\sigma)E^M_\sigma \in \mathscr{B}(\mathscr{H}^{\otimes M}),\ X_M \subset \mathscr{S}(\mathscr{H}) \tag{7.58}$$

[23] The generalization of the following considerations to continuous sets and a measure theoretic setup is straightforward and does not lead to a different result; i.e. we cannot improve the estimation quality with continuous observables.

of estimators, i.e. the error probabilities (4.17) vanish in the limit $N \to \infty$. If the $E_\sigma^M \in \mathscr{B}(\mathscr{H}^{\otimes M})$ are pure tensor products (i.e. the E^M are realized by a "quorum" of observables as described in Section 4.2.1) they cannot distinguish between the output state $\hat{T}^*(\rho^{\otimes N})$ (which is highly correlated) and the pure product state $\tilde{\rho}^{\otimes M}$ where $\tilde{\rho} \in \mathscr{B}^*(\mathscr{H})$ denotes the partial trace over $M-1$ tensor factors (due to permutation invariance it does not matter which factors we trace away here). Hence if we apply E^M to the output of the optimal N to M cloner $\hat{T}_{N \to M}$ we get an estimate for $\tilde{\rho}$ and in the limit $M \to \infty$ this estimate is exact. The fidelity $\langle \psi, \tilde{\rho}\psi \rangle$ of $\tilde{\rho}$ with respect to the pure input state ψ of $\hat{T}_{N \to M}$ coincides however with $\mathscr{F}_{c,1}(N,M)$. Hence the composition of $\hat{T}_{N \to M}$ with E^M converges[24] to an estimator E with $\mathscr{F}_e(E) = \mathscr{F}_{c,1}(N,\infty)$. We can rephrase this result roughly in the from: "producing infinitely many optimal clones of a pure state ψ is the same as estimating ψ optimally".

7.2.4. The UNOT gate

The discussion of the last subsection shows that the optimal cloner $\hat{T}_{N \to M}$ produces better clones than any estimation-based scheme (as in Eq. (7.56)), as long as we are interested only in *finitely many* copies. Loosely speaking we can say that the detour via classical information is wasteful and destroys too much quantum information. The same is true for the optimal purifier: We can first run an estimator on the mixed input state $\rho(\beta)^{\otimes N}$, apply the inverse $(R^*)^{-1}$ of the channel map to the attained classical data and reprepare arbitrarily many purified qubits accordingly. The quality of output systems attained this way is, however, worse than those of the optimal purifier from Eq. (7.49) as long as the number M of output systems is finite; this can be seen easily from Fig. 7.3. In this sense the UNOT gate is a harder task than cloning and purification, because there is *no quantum operation* which performs better than the estimation-based strategy. The following theorem can be proved again with the group theoretical scheme from Section 7.1.3 [36].

Theorem 7.3. *Let $\mathscr{H} = \mathbb{C}^2$. Among all channels $T : \mathscr{B}(\mathscr{H}) \to \mathscr{B}(\mathscr{H}_+^{\otimes N})$ the estimation-based scheme just described attains the biggest possible value for the fidelity $\mathscr{F}_{\theta,\#}$, namely*

$$\mathscr{F}_{\theta,1}(N,1) = \mathscr{F}_{\theta,\mathrm{all}}(N,1) = 1 - \frac{1}{N+2} . \tag{7.59}$$

The dependence on the number M of outputs is not interesting here, because the optimal device produces arbitrarily many copies of the same quality.

7.3. Asymptotic behaviour

If a device, such as the optimal cloner, is given which produces M output system from N inputs it is interesting to ask for the maximal rate, i.e. the maximal ratio $M(N)/N$ in the limit $N \to \infty$ such that the asymptotic fidelity $\lim_{N\to\infty} \mathscr{F}(N, M(N))$ is above a certain threshold (preferably equal to one). Note that this type of question was very important as well for distillation of entanglement and channel capacities, but almost not computable in there. In the current context this type of question is somewhat easier to answer. This relies on the one hand on the group theoretical structure presented

[24] Basically convergence must be shown here. It follows however easily from the corresponding property of the E^M.

in the last section and on the other on the close relation to quantum state estimation. We start this section therefore with a look on some aspects of the asymptotics of mixed state estimation.

7.3.1. Estimating mixed state

If we do not know a priori that the input systems are in a pure state much less is known about estimating and cloning. It is, in particular, almost impossible to say anything about optimality for finitely many input systems (only if N is very small e.g. [156]). Nevertheless, some strong results are available for the behavior in the limit $N \to \infty$ and we will give here a short review of some of them.

One quantity, interesting to be analyzed for a family of estimators E^N in the limit $N \to \infty$ is the variance of the E^N. To state some results in this context it is convenient to parameterize the state space $\mathcal{S}(\mathcal{H})$ or parts of it in terms of n real parameters $x = (x_1, \ldots, x_n) = \Sigma \subset \mathbb{R}^n$ and to write $\rho(x)$ as the corresponding state. If we want to cover all states, one particular parameterization is e.g. the generalized Bloch ball from Section 2.1.2. An estimator taking N input systems is now a (discrete) observable $E^N_x \in \mathcal{B}(\mathcal{H}^{\otimes N})$, $x \in X_N$ with values in a (finite) subset X_N of Σ. The expectation value of E^N in the state $\rho(x)^{\otimes N}$ is therefore the vector $\langle E^N \rangle_x$ with components $\langle E^N \rangle_x, j$, $j = 1, \ldots, n$ given by

$$\langle E^N \rangle_{x,j} = \sum_{y \in X_N} y_j \operatorname{tr}(E^N_y \rho(x)^{\otimes N}) \tag{7.60}$$

and the *mean quadratic error* is described by the matrix

$$V^N_{jk}(x) = \sum_{y \in X_N} (\langle E_N \rangle_{x,j} - y_j)(\langle E_N \rangle_{x,k} - y_k) \operatorname{tr}(E^N_y \rho(x)^{\otimes N}) \,. \tag{7.61}$$

For a good estimation strategy we expect that $V_{jk}(x)$ decreases as $1/N$, i.e.

$$V^N_{jk}(x) \simeq \frac{W_{jk}(x)}{N} \,, \tag{7.62}$$

where the *scaled* mean quadratic error matrix $W_{jk}(x)$ does not depend on N. The task is now to find bounds on this matrix. We will state here one result taken from [66]. To this end we need the *Hellström quantum information matrix*

$$H_{jk}(x) = \operatorname{tr}\left[\rho(x) \frac{\lambda j(x)\lambda_k(x) - \lambda_k(x)\lambda_j(x)}{2}\right] \,, \tag{7.63}$$

which is defined in terms of *symmetric logarithmic derivatives* λ_j, which in turn are implicitly given by

$$\frac{\partial \rho(x)}{\partial x_j} = \frac{\lambda_j(x)\rho(x) + \rho(x)\lambda_j(x)}{2} \,. \tag{7.64}$$

Now we have the following theorem [66]:

Theorem 7.4. *Consider a family of estimators E^N, $N \in \mathbb{N}$ as described above such that the following conditions hold*:

1. *The scaled mean quadratic error matrix $NV^N_{jk}(x)$ converges uniformly in x to $W_{jk}(x)$ as $N \to \infty$.*

2. *$W_{jk}(x)$ is continuous at a point $x_0 = x$.*
3. *$H_{jk}(x)$ and its derivatives are bounded in a neighborhood of x_0.*

Then we have

$$\mathrm{tr}[H^{-1}(x_0)W^{-1}(x_0)] \leqslant (d-1) . \tag{7.65}$$

For qubits this bound can be attained by a particular estimation strategy which measures on each qubit separately. We refer to [66] for details.

A second quantity interesting to study in the limit $N \to \infty$ is the error probability defined in Section 4.2; cf. Eq. (4.17). For a good estimation strategy it should go to zero of course, an additional question, however, concerns the rate with which this happens. We will review here a result from [99] which concerns the subproblem of *estimating the spectrum*. Hence we are looking now at a family of observables $E^N : \mathscr{C}(X_N) \to \mathscr{B}(\mathscr{H}^{\otimes N})$, $N \in \mathbb{N}$ taking their values in a finite subset X_N of the set

$$\Sigma = \left\{ (x_1, \dots, x_d) \in \mathbb{R}^d \,|\, x_1 \geqslant \cdots \geqslant x_d \geqslant 0, \sum_j x_j = 1 \right\} \tag{7.66}$$

of ordered spectra of density operators on $\mathscr{H} = \mathbb{C}^d$. Our aim is to determine the behavior of the error probabilities (cf. Eq. (4.17)

$$K_N(\Delta) = \sum_{x \in \Delta \cap X_N} \mathrm{tr}(E_x^N \rho^{\otimes N}) \tag{7.67}$$

in the limit $N \to \infty$. Following the general arguments in Section 7.1.2 we can restrict our attention here to covariant observables, i.e. we can assume without loss of cloning quality that the E_x^N commute with all permutation unitaries V_p, $p \in S_N$ and all local unitaries $U^{\otimes N}$, $U \in U(d)$. If we restrict our attention in addition to projection-valued measures, which is suggestive for ruling out unnecessary fuzziness, we see that each E_x^N must coincide with a (sum of) projections P_Y from $\mathscr{H}^{\otimes N}$ onto the $U(d)$, respectively V_p, invariant subspace $\mathscr{H}_Y \otimes \mathscr{K}_Y$, which is defined in Eq. (7.32), where $Y = (Y_1, \dots, Y_d)$ refers here to Young frames with d rows and N boxes. The only remaining freedom for the E^N is the assignment $x(Y) \in \Sigma$ of Young frames (and therefore projections E_N) to points in Σ. Since the Young frames themselves have up to normalization the same structure as the elements of Σ, one possibility for $s(Y)$ is just $s(Y) = Y/N$. Written as quantum to classical channel this is

$$\mathscr{C}(X_N) \ni f \mapsto \sum_Y f(Y/N) P_Y \in \mathscr{B}(\mathscr{H}^{\otimes N}) , \tag{7.68}$$

where $X_N \subset \Sigma$ is the set of normalized Young frames, i.e. all Y/N if Y has d rows and N boxes. It turns out, somewhat surprisingly that this choice leads indeed to an asymptotically exact estimation strategy with exponentially decaying error probability (7.67). The following theorem can be proven with methods from the theory of large deviations:

Theorem 7.5. *The family of estimators E^N, $N \in \mathbb{N}$ given in Eq.* (7.68) *is* asymptotically exact, *i.e. the error probabilities $K_N(\Delta)$ vanish in the limit $N \to \infty$ if Δ is a complement of a ball around* *the spectrum $r \in \Sigma$ of ρ. If Δ is a set (possibly containing r) whose interior is dense in its closure we have the asymptotic estimate for $K_N(\Delta)$:*

$$\lim_{N\to\infty} \frac{1}{N} \ln K_N(\Delta) = \inf_{s\in\Delta} I(s) , \tag{7.69}$$

where the "rate function" $I : \Sigma \to \mathbb{R}$ is just the relative entropy between the two probability vectors s and r

$$I(s) = \sum_j s_j(\ln s_j - \ln r_j) . \tag{7.70}$$

To make this statement more transparent, note that we can rephrase (7.69) as

$$K_N(\Delta) \approx \exp\left(-N \inf_{s\in\Delta} I(s) \right) . \tag{7.71}$$

Since the rate function I vanishes only for $s = r$ we see that the probability measures K_N converge (weakly) to a point measure concentrated at $r \in \Sigma$. The rate of this convergence is exponential and measured exactly by the function I.

7.3.2. Purification and cloning

Let us come back now to the discussion of purification started in Section 7.2.2 (consequently we have $\mathscr{H} = \mathbb{C}^2$ again). Our aim is now to calculate the fidelities $\mathscr{F}_{R,\#}(N, M(N))$ in the limit $N \to \infty$ for a sequence $M(N)$, $N \in \mathbb{N}$ such that $M(N)/N$ converges to a value $c \in \mathbb{R}$. The crucial step to do this is the application of Theorem 7.5. The density matrices $\rho_s(\beta)$ from Eq. (7.46) can be defined alternatively by

$$\rho_s(\beta) \otimes \frac{\mathbb{1}}{\dim \mathscr{K}_{N,s}} = w_N(s)^{-1} P_s \rho(\beta)^{\otimes N} P_s, \quad w_N(s) = \operatorname{tr}(\rho(\beta)^{\otimes N} P_s) , \tag{7.72}$$

where P_s is the projection from $\mathscr{H}^{\otimes N}$ to $\mathscr{H}_s \otimes \mathscr{K}_{N,s}$. In other words P_s is equal to P_Y from Eq. (7.68) if we apply the reparametrization

$$(Y_1, Y_2) \mapsto (s, N) = ((Y_1 - Y_2)/2, Y_1 + Y_2) . \tag{7.73}$$

In a similar way we can rewrite the set of ordered spectra by $\Sigma \ni (x_1, x_2) \mapsto x_1 - x_2 \in [0, 1]$ and $K_N(\Delta)$ becomes a measure on $[0, 1]$ (i.e. $\Delta \subset [0, 1]$):

$$K_N(\Delta) = \sum_{2s/N\in\Delta} \operatorname{tr}(\rho(\beta)^{\otimes N} P_s) = \sum_{2s/N\in\Delta} w_N(s) \tag{7.74}$$

and the sum

$$\mathscr{F}_{R,\#}(N, M(N)) = \sum_s w_N(s) f_\#(M(N), \beta, s) \tag{7.75}$$

can be rephrased as the integral of a function $[0, 1] \ni x \mapsto \tilde{f}_\#(N, \beta, x) \in \mathbb{R}$ with respect to this measure, provided $\tilde{f}_\#$ is related to $f_\#$ by $\tilde{f}_\#(N, \beta, 2s/N) = f_\#(M(N), \beta, s)$. According to Theorem 7.5 the K_N converge to a point measure concentrated at the ordered spectrum of $\rho(\beta)$; but the latter corresponds, according to the reparametrization above, to the noise parameter $\vartheta = \tanh \beta$. Hence, if

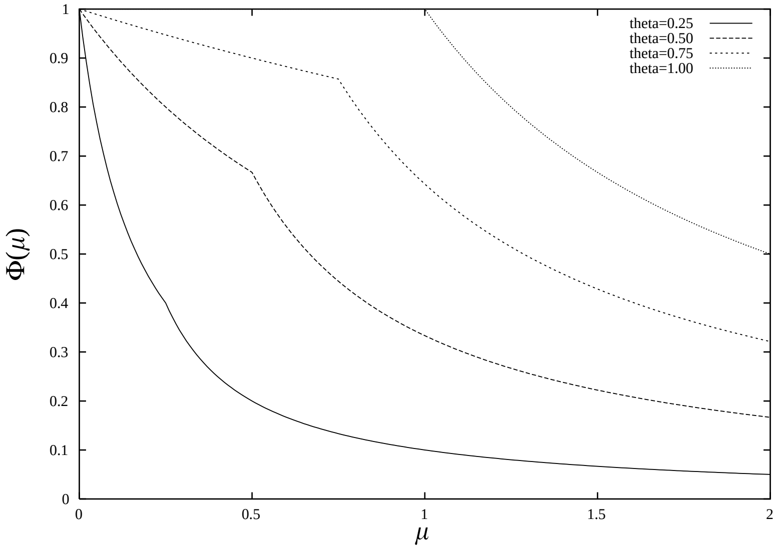


Fig. 7.4. Asymptotic all-qubit fidelity $\Phi(\mu)$ plotted as function of the rate μ.

the sequence of functions $\tilde{f}_{\#}(N,\beta,\cdot)$ converges for $N \to \infty$ uniformly (or at least uniformly on a neighborhood of ϑ) to $\tilde{f}_{\#}(\beta,\cdot)$ we get

$$\lim_{N\to\infty} \mathscr{F}(N,M(N)) = \lim_{N\to\infty} \sum_s \tilde{f}_{\#}(N,\beta,s) = \tilde{f}_{\#}(\beta,\vartheta) \tag{7.76}$$

for the limit of the fidelities. A precise formulation of this idea leads to the following theorem [100].

Theorem 7.6. *The two purification fidelities $\mathscr{F}_{R,\#}$ have the following limits*:

$$\lim_{N\to\infty}\lim_{M\to\infty} \mathscr{F}_{R,1}(N,M) = 1 \tag{7.77}$$

and

$$\Phi(\mu) = \lim_{\substack{N\to\infty \\ M/N\to\mu}} \mathscr{F}_{R,\mathrm{all}}(N,M) = \begin{cases} \dfrac{2\vartheta^2}{2\vartheta^2 + \mu(1-\vartheta)} & if\ \mu \leqslant \vartheta, \\ \dfrac{2\vartheta^2}{\mu(1+\vartheta)} & if\ \mu \geqslant \vartheta. \end{cases} \tag{7.78}$$

If we are only interested in the quality of each qubit separately we can produce arbitrarily good purified qubits at any rate. If on the other hand the correlations between the output systems should vanish in the limit the rate is always zero. This can be seen from the function Φ, which is the asymptotic all-qubit fidelity which can be reached by a given rate μ. We have plotted it in Fig. 7.4. Note finally that the results just stated contain the rates of optimal cloning machines as a special case; we only have to set $\vartheta = 1$.

References

[1] A. Acín, A. Andrianov, L. Costa, E. Jané, J.I. Latorre, R. Tarrach, Schmidt decomposition and classification of three-quantum-bit states, Phys. Rev. Lett. 85 (7) (2000) 1560–1563.

[2] C. Adami, N.J. Cerf, Von Neumann capacity of noisy quantum channels, Phys. Rev. A 56 (5) (1997) 3470–3483.

[3] G. Alber, T. Beth, M. Horodecki, R. Horodecki, M. Rötteler, H. Weinfurter, R. Werner, A. Zeilinger (Eds.), Quantum Information, Springer, Berlin, 2001.

[4] A. Ashikhmin, E. Knill, Nonbinary quantum stabilizer codes, IEEE Trans. Inf. Theory 47 (7) (2001) 3065–3072.

[5] A. Aspect, J. Dalibard, G. Roger, Experimental test of Bell's inequalities using time-varying analyzers, Phys. Rev. Lett. 49 (1982) 1804–1807.

[6] H. Barnum, E. Knill, M.A. Nielsen, On quantum fidelities and channel capacities, IEEE Trans. Inf. Theory 46 (2000) 1317–1329.

[7] H. Barnum, M.A. Nielsen, B. Schumacher, Information transmission through a noisy quantum channel, Phys. Rev. A 57 (6) (1998) 4153–4175.

[8] H. Barnum, J.A. Smolin, B.M. Terhal, Quantum capacity is properly defined without encodings, Phys. Rev. A 58 (5) (1998) 3496–3501.

[9] C.H. Bennett, H.J. Bernstein, S. Popescu, B. Schumacher, Concentrating partial entanglement by local operations, Phys. Rev. A 53 (4) (1996) 2046–2052.

[10] C.H. Bennett, G. Brassard, Quantum key distribution and coin tossing, in: Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing, Bangalore, India, IEEE, New York, 1984, pp. 175–179.

[11] C.H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, W.K. Wootters, Teleporting an unknown quantum state via dual classical and Einstein–Podolsky–Rosen channels, Phys. Rev. Lett. 70 (1993) 1895–1899.

[12] C.H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J.A. Smolin, W.K. Wootters, Purification of noisy entanglement and faithful teleportation via noisy channels, Phys. Rev. Lett. 76 (5) (1996) 722–725;
C.H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J.A. Smolin, W.K. Wootters, Erratum, Phys. Rev. Lett. 78 (10) (1997) 2031.

[13] C.H. Bennett, D.P. DiVincenzo, C.A. Fuchs, T. Mor, E.M. Rains, P.W. Shor, J.A. Smolin, W.K. Wootters, Quantum nonlocality without entanglement, Phys. Rev. A 59 (2) (1999) 1070–1091.

[14] C.H. Bennett, D.P. DiVincenzo, T. Mor, P.W. Shor, J.A. Smolin, B.M. Terhal, Unextendible product bases and bound entanglement, Phys. Rev. Lett. 82 (26) (1999) 5385–5388.

[15] C.H. Bennett, D.P. DiVincenzo, J.A. Smolin, Capacities of quantum erasure channels, Phys. Rev. Lett. 78 (16) (1997) 3217–3220.

[16] C.H. Bennett, D.P. DiVincenzo, J.A. Smolin, W.K. Wootters, Mixed-state entanglement and quantum error correction, Phys. Rev. A 54 (4) (1996) 3824–3851.

[17] C.H. Bennett, P.W. Shor, J.A. Smolin, A.V. Thapliyal, Entanglement-assisted classical capacity of noisy quantum channels, Phys. Rev. Lett. 83 (15) (1999) 3081–3084.

[18] C.H. Bennett, P.W. Shor, J.A. Smolin, A.V. Thapliyal, Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem, 2001, quant-ph/0106052.

[19] C.H. Bennett, S.J. Wiesner, Communication via one- and two-particle operators on Einstein–Podolsky–Rosen states, Phys. Rev. Lett. 20 (1992) 2881–2884.

[20] T. Beth, M. Rötteler, Quantum algorithms: applicable algebra and quantum physics, in: G. Alber, et al., (Eds.), Quantum Information, Springer, Berlin, 2001, pp. 97–150.

[21] E. Biolatti, R.C. Iotti, P. Zanardi, F. Rossi, Quantum information processing with semiconductor macroatoms, Phys. Rev. Lett. 85 (26) (2000) 5647–5650.

[22] D. Boschi, S. Branca, F. De Martini, L. Hardy, S. Popescu, Experimental realization of teleporting an unknown pure quantum state via dual classical an Einstein–Podolsky–Rosen channels, Phys. Rev. Lett. 80 (6) (1998) 1121–1125.

[23] D. Bouwmeester, A.K. Ekert, A. Zeilinger (Eds.), The Physics of Quantum Information: Quantum Cryptography, Quantum Teleportation, Quantum Computation, Springer, Berlin, 2000.

[24] D. Bouwmeester, J.-W. Pan, K. Mattle, M. Eibl, H. Weinfurter, A. Zeilinger, Experimental quantum teleportation, Nature 390 (1997) 575–579.

[25] O. Bratteli, D.W. Robinson, Operator Algebras and Quantum Statistical Mechanics I, Springer, New York, 1979.

[26] O. Bratteli, D.W. Robinson, Operator Algebras and Quantum Statistical Mechanics II, Springer, Berlin, 1997.

[27] S.L. Braunstein, C.M. Caves, R. Jozsa, N. Linden, S. Popescu, R. Schack, Separability of very noisy mixed states and implications for NMR quantum computing, Phys. Rev. Lett. 83 (5) (1999) 1054–1057.

[28] G.K. Brennen, C.M. Caves, I.H. Deutsch, F.S. Jessen, Quantum logic gates in optical lattices, Phys. Rev. Lett. 82 (5) (1999) 1060–1063.

[29] K.R. Brown, D.A. Lidar, K.B. Whaley, Quantum computing with quantum dots on linear supports, 2001, quant-ph/0105102.

[30] T.A. Brun, H.L. Wang, Coupling nanocrystals to a high-q silica microsphere: entanglement in quantum dots via photon exchange, Phys. Rev. A 61 (2000) 032307.

[31] D. Bruß, D.P. DiVincenzo, A. Ekert, C.A. Fuchs, C. Machiavello, J.A. Smolin, Optimal universal and state-dependent cloning, Phys. Rev. A 57 (4) (1998) 2368–2378.

[32] D. Bruß, A.K. Ekert, C. Macchiavello, Optimal universal quantum cloning and state estimation, Phys. Rev. Lett. 81 (12) (1998) 2598–2601.

[33] D. Bruß, C. Macchiavello, Optimal state estimation for d-dimensional quantum systems, Phys. Lett. A 253 (1999) 249–251.

[34] W.T. Buttler, R.J. Hughes, S.K. Lamoreaux, G.L. Morgan, J.E. Nordholt, C.G. Peterson, Daylight quantum key distribution over 1.6 km, Phys. Rev. Lett. 84 (2000) 5652–5655.

[35] V. Bužek, M. Hillery, Universal optimal cloning of qubits and quantum registers, Phys. Rev. Lett. 81 (22) (1998) 5003–5006.

[36] V. Bužek, M. Hillery, R.F. Werner, Optimal manipulations with qubits: universal-not gate, Phys. Rev. A 60 (4) (1999) R2626–R2629.

[37] A. Cabello, Bibliographic guide to the foundations of quantum mechanics and quantum information, 2000, quant-ph/0012089.

[38] A.R. Calderbank, E.M. Rains, P.W. Shor, N.J.A. Sloane, Quantum error correction and orthogonal geometry, Phys. Rev. Lett. 78 (3) (1997) 405–408.

[39] A.R. Calderbank, P.W. Shor, Good quantum error-correcting codes exist, Phys. Rev. A 54 (1996) 1098–1105.

[40] N.J. Cerf, Asymmetric quantum cloning in any direction, J. Mod. Opt. 47 (2) (2000) 187–209.

[41] N.J. Cerf, C. Adami, Negative entropy and information in quantum mechanics, Phys. Rev. Lett. 79 (26) (1997) 5194–5197.

[42] N.J. Cerf, C. Adami, R.M. Gingrich, Reduction criterion for separability, Phys. Rev. A 60 (2) (1999) 898–909.

[43] N.J. Cerf, S. Iblisdir, G. van Assche, Cloning and cryptography with quantum continuous variables, 2001, quant-ph/0107077.

[44] I.L. Chuang, L.M.K. Vandersypen, X.L. Zhou, D.W. Leung, S. Lloyd, Experimental realization of a quantum algorithm, Nature 393 (1998) 143–146.

[45] A. Church, An unsolved problem of elementary number theory, Am. J. Math. 58 (1936) 345–363.

[46] J.I. Cirac, A.K. Ekert, C. Macchiavello, Optimal purification of single qubits, Phys. Rev. Lett. 82 (1999) 4344–4347.

[47] J.F. Clauser, M.A. Horne, A. Shimony, R.A. Holt, Proposed experiment to test local hidden-variable theories, Phys. Rev. Lett. 23 (15) (1969) 880–884.

[48] J.F. Cornwell, Group Theory in Physics II, Academic Press, London, 1984.

[49] T.M. Cover, J.A. Thomas, Elements of Information Theory, Wiley, Chichester, 1991.

[50] E.B. Davies, Quantum Theory of Open Systems, Academic Press, London, 1976.

[51] B. Demoen, P. Vanheuverzwijn, A. Verbeure, Completely positive maps on the CCR-algebra, Lett. Math. Phys. 2 (1977) 161–166.

[52] D. Deutsch, Quantum theory, the Church–Turing principle and the universal quantum computer, Proc. R. Soc. London A 400 (1985) 97–117.

[53] D. Deutsch, R. Jozsa, Rapid solution of problems by quantum computation, Proc. R. Soc. London A 439 (1992) 553–558.

[54] D.P. DiVincenzo, P.W. Shor, J.A. Smolin, Quantum-channel capacity of very noisy channels, Phys. Rev. A 57 (2) (1998) 830–839;
D.P. DiVincenzo, P.W. Shor, J.A. Smolin, Erratum, Phys. Rev. A 59 (2) (1999) 1717.

[55] D.P. DiVincenzo, P.W. Shor, J.A. Smolin, B.M. Terhal, A.V. Thapliyal, Evidence for bound entangled states with negative partial transpose, Phys. Rev. A 61 (6) (2000) 062312.

[56] M.J. Donald, M. Horodecki, Continuity of relative entropy of entanglement, Phys. Lett. A 264 (4) (1999) 257–260.
[57] M.J. Donald, M. Horodecki, O. Rudolph, The uniqueness theorem for entanglement measures, 2001, quant-ph/0105017.
[58] W. Dür, J.I. Cirac, M. Lewenstein, D. Bruss, Distillability and partial transposition in bipartite systems, Phys. Rev. A 61 (6) (2000) 062313.
[59] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall, New York, 1993.
[60] T. Eggeling, K.G.H. Vollbrecht, R.F. Werner, M.M. Wolf, Distillability via protocols respecting the positivity of the partial transpose, Phys. Rev. Lett. 87 (2001) 257902.
[61] T. Eggeling, R.F. Werner, Separability properties of tripartite states with $U \times U \times U$-symmetry, Phys. Rev. A 63 (4) (2001) 042111.
[62] A. Feinstein, Foundations of Informations Theory, McGraw-Hill, New York, 1958.
[63] D.G. Fischer, M. Freyberger, Estimating mixed quantum states, Phys. Lett. A 273 (2000) 293–302.
[64] G. Giedke, L.-M. Duan, J.I. Cirac, P. Zoller, Distillability criterion for all bipartite gaussian states, Quant. Inf. Comput. 1 (3) (2001).
[65] G. Giedke, B. Kraus, M. Lewenstein, J.I. Cirac, Separability properties of three-mode gaussian states, Phys. Rev. A 64 (5) (2001) 052303.
[66] R.D. Gill, S. Massar, State estimation for large ensembles, Phys. Rev. A 61 (2000) 2312–2327.
[67] N. Gisin, Hidden quantum nonlocality revealed by local filters, Phys. Lett. A 210 (3) (1996) 151–156.
[68] N. Gisin, S. Massar, Optimal quantum cloning machines, Phys. Rev. Lett. 79 (11) (1997) 2153–2156.
[69] N. Gisin, G. Ribordy, W. Tittel, H. Zbinden, Quantum Cryptography, 2001, quant-ph/0101098.
[70] D. Gottesman, Class of quantum error-correcting codes saturating the quantum hamming bound, Phys. Rev. A 54 (1996) 1862–1868.
[71] D. Gottesman, Stabilizer codes and quantum error correction, Ph.D. Thesis, California Institute of Technology, 1997, quant-ph/9705052.
[72] M. Grassl, T. Beth, T. Pellizzari, Codes for the quantum erasure channel, Phys. Rev. A 56 (1) (1997) 33–38.
[73] D.M. Greenberger, M.A. Horne, A. Zeilinger, Going beyond bell's theorem, in: M. Kafatos (Ed.), Bell's Theorem, Quantum Theory, and Conceptions of the Universe, Kluwer Academic Publishers, Dordrecht, 1989, pp. 69–72.
[74] L.K. Grover, Quantum computers can search arbitrarily large databases by a single query, Phys. Rev. A 56 (23) (1997) 4709–4712.
[75] L.K. Grover, Quantum mechanics helps in searching for a needle in a haystack, Phys. Rev. Lett. 79 (2) (1997) 325–328.
[76] J. Gruska, Quantum Computing, McGraw-Hill, New York, 1999.
[77] J. Harrington, J. Preskill, Achievable rates for the gaussian quantum channel, Phys. Rev. A 64 (6) (2001) 062301.
[78] P.M. Hayden, M. Horodecki, B.M. Terhal, The asymptotic entanglement cost of preparing a quantum state, J. Phys. A. Math. Gen. 34 (35) (2001) 6891–6898.
[79] A.S. Holevo, Probabilistic and Statistical Aspects of Quantum Theory, North-Holland, Amsterdam, 1982.
[80] A.S. Holevo, Coding theorems for quantum channels, Tamagawa University Research Review no. 4, 1998, quant-ph/9809023.
[81] A.S. Holevo, Sending quantum information with gaussian states, in: Proceedings of the Fourth International Conference on Quantum Communication, Measurement and Computing, Evanston, 1998, quant-ph/9809022.
[82] A.S. Holevo, On entanglement-assisted classical capacity, 2001, quant-ph/0106075.
[83] A.S. Holevo, Statistical Structure of Quantum Theory, Springer, Berlin, 2001.
[84] A.S. Holevo, R.F. Werner, Evaluating capacities of bosonic gaussian channels, Phys. Rev. A 63 (3) (2001) 032312.
[85] M. Horodecki, P. Horodecki, Reduction criterion of separability and limits for a class of distillation protocols, Phys. Rev. A 59 (6) (1999) 4206–4216.
[86] M. Horodecki, P. Horodecki, R. Horodecki, Separability of mixed states: necessary and sufficient conditions, Phys. Lett. A 223 (1–2) (1996) 1–8.
[87] M. Horodecki, P. Horodecki, R. Horodecki, Mixed-state entanglement and distillation: is there a "bound" entanglement in nature? Phys. Rev. Lett. 80 (24) (1998) 5239–5242.
[88] M. Horodecki, P. Horodecki, R. Horodecki, General teleportation channel, singlet fraction, and quasidistillation, Phys. Rev. A 60 (3) (1999) 1888–1898.

[89] M. Horodecki, P. Horodecki, R. Horodecki, Limits for entanglement measures, Phys. Rev. Lett. 84 (9) (2000) 2014–2017.

[90] M. Horodecki, P. Horodecki, R. Horodecki, Unified approach to quantum capacities: towards quantum noisy coding theorem, Phys. Rev. Lett. 85 (2) (2000) 433–436.

[91] M. Horodecki, P. Horodecki, R. Horodecki, Mixed-state entanglement and quantum communication, in: G. Alber, et al., (Eds.), Quantum Information, Springer, Berlin, 2001, pp. 151–195.

[92] P. Horodecki, M. Horodecki, R. Horodecki, Bound entanglement can be activated, Phys. Rev. Lett. 82 (5) (1999) 1056–1059.

[93] R.J. Hughes, G.L. Morgan, C.G. Peterson, Quantum key distribution over a 48 km optical fibre network, J. Mod. Opt. 47 (2–3) (2000) 533–547.

[94] A. Jamiołkowski, Linear transformations which preserve trace and positive semidefiniteness of operators, Rep. Math. Phys. 3 (1972) 275–278.

[95] T. Jennewein, C. Simon, G. Weihs, H. Weinfurter, A. Zeilinger, Quantum cryptography with entangled photons, Phys. Rev. Lett. 84 (2000) 4729–4732.

[96] J.A. Jones, M. Mosca, R.H. Hansen, Implementation of a quantum search algorithm on a quantum computer, Nature 393 (1998) 344–346.

[97] M. Keyl, D. Schlingemann, R.F. Werner, Infinitely entangled states, in preparation.

[98] M. Keyl, R.F. Werner, Optimal cloning of pure states, testing single clones, J. Math. Phys. 40 (1999) 3283–3299.

[99] M. Keyl, R.F. Werner, Estimating the spectrum of a density operator, Phys. Rev. A 64 (5) (2001) 052311.

[100] M. Keyl, R.F. Werner, The rate of optimal purification procedures, Ann H. Poincaré 2 (2001) 1–26.

[101] A.I. Khinchin, Mathematical Foundations of Information Theory, Dover Publications, New York, 1957.

[102] B.E. King, C.S. Wood, C.J. Myatt, Q.A. Turchette, D. Leibfried, W.M. Itano, C. Monroe, D.J. Wineland, Cooling the collective motion of trapped ions to initialize a quantum register, Phys. Rev. Lett. 81 (7) (1998) 1525–1528.

[103] E. Knill, R. Laflamme, Theory of quantum error-correcting codes, Phys. Rev. A 55 (2) (1997) 900–911.

[104] B. Kraus, M. Lewenstein, J.I. Cirac, Characterization of distillable and activable states using entanglement witnesses, 2001, quant-ph/0110174.

[105] K. Kraus, States Effects and Operations, Springer, Berlin, 1983.

[106] R. Landauer, Irreversibility and heat generation in the computing process, IBM J. Res. Dev. 5 (1961) 183.

[107] U. Leonhardt, Measuring the Quantum State of Light, Cambridge University Press, Cambridge, 1997.

[108] M. Lewenstein, A. Sanpera, Separability and entanglement of composite quantum systems, Phys. Rev. Lett. 80 (11) (1998) 2261–2264.

[109] N. Linden, H. Barjat, R. Freeman, An implementation of the Deutsch–Jozsa algorithm on a three-qubit NMR quantum computer, Chem. Phys. Lett. 296 (1–2) (1998) 61–67.

[110] S. Lloyd, Capacity of the noisy quantum channel, Phys. Rev. A 55 (3) (1997) 1613–1622.

[111] H.-K. Lo, T. Spiller, S. Popescu (Eds.), Introduction to Quantum Computation and Information, World Scientific, Singapore, 1998.

[112] Y. Makhlin, G. Schön, A. Shnirman, Quantum-state engineering with Josephson-junction devices, Rev. Mod. Phys. 73 (2) (2001) 357–400.

[113] R. Marx, A.F. Fahmy, J.M. Myers, W. Bermel, S.J. Glaser, Approaching five-bit NMR quantum computing, Phys. Rev. A 62 (1) (2000) 012310.

[114] R. Matsumoto and T. Uyematsu, Lower bound for the quantum capacity of a discrete memoryless quantum channel, 2001, quant-ph/0105151.

[115] K. Mattle, H. Weinfurter, P.G. Kwiat, A. Zeilinger, Dense coding in experimental quantum communication, Phys. Rev. Lett. 76 (25) (1996) 4656–4659.

[116] N.D. Mermin, Quantum mysteries revisited, Am. J. Phys. 58 (8) (1990) 731–734.

[117] N.D. Mermin, What's wrong with these elements of reality? Phys. Today 43 (6) (1990) 9–11.

[118] H.C. Nagerl, W. Bechter, J. Eschner, F. Schmidt-Kaler, R. Blatt, Ion strings for quantum gates, Appl. Phys. B 66 (5) (1998) 603–608.

[119] M. A. Nielsen, Conditions for a class of entanglement transformations, Phys. Rev. Lett. 83 (2) (1999) 436–439.

[120] M.A. Nielsen, Continuity bounds for entanglement, Phys. Rev. A 61 (6) (2000) 064301.

[121] M.A. Nielsen, Characterizing mixing and measurement in quantum mechanics, Phys. Rev. A 63 (2) (2001) 022114.

[122] M.A. Nielsen, I.L. Chuang, Quantum Computation and Quantum Information, Cambridge University Press, Cambridge, 2000.

[123] M. Ohya, D. Petz, Quantum Entropy and its Use, Springer, Berlin, 1993.

[124] C.M. Papadimitriou, Computational Complexity, Addison-Wesley, Reading, MA, 1994.

[125] V.I. Paulsen, Completely Bounded Maps and Dilations, Longman Scientific & Technical, New York, 1986.

[126] A. Peres, Higher order schmidt decompositions, Phys. Lett. A 202 (1) (1995) 16–17.

[127] A. Peres, Separability criterion for density matrices, Phys. Rev. Lett. 77 (8) (1996) 1413–1415.

[128] S. Popescu, Bell's inequalities versus teleportation: what is nonlocality? Phys. Rev. Lett. 72 (6) (1994) 797–799.

[129] S. Popescu, D. Rohrlich, Thermodynamics and the measure of entanglement, Phys. Rev. A 56 (5) (1997) R3319–R3321.

[130] J. Preskill, Lecture notes for the course 'Information for Physics 219/Computer Science 219, Quantum Computation,' Caltech, Pasadena, California, 1999, www.theory.caltech.edu/people/preskill/ph229.

[131] M. Purser, Introduction to Error-Correcting Codes, Artech House, Boston, 1995.

[132] E.M. Rains, Bound on distillable entanglement, Phys. Rev. A 60 (1) (1999) 179–184;
E.M. Rains, Erratum, Phys. Rev. A 63 (1) (2001) 019902(E).

[133] E.M. Rains, A semidefinite program for distillable entanglement, IEEE Trans. Inf. Theory 47 (7) (2001) 2921–2933.

[134] M. Reed, B. Simon, Methods of Modern Mathematical Physics I, Academic Press, San Diego, 1980.

[135] W. Rudin, Functional Analysis, McGraw-Hill, New-York, 1973.

[136] O. Rudolph, A separability criterion for density operators, J. Phys. A 33 (21) (2000) 3951–3955.

[137] D. Schlingemann, R.F. Werner, Quantum error-correcting codes associated with graphs, 2000, quant-ph/0012111.

[138] C.E. Shannon, A mathematical theory of communication, Bell. Syst. Tech. J. 27 (1948) 379–423, 623–656.

[139] P.W. Shor, Algorithms for quantum computation: discrete logarithms and factoring, in: S. Goldwasser (Ed.), Proceedings of the 35th Annual Symposium on the Foundations of Computer Science, IEEE Computer Science, Society Press, Los Alamitos, CA, 1994, pp. 124–134.

[140] P.W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, Soc. Ind. Appl. Math. J. Comput. 26 (1997) 1484–1509.

[141] P.W. Shor, J.A. Smolin, B.M. Terhal, Nonadditivity of bipartite distillable entanglement follows from a conjecture on bound entangled Werner states, Phys. Rev. Lett. 86 (12) (2001) 2681–2684.

[142] B. Simon, Representations of Finite and Compact Groups, American Mathematical Society, Providence, RI, 1996.

[143] D. Simon, On the power of quantum computation, in: Proceedings of the 35th Annual Symposium on Foundations of Computer Science, IEEE Computer Society Press, Los Alamitos, 1994, pp. 124–134.

[144] R. Simon, Peres-Horodecki separability criterion for continuous variable systems, Phys. Rev. Lett. 84 (12) (2000) 2726–2729.

[145] S. Singh, The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography, Fourth Estate, London, 1999.

[146] A.M. Steane, Multiple particle interference and quantum error correction, Proc. Roy. Soc. London A 452 (1996) 2551–2577.

[147] W.F. Stinespring, Positive functions on C*-algebras, Proc. Am. Math. Soc. (1955) 211–216.

[148] E. Størmer, Positive linear maps of operator algebras, Acta Math. 110 (1693) 233–278.

[149] T. Tanamoto, Quantum gates by coupled asymmetric quantum dots and controlled-not-gate operation, Phys. Rev. A 61 (2000) 022305.

[150] B.M. Terhal, K.G.H. Vollbrecht, Entanglement of formation for isotropic states, Phys. Rev. Lett. 85 (12) (2000) 2625–2628.

[151] W. Tittel, J. Brendel, H. Zbinden, N. Gisin, Violation of Bell inequalities by photons more than 10 km apart, Phys. Rev. Lett. 81 (17) (1998) 3563–3566.

[152] A.M. Turing, On computable numbers, with an application to the entscheidungsproblem, Proc. London Math. Soc. Ser. 2 42 (1936) 230–265.

[153] V. Vedral, M.B. Plenio, Entanglement measures and purification procedures, Phys. Rev. A 54 (3) (1998) 1619–1633.

[154] V. Vedral, M.B. Plenio, M.A. Rippin, P.L. Knight, Quantifying entanglement, Phys. Rev. Lett. 78 (12) (1997) 2275–2279.

[155] G. Vidal, Entanglement monotones, J. Mod. Opt. 47 (2–3) (2000) 355–376.

[156] G. Vidal, J.I. Latorre, P. Pascual, R. Tarrach, Optimal minimal measurements of mixed states, Phys. Rev. A 60 (1999) 126–135.

[157] G. Vidal, R. Tarrach, Robustness of entanglement, Phys. Rev. A 59 (1) (1999) 141–155.

[158] G. Vidal, R.F. Werner, A computable measure of entanglement, 2001, quant-ph/0102117.

[159] K.G.H. Vollbrecht, R.F. Werner, Entanglement measures under symmetry, 2000, quant-ph/0010095.

[160] K.G.H. Vollbrecht, R.F. Werner, Why two qubits are special, J. Math. Phys. 41 (10) (2000) 6772–6782.

[161] I. Wegener, The Complexity of Boolean Functions, Teubner, Stuttgart, 1987.

[162] S. Weigert, Reconstruction of quantum states and its conceptual implications, in: H.D. Doebner, S.T. Ali, M. Keyl, R.F. Werner (Eds.), Trends in Quantum Mechanics, World Scientific, Singapore, 2000, pp. 146–156.

[163] H. Weinfurter, A. Zeilinger, Quantum communication, in: G. Alber, et al., (Eds.), Quantum Information, Springer, Berlin, 2001, pp. 58–95.

[164] R.F. Werner, Quantum harmonic analysis on phase space, J. Math. Phys. 25 (1984) 1404–1411.

[165] R.F. Werner, Quantum states with Einstein–Podolsky–Rosen correlations admitting a hidden-variable model, Phys. Rev. A 40 (8) (1989) 4277–4281.

[166] R.F. Werner, Optimal cloning of pure states, Phys. Rev. A 58 (1998) 980–1003.

[167] R.F. Werner, All teleportation and dense coding schemes, 2000, quant-ph/0003070.

[168] R.F. Werner, Quantum information theory—an invitation, in: G. Alber, et al., (Eds.), Quantum Information, Springer, Berlin, 2001, pp. 14–59.

[169] R.F. Werner, M.M. Wolf, Bell inequalities and entanglement, Quant. Inf. Comput. 1 (3) (2001) 1–25.

[170] R.F. Werner, M.M. Wolf, Bound entangled gaussian states, Phys. Rev. Lett. 86 (16) (2001) 3658–3661.

[171] H. Weyl, The Classical Groups, Princeton University, Princeton, NJ, 1946.

[172] W.K. Wooters, Entanglement of formation of an arbitrary state of two qubits, Phys. Rev. Lett. 80 (10) (1998) 2245–2248.

[173] W.K. Wootters, W.H. Zurek, A single quantum cannot be cloned, Nature 299 (1982) 802–803.

[174] S.L. Woronowicz, Positive maps of low dimensional matrix algebras, Rep. Math. Phys. 10 (1976) 165–183.