Chapter 8

INFORMATION

INTRODUCTION

The development and the understanding of the basic concepts, methods and results of quantum information theory and of the faithful transmission of quantum information in time and space is the most fundamental problem of quantum information processing. In order to be able to understand and to utilize fully information processing available in nature the concepts of classical information theory need to be expanded to accumulate quantum information carriers. Three central problems concerning quantum information and its communication are dealt with, very briefly, in this chapter.

1. **Quantum information theory.** How to rebuild classical information theory on quantum foundations. How to define fundamental concepts of quantum information theory. How to identify and explore the inherently quantum elements of such a theory.
2. **Quantum transmissions theory.** How to use optimally quantum channels to send classical information and how to use optimally quantum and classical channels to transmit quantum information. How to define and determine the capacity of different quantum channels.
3. **Quantum entanglement theory.** How to quantify and manipulate entanglement and how to produce good entanglement.

LEARNING OBJECTIVES

The aim of the chapter is to learn:

1. the basic concepts of quantum information theory;
2. the basic concepts of quantum transmission;
3. the basic techniques of quantum data compression;
4. the basic techniques of communication through a noisy channel;
5. the basic modes and measures of entanglement;
6. the basic quantum entanglement concentration and purification techniques.

To do good and to cummunicate forget not

Hebrews ch 13, v 16

Quantum information theory and quantum information transmission problems are another area where quantum phenomena are of importance and bring new and surprising potentials, methods and results.

Quantum information theory exhibits on one side strong parallels and on the other side surprising and deep differences with respect to the classical information theory. The main outcomes of the classical information theory are the concepts of the source, channel, methods of encodings of information and results concerning capacities of quantum channels. Quantum information theory deals with quantum analogues of classical information theory concepts, methods and results. However, problems are this time much more complex. For example, there are two quantifiable kinds of information: classical and entanglement; there are at least four types of channel capacities and a variety of measures of entanglement.

In the classical communications we have only one communication primitive: transmission of a classical bit. In the quantum communication there are three communication primitives: transmiting a bit, transmitting a qubit and sharing an EPR-pair of qubits. Entanglement plays the central role in quantum information theory.

There are several fundamental problems to deal with in quantum information theory.

- How to use optimally quantum channels to send reliably classical information?
- How to use quantum and classical channels to transmit reliably quantum states?
- How classical and quantum information interact?
- How to develop qualitative and quantitative theory of entanglement?
- How to develop quantum error correction and fault-tolerant techniques?

With the first four problems we deal briefly in this chapter. Quantum error correction codes and networks as well as quantum computation fault-tolerant techniques were dealt with in Sections 7.4 and 7.5.

A special type of quantum/classical communication, on the basis of quantum teleportation and superdense coding, were dealt with in Section 6.4.

8.1 Quantum Entropy and Information

The basic problems of information theory are how to define such concepts as information and entropy formally, how to store information efficiently and faithfully and how to transmit efficiently information through a noiseless or noisy channel.

Classical information theory solved these basic problems elegantly and achieved enormous practical applicability. Modern communications, space exploration, and very high quality sound reproductions, for example, would be impossible without it.

8.1.1 Basic concepts of classical information theory

There are two basic approaches how to define the amount of information in an object. Shannon's approach defines the amount of information in an object as the number of bits that need to be transmitted in order to select the object from the previously agreed upon set of elements. Kolmogorov/Chaitin's approach considers the amount of information in an object as the number of bits needed to describe (to compress) the object. Theory of information developed using the second approach is called *algorithmic*. Even if some of the first attempts to apply the complexity point of view to physics (see Bennett, 1988a) used ideas close to Kolmogorov/Chaitin's approach so far Shannon's approach have been the basis of the concepts of quantum information theory.

If X is a random variable (a source), which takes a value x with probability $p(x)$, then the **(Shannon) entropy** of X is defined by

$$S(X) = -\sum_x p(x)\lg p(x)$$

and it is considered to be the **information content** of X or a measure of the lack of information about X. The maximum information which can be stored by an n-value variable is then $\lg n$.

It is natural to take as a unit of information the case that $S(X) = 1$, i.e., X can take two values, both with the same probability. A two-valued, or binary, variable can therefore contain one unit of information (called **bit**).

In the special case of binary variable, with two values, 1 with probability p and 0 with probability $1-p$, the **(binary) entropy function** $H(p)$ is defined by

$$H(p) = -p\lg p - (1-p)\lg(1-p).$$

The concepts of conditional probability $p(x|y)$ and mutual probability $p(x,y)$ influenced the definition of conditional and mutual entropy.

If X, Y are two random variables, then the **conditional entropy** $S(X|Y)$ is defined by

$$\begin{aligned} S(X|Y) &= -\sum_y p(y)\sum_x p(x|y)\lg p(x|y) \\ &= -\sum_y\sum_x p(x,y)\lg p(x|y), \end{aligned}$$

where $p(x,y) = p(y)p(x|y)$. $S(X|Y)$ can be seen as a measure of how much information, on average, would remain in X if we were to learn Y.

From the above relations one can easily deduce that

$$S(X|Y) = S(X,Y) - S(Y), \tag{8.1}$$

where $S(X,Y)$ is the information content of X and Y (i.e., the information one gains if, having a priori knowledge of neither, one learns the values of X and Y).

Another important concept is that of **mutual information** $I(X:Y)$ defined by

$$I(X:Y) = -\sum_x\sum_y p(x,y)\lg p(x:y),$$

where $p(x:y)$ is the mutual probability defined by $(p(x)p(y))/p(x,y)$. $I(X:Y)$ is a measure of how much X and Y contain information about each other. One can also say that $I(X:Y)$

is the amount of information about X that is obtained by determining the value of Y. It holds

$$\begin{aligned} I(X:Y) &= S(X) - S(X|Y) \\ &= S(X) + S(Y) - S(X,Y) \end{aligned}$$

and therefore $I(X:Y) = 0$ if and only if X and Y are independent.

Remark 8.1.1 The last identity can be used to describe the trade-off between entropy and information in measurements, if we write the above identity in the form

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

Indeed, consider two random variables X, Y and a measurement of X and Y. $S(X|Y)$ can be seen as the entropy of X after having measured the values of X that become correlated by values of Y. $I(X:Y)$ is the information obtained about X via the measurement of Y. As it follows from the equality $S(X) = S(X|Y) + I(X:Y)$, $S(X|Y)$ and $I(X:Y)$ compensate each other such that $S(X)$ is unchanged. The second law of thermodynamic is not violated in a measurement—in spite of the decrease of $S(X|Y)$ (see Cerf and Adami, 1996).

Several simple inequalities are often useful.

1. Classical data processing inequality

$$X \to Y \to Z \Longrightarrow I(X:Z) \leq I(Y:Z)$$

says that if in a (Markov) process X develops to Y and later to Z, then Y cannot pass to Z more information about X than it has received.

2. If $X = X_1, \ldots, X_m$, $Z = Z_1, \ldots, Z_m$. then

$$S(X|Z) \leq \sum_{i=1}^{m} S(X_i|Z) \leq \sum_{i=1}^{m} S(X_i|Z_i).$$

8.1.2 Quantum entropy and information

The very basic concept of quantum information theory goes back to von Neumann (1927). Suppose that a quantum source (variable) X produces pure states $|\phi_1\rangle, \ldots, |\phi_m\rangle$ with probabilities $p_1, \ldots, p_m$. If ρ_X is the density matrix of the source, then[1]

$$QS(X) = QS(\rho_X) = -\mathrm{Tr}\rho_X \lg \rho_X$$

is called **von Neumann (quantum) entropy**[2] of ρ_X, or of X.[3] Quantum entropy of ρ expresses the degree of ignorance embodied in ρ_X.

[1]In this section notation for quantum analogues of classical information and entropy concepts are derived from the classical ones using the prefix "Q". This Q is often ommitted in literature once the quantum case is clear from the context.

[2]Actually, quantum entropy is defined as $QS = -k_B Tr\rho_X \lg \rho_X$, where k_B is Boltzman constant. By replacing k_B with 1 the entropy becomes dimensionless and has no direct physical meaning.

[3]Suppose that a classical random variable X has probability distribution $p(x)$. Let a quantum system be prepared in the state $|x\rangle$, dictated by the value of X, with the probability $p(x)$. In such a case it holds for the corresponding density matrix ρ that $QS(\rho)$ is an upper bound on the classical mutual information $I(X:Y)$ between X and the result Y of a measurement of an PVOM observable on the system (see Holevo, 1973, and Levitin, 1969).

Observe that in the case the states $|\phi_1\rangle, \ldots, |\phi_m\rangle$ are mutually orthogonal, then ρ is a diagonal matrix. Let $p_1, \ldots, p_m$ be elements of its diagonal. In such a case $QS(\rho) = S(p)$ for $p = (p_1, \ldots, p_m)$, i.e. the quantum entropy is exactly the classical entropy. In general, however, the density matrix ρ has also off-diagonal elements and $QS(\rho_X) \leq S(p)$. Clearly $QS(\rho) = 0$ if and only if ρ is a pure state. and $QS(\rho) \geq 0$ in general. Quantum entropy can be arbitrarily large.

Basic properties of quantum entropy (see Wehrl, 1978): (1) **invariance:** $QS(\rho) = QS(U^*\rho U)$ for any unitary matrix U; (ii) **concavity:** if $\rho = \lambda_1\rho_1 + \lambda_2\rho_2$, $\lambda_1 + \lambda_2 = 1$, $\lambda_1 \geq, \lambda_2 \geq 0$, then $QS(\rho) \geq \lambda_1 QS(\rho_1) + \lambda_2 QS(\rho_2)$; (iii) **additivity:** $QS(\rho_1 \otimes \rho_2) = QS(\rho_1) + QS(\rho_2)$; (iv) **subadditivity**: if ρ is a density matrix of a bipartite system $A \otimes B$, $\rho_1 = TR_B\rho$, $\rho_2 = TR_A\rho$, then $QS(\rho) \leq QS(\rho_1) + QS(\rho_2) \leq QS(\rho_1 \otimes \rho_2)$; (v) **lower bound**: $S(\rho) \geq -\lg \lambda_1$, where λ_1 is the largest eigenvalue of ρ.

The relative quantum entropy $S(\rho_1|\rho_2)$ of two density matrices is defined as $S(\rho_1|\rho_2) = Tr(\rho_1(\lg \rho_1 - \lg \rho_2))$. For all density matrices ρ_1, ρ_2, $S(\rho_1|\rho_2) \geq 0$.

Though the concept of von Neumann quantum entropy has already been a standard one a for a long time a fully developed theory of quantum information does not yet exist. Only quite recent are attempts to provide quantum analogues of other basic concepts of information and entropy (see Cerf and Adami, 1996). Their approach will now be dealt with.

The **quantum conditional entropy** $QS(X|Y)$ is defined by a natural generalization of the classical case as

$$QS(X|Y) = -\mathrm{Tr}_{XY}[\rho_{XY} \lg \rho_{X|Y}],$$

where

$$\rho_{X|Y} = \lim_{n\to\infty} [\rho_{XY}^{\frac{1}{n}} (\mathbf{I}_X \otimes \rho_Y)^{-\frac{1}{n}}]^n \tag{8.2}$$

which should be an analogue to the conditional probability $p(x|y)$. In (8.2) $\mathbf{I}_A$ is the unit matrix in the Hilbert space for X and $\rho_Y = Tr[\rho_{XY}]$ denotes a "marginal" density matrix—an analogue to the "marginal" probability $p_y = \sum_x p(x,y)$.

The above definition of $QS(X|Y)$ allows to show that an analogue of the identity 8.1 holds also for quantum entropy.[4]

$$QS(X|Y) = QS(X,Y) - QS(Y).$$

In spite of the apparent similarity between quantum $QS(X|Y)$ and classical $S(X|Y)$ the fact that in the quantum case we deal with (density) matrices, rather than with numbers, as in the classical case, brings quite a different situation for quantum information theory and potential far exceeding the classical one.

The main point is that while $p(x|y)$ is a probability distribution in x (i.e. $0 \leq p(|y) \leq 1$), its quantum analogue $\rho(X|Y)$ is not a density matrix. It is Hermitian and positive but its eigenvalues can be larger than 1 and, consequently, the conditional entropy can be negative. This helps to explain the well-known fact that quantum entropy is non-monotonic, and it can be the case that $QS(X,Y) < QS(Y)$, i.e. the quantum entropy of the entire system can be smaller than the entropy of one of its subparts (what is not possible in the classical case). This happens, for example, in the case of quantum entanglement as shown in the following example.

[4]The quite complicated expression in (8.2), and the fact that a limit has to be used are due to the fact that the joint density matrix (ρ_{XY}) and the marginal matrices $\mathbf{I}_X \otimes \rho_Y$ do not commute in general. If they do commute, the whole expression gets much simpler, as discussed later.

Example 8.1.2 *(Cerf and Adami, 1996) Consider the Bell state* $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ *of the Hilbert space* $H_{AB} = H_A \otimes H_B$*, where* $H_A = H_B = H_2$*. The density matrix* $\rho_{AB} = |\psi\rangle\langle\psi|$ *is shown in Figure 8.1a.* $\rho_A = \rho_B = Tr_{H_B}[\rho_{AB}] = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$*, see Figure 8.1b.*

$$\rho_{AB} = \begin{pmatrix} 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{pmatrix} \quad \rho_A = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \quad \rho_{A|B} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

(a) (b) (c)

Figure 8.1: Density matrices

Hence $QS(A) = QS(B) = 1$*. The density matrix* $\rho_{A|B} = \rho_{AB}(\mathbf{I}_A \otimes \rho_B)^{-1}$ *(because in this case the joint and marginal matrices* ρ_{AB} *and* $I_A \otimes \rho_B$ *commute) is shown in Figure 8.1c. Hence* $QS(AB) = QS(B) + QS(A|B) = 1 - 1 = 0$*, because* $QS(A|B) = -1$*.*

Quantum mutual information $QI(X : Y)$, as an analogue of the classical mutual information $I(X : Y)$ is defined on the base of "mutual" density matrix

$$\rho_{X:Y} = \lim_{n\to\infty}[(\rho_X \otimes \rho_Y)^{\frac{1}{n}} \rho_{XY}^{-\frac{1}{n}}]^n.$$

and the definition implies the standard relation

$$QI(X : Y) = QS(X) - QS(X|Y) = QS(X) + QS(Y) - QS(XY).$$

Observe that this definition is reduced to the classical one for the case ρ_{XY} is a diagonal matrix.

However, not all basic equalities and inequalities of the classical information theory transfer from the classical to the quantum case. For example, in the classical case we have

$$I(x : y) \leq \min[S(x), S(y)],$$

but in the quantum case the best upper bound possible is

$$QS(X : Y) \leq 2\min[QS(X), QS(Y)].$$

8.2 Quantum Channels and Data Compression

The question which is the most efficient way to code, to store and to transfer information is of large theoretical and especially practical interest. Coding is actually the central problem of the classical information theory.

8.2.1 Quantum sources, channels and transmissions

A classical (discrete, memoryless) channel is described by conditional probabilities p_{ij} of the output j for the input i. A quantum channel is described by a superoperator.

Let us now introduce quantum analogues of the classical concepts of the source and channel. A quantum **source** produces mixed states $\rho_1, \rho_2, \ldots$ with probabilities $p_1, p_2, \ldots$. (one also says it is an ensemble of mixed states each produced with certain probability.)

A **quantum noiseless channel** is any quantum system (sender) with the following properties:

1. it can be in any state of a finite dimensional Hilbert space;

2. it is entangled with some other (similar) quantum system (receiver);

3. it remains stable in this entangled state (while enroute the sender and the receiver).

In a similar way as a sequence of n bits can be used to transmit through a classical channel any of the 2^n messages from $\{0,1\}^n$, a sequence of n qubits can be used to transmit any state of H_{2^n}.

A **noisy quantum channel** is a quantum channel that interacts unitarily with an environment while enroute from the sender to the receiver. A noiseless channel is a special case of noisy channel. An equivalent way to define a noisy channel, as already mentioned, is that of an superoperator.

Quantum channels have a variety of implementations. A quantum channel can be a regular communication channel, for example an optic fibre, which gets on input a pure state $|\psi\rangle$, preserves coherence to some degree, and produces a mixed state $[\phi\rangle$ on the output. However, a quantum channel can be just "a flow of time" for a qubit "sitting in a quantum memory and interacting with its environment". It can also be the result of the operation of a noisy quantum gate, see Figure 8.2.

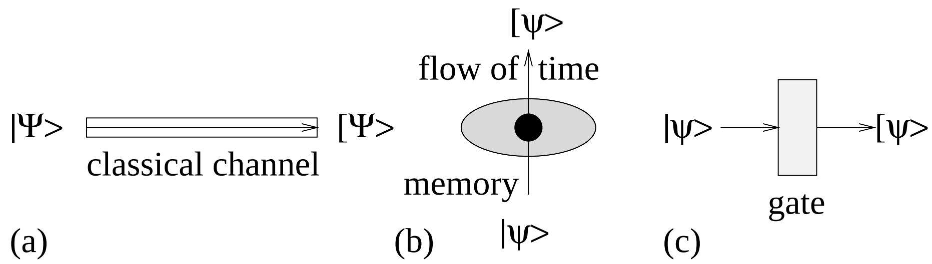

Figure 8.2: Quantum channels-80%

.

Transmissions through quantum channels

A quantum channel can be seen as a mean to transmit to the receiver states $\rho_i, i = 1,2,\ldots$, produced by the source (or sent by the sender). Three cases are of special interest.

If the states ρ_i of the quantum source are orthogonal, then the source and the channel can be considered as classical. Indeed, in such a case complete information about states produced by the source can be extracted by a measurement of the sender, transmitted classically to the receiver who can then make arbitrarily many copies of the source states.

If the source states are pure and nonorthogonal, then no projection measurement can extract full information about a state. In addition, whenever a source state is transmitted through a quantum channel at most one faithful copy of the source state can be produced, and only if no faithful copy remains at the sender.

An interesting intermediate case is that the source states are nonorthogonal but commuting mixed states. Such source states ρ_i can be broadcast in the following sense. The two systems, A and B—that of sender and receiver—can be prepared in the joint state $\rho_i(AB)$ which is not a clone of the source states, i.e., $\rho_i(AB) \neq \rho_i(A) \otimes \rho_i(B)$, but can be obtained by tracing out each of the subsystems. Namely,

$$\rho_i = Tr_A \rho_i(AB) = Tr_B \rho_i(AB).$$

If density matrices of the source states do not commute, then the source can neither be cloned nor broadcasted (see Barnum et al. 1996), Bennett (1998a).

If a quantum channel is to transmit nonorthogonal states faithfully, it must operate on states that are sent through the channel "blindly" (in an "oblivious way")—without learning anything about them. This is due to the fact that quantum information can be neither read nor copied without disturbance.

Evolution of any quantum system can be seen as being done in the quantum channel of the time flow. This evolution remains deterministic till some information starts to leak into the environment. If this happens the channel gets noisy. The quantum state of the system gets randomized and entangled with the environment.

Transmission of an unknown quantum state requires the following resources.

1. A quantum resource (that cannot be cloned).
2. A directed communication resource (that cannot propagate faster than light).

There are two basic ways of transmission of quantum states. A direct unrestricted transmission of qubits at which the same particle provides both resources and functions. An indirect one, through teleportation, see Section 6.4, at which both functions are performed by an entangled pair of particles and by a communication through the classical channel.

Fidelity of transmissions

An important issue is how to measure quality of transmissions. (There are at least two reasons why a transmission will generally be imperfect: (i) data compression is needed because of the limited resources; (ii) noisy quantum channel corrupts the state being transmitted.) If source states are pure and a quantum channel produces on each pure input state $|\phi_i\rangle$, produced with probability p_i, an output (mixed state) W_i, then the quality of transmissions is measured by the fidelity (Jozsa and Schumacher, 1994)

$$F = \sum_i p_i \langle \phi_i | W_i | \phi_i \rangle. \tag{8.3}$$

F is the probability that a channel output state passes a test (conducted by someone who knows the inputs) for being the same as the input.

In the case the input states ρ_i are mixed the fidelity is defined as follows:

$$F = \sum_i p_i \left(Tr \left(\sqrt{\sqrt{\rho_i} W_I \sqrt{\rho_i}} \right) \right)^2 . \tag{8.4}$$

As shown by Uhlmann (1976) and Jozsa (1994), (8.4) is a natural generalization of (8.3) and represents maximum of (8.3) over all purifications $|\phi_i\rangle$ of ρ_i. More formally, if we define $F(\rho_1, \rho_2) = (Tr\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}})^2$, then $F(\rho_1, \rho_2) = \max\{|\langle \phi_1 | \phi_2 \rangle|^2 \mid \phi_i \text{ is a purification of } \rho_i\}$.

Exercise 8.2.1 (Jozsa (1994)) *Show the following properties of the fidelity function defined above: (a) $0 \le F(\rho_1, \rho_2) \le 1$; (b) If $\rho_1 = |\phi\rangle\langle\phi|$ is a pure state, then $F(\rho_1, \rho_2) = \langle \phi | \rho_i | \phi \rangle$; if $\rho_1, \rho_2 \ge 0$, $p_1 + p_2 = 1$, then $F(\rho, p_1\rho_1 + p_2\rho_2) \ge p_1 F(\rho, \rho_1) + p_2 F(\rho, \rho_2)$; (c) $F(\rho_1 \otimes \rho_2, \rho_3 \otimes \rho_4) = F(\rho_1, \rho_3) F(\rho_2, \rho_4)$; (d) $F(\rho_1, \rho_2)$ is preserved by unitary transformations.*

A quantum channel will be considered as faithful if, in an appropriate limit, the expected fidelity of transmission tends to unity. In other words the chance to distinguish channel outputs from inputs by any quantum measurement should tend to zero.

Remark 8.2.2 It may happen that a quantum state to be transmitted is entangled with some other state, We may now be intertested that entanglement is preserved by the transmission. Because of that it was not clear how to define capacity of quantum channels. To deal with the problem of the quality of transmission of entanglement the concept of **entanglement fidelity** has been introduced by Barnum et al. (1998) and they have shown the equivalence of the concept of quantum capacity based on entanglement fidelity and transmission fidelity—to be discussed later.

8.2.2 Shannon's coding theorems

The concept of entropy $S(X)$ of a random classical source (variable) X allows to state in a simple and elegant way one of the basic results of the classical data compression, which solves the following fundamental transmission problem: how many bits are needed to transmit through a noiseless channel, n values of X. The answer is elegant: $nS(X)$ bits. (They are sufficient in the limit and one cannot do better.) More precisely:

Theorem 8.2.3 (Shannon's noiseless coding theorem) *If $x_1, \dots, x_n$ are independently obtained values of a random source X and x_i is obtained with probability p_i, then any string $x_1 \dots x_n$ can be compressed to a $nS(p)$-bit string, $(p = (p_1, \dots, p_n))$—and no more–in the following asymptotic sense.*

For any $\varepsilon, \delta > 0$, for sufficiently large n, for any $\lambda(n) \geq n[S(p) + \delta]$, $\lambda(n) \in \{1, \dots, n\}$, there exists an encoding (compression method) that encodes (compresses) each binary string $x_1 \dots x_n$ to $y_1 \dots y_{\lambda(n)}$, in such a way that $x_1 \dots x_n$ can be successfully recovered from $y_1 \dots y_{\lambda(n)}$, with probability greater than $1 - \varepsilon$. Moreover, the above compression method produces the minimal compression in the sense that for any $\varepsilon, \delta > 0$, for sufficiently large n, and $\lambda(n) \leq n[S(p) - \delta]$, for any compression method that maps $x_1 \dots x_n$ into $y_1 \dots y_{\lambda(n)}$, the probability that $x_1 \dots x_n$ can be successfully recovered from $y_1 \dots y_{\lambda(n)}$ is less than $1 - \varepsilon$.

Since one of the main aims of this section is to present an analogous result for quantum data compression, let us discuss the basic ideas behind the proof of Shannon's theorem for the case of binary variables (Steane, 1997).

Let us assume that Alice wants to communicate to Bob a sequence X of n values $x_1, \dots, x_n$ of a binary variable X that takes the value 1 with probability p and the value 0 with the probability $1 - p$. The mean number of 1's in X is in such a case np.

If np is an integer, then the probability of obtaining a binary sequence of length n with np 1's is

$$p^{np}(1-p)^{n-np} = 2^{-nH(p)} \tag{8.5}$$

Exercise 8.2.4 *Show in detail the equality (8.5).*

The result in (8.5) naturally leads to the following encoding strategy. A sequence X of n binary values is said to be an ε-**typical sequence**, $\varepsilon > 0$, if its probability $p(X)$ satisfies the inequality

$$2^{-n(H(p)+\varepsilon)} \leq p(X) \leq 2^{-n(H(p)-\varepsilon)}.$$

It can be shown that the probability that the sequence Alice wants to send is an ε-typical sequence is greater than $1-\varepsilon$, for sufficiently large n, no matter how small ε is.

The above facts lead naturally to the following communication decision. Alice does not need to communicate X to Bob directly. It is enough that Alice sends Bob information which of the typical sequences she wants to send. (Of course, they have to agree beforehand how to number all typical sequences.) Alice therefore sends only the number of the typical sequence instead of the sequence itself (and any nontypical , unlikely sequence, directly).

How good is the method? It can be shown that all of $2^{nH(p)}$ typical sequences have the same probability. In order to communicate a number of one of them, it is sufficient to send $nH(p) < n$ bits. In addition, there is no way for Alice to do better because all typical sequences have the same probability.

Shannon's noiseless theorem provides theoretical limitations how well can (classical) data be compressed. The next important task is to develop simple methods that can either achieve optimality or to get close to it.

Example 8.2.5 (Huffman code) *For $n = 4$ and $p = \frac{1}{4}$ the best possible data compression, according to Shannon's theorem, requires sending on average $4H(\frac{1}{4}) = 3.245$ bits for communicating 4 values of a binary variable X. A simple and practical method known as Huffman code, see table 8.3, in which less probable sequences are encoded by longer words and more probable sequences by shorter words, requires sending on average 3.273 bits per message.*

message	code	message	code	message	code	message	code
0000	10	0100	010	1000	011	1100	11101
0001	000	0101	11001	1001	11011	1101	111110
0010	001	0110	11010	1010	1100	1110	111101
0011	11000	0111	1111000	1011	111111	1111	1111001

Figure 8.3: Huffman code for $n = 4$

Remark 8.2.6 Shannon's noiseless coding theorem also provides an interpretation of Shannon entropy $S(X)$ as the mean number of bits necessary to code the outputs of a random variable X using an ideal code.

8.2.3 Schumacher's noiseless coding theorem

The basic quantum data compression problem and its solution are quantum analogues of the problem discussed in the previous section. This time we have a quantum source X producing a mixed state with the density matrix ρ and the task is to find an achievable limit on the number of qubits needed to transmit n pure states produced by the source. The answer is that $nQS(\rho)$ qubits are needed and sufficient in average to transmit n states of the source (Schumacher (1995). Let us now discuss both the problem and its solution in more detail, adopting the proof and example of Jozsa and Schumacher (1994).

Let a source X produce pure states (as signals) $|\psi_1\rangle, \ldots, |\psi_n\rangle$ in H_n, with probabilties $p_1, \ldots, p_n$. In addition, let the corresponding density matrix of the source X be

$$\rho = \sum_{i=1}^{n} p_i |\psi_i\rangle\langle\psi_i|.$$

Example 8.2.7 *Let the source produce states $|\phi\rangle$ and $|\psi\rangle$ both with probability $\frac{1}{2}$. Let $|\phi\rangle = \sqrt{0.9}|a\rangle + \sqrt{0.1}|b\rangle$ and $|\psi\rangle = \sqrt{0.9}|a\rangle - \sqrt{0.1}|b\rangle$, where $\{|a\rangle, |b\rangle\}$ is an orthonormal basis of H_2 . The density matrix of the source is $\begin{pmatrix} 0.9 & 0 \\ 0 & 0.1 \end{pmatrix}$ and $S(\rho) = H(0.9) = 0.469$.*

The basic idea of the encoding-transmission-decoding process goes as follows: n outputs (signals) of the quantum source X with the density matrix ρ are encoded by a state in a Hilbert space of a smaller dimension (than that of n-signals), transmitted and decoded into n signals again—this time of another source X' with the density matrix ρ'. It may happen that some information is lost in the encoding process, or during the transmission, and therefore we have to expect that in general $\rho \neq \rho'$. For the whole encoding-transfer-decoding process to be successful we require that their transmission fidelity $F(\rho, \rho')$ is as close to 1 as required.

It can be shown that for any $0 < \varepsilon \leq 1$ there is an encoding/decoding scheme requiring for transmission slightly more than $nQS(\rho)$ qubits with transmission fidelity $1-\varepsilon$. This can be done as follows (if we restrict ourselves to the case that the pure states of the source are qubits).

If n signals are to be transmitted, then the total composed state of n symbol is represented by a vector of 2^n-dimensional Hilbert space. However, in case $QS(\rho) < 1$, it is very likely (and the probability goes to 1 for $n \to \infty$), that in any given realization the total state of n signals actually falls in a **typical subspace** of the Hilbert space of the composed block of signals (an analogue of typical sequences from the classical case). It has been shown by Jozsa and Schumacher (1994), that the dimension of the typical subspace is $2^{nQS(\rho)}$. As a consequence only $nQS(\rho)$ qubits are required to represent quantum information in n signals.

Example 8.2.8 *For blocks $s_1s_2s_3$ of three signals of the source X from Example 8.2.7 let us encode s_1 and s_2 using the mapping $|a\rangle \to |0\rangle$, $|b\rangle \to |1\rangle$ and let us ignore s_3. During the decoding process let us decode s_1 and s_2 fully and s_3 always as $|a\rangle$. Encodings are therefore into a 4-dimensional space and the fidelity of this encoding/decoding scheme is 0.9.*

Theorem 8.2.9 (Schumacher's noiseless coding theorem) *For any quantum source X with the density matrix ρ and any $\varepsilon, \delta > 0$ it holds:*

(a) If $S(\rho) + \delta$ qubits are available per signal, then for large n, there exists a coding/decoding scheme of fidelity $F \geq 1 - \varepsilon$ for strings of the signals of length n.

(b) If $S(\rho) - \delta$ qubits are available per signal, then, for any encoding/decoding scheme, some strings of length n will be decoded with the fidelity $F < \varepsilon$ for n sufficiently large.

Proof. We assume a quantum source X producing pure states (sequences of pure states) $|\psi_1\rangle, \ldots, |\psi_m\rangle$ of a Hilbert space H_n with the probability p_i for $|\psi_i\rangle$. The density matrix ρ of X is $\rho = \sum_{i=1}^{m} p_i |\phi_i\rangle\langle\phi_i|$. In the proof we use two lemmas.

Lemma 8.2.10 *Let $|\phi_i\rangle\langle\phi_i| \longleftrightarrow W_i$, $1 \leq i \leq m$, be an association of density matrices to signals, where each W_i is a density matrix of a mixed state with pure states over a d-dimensional subspace D of H_n. Let the sum of d largest eigenvalues of ρ be θ. Then the fidelity of transmissions $|\phi_i\rangle \longrightarrow W_i$ is at most θ.*

Proof of Lemma 8.2.10. Since each W_i is "supported" by D there is, for each $1 \leq i \leq m$, an orthonormal basis $|\psi_1^{(i)}\rangle, \ldots, |\psi_d^{(i)}\rangle$ of D consisting of eigenvectors of W_i such that

$$W_i = \sum_{j=1}^{d} g_j^{(i)} |\psi_j^i\rangle\langle\psi_j^{(i)}|,$$

where $0 \leq g_j^{(i)} \leq 1$. Denote by Γ the projection into D. Hence for any $1 \leq i \leq m$, $\Gamma = \sum_{j=1}^{d} |\psi_j^{(i)}\rangle\langle\psi_j^{(i)}|$. In addition, for any $1 \leq i \leq m$, we have

$$\langle\phi_i|W_i|\phi_i\rangle = \sum_{i=1}^{d} g_j^{(i)} \langle\phi_i|\psi_j^{(i)}\rangle\langle\psi_j^{(i)}|\phi_i\rangle = Tr(|\phi_i\rangle\langle\phi_i|\Gamma).$$

Hence, for the fidelity F it holds

$$F \leq \sum_{i=1}^{m} p_i\, Tr(|\phi_i\rangle\langle\phi_i|\Gamma) = Tr(\rho\Gamma).$$

Let now $|e_1\rangle, \ldots, |e_n\rangle$ be an orthonormal basis of eigenvectors of ρ with the corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Then

$$Tr(\rho\Gamma) = \sum_{i=1}^{n} \lambda_i \langle e_i|\Gamma|e_i\rangle,$$

where $0 \leq \langle e_i|\Gamma|e_i\rangle \leq 1$, and $\sum_{i=1}^{n} \langle e_i|\Gamma|e_i\rangle = Tr(\Gamma) = d$, see Exercises 9.2.24. From the minimax properties of eigenvalues (see Kato, 1978), this implies that $Tr(\rho\Gamma) \leq \theta$.

Observe that the above upper bound for $Tr(\rho\Gamma)$ is obtained if D is the subspace spanned by d eigenvectors corresponding to d largest eigenvalues. ▯

The lemma claims that for a fixed d we get the highest fidelity by taking as D the space generated by eigenvectors corresponding to d largest eigenvalues. Let us denote by D_X such a subspace of H_n. We present now an encoding-decoding method based on the choice od D as D_X.

Let $|0\rangle$ be any fixed state in D_X and let $D_X^{\perp}$ be the orthogonal complement of D_X in H_n. For each $|\phi_i\rangle$ we construct W_i as follows:

If $|\phi_i\rangle = \alpha_i|l_i\rangle + \beta_i|r_i\rangle$, where $|l_i\rangle \in D, |r_i\rangle \in D_X^{\perp}$ are unit vectors, then

$$W_i = |\alpha_i|^2 |l_i\rangle\langle l_i| + |\beta_i|^2 |0\rangle\langle 0|. \tag{8.6}$$

Technically, W_i can be designed by applying observable P_{D_X} on $|\phi_i\rangle$ and, if the result is 0, i.e. $|\psi_i\rangle$ projects to $D_X^{\perp}$, $|0\rangle$ is taken as the post-measurement state.

Lemma 8.2.11 *Suppose that the sum of the d largest eigenvalues of ρ is greater than $1-\varepsilon$. Then the coding/decoding process defined by the mapping $|\phi_i\rangle \longleftrightarrow W_i$ has fidelity $F > 1-2\varepsilon$.*

Proof. On the base of equalities and inequalities

$$\langle\phi_i|W_i|\phi_i\rangle = |\alpha_i|^2\langle\phi_i|l_i\rangle^2 + |\beta_i|^2|\langle\phi_i|0\rangle|^2 \geq |\alpha_i|^2|\langle\phi_i|l_i\rangle|^2 = |\alpha_i|^4 \geq 2|\alpha_i|^2 - 1$$

we get

$$F = \sum_{i=1}^{m} p_i \langle \phi_i | W_i | \phi_i \rangle \geq 2 \sum_{i=1}^{m} (p_i |\phi_i|^2) - 1 = 2\,Tr(\rho P_{D_X^{\perp}}) - 1 \geq 1 - 2\varepsilon,$$

if we make use of the fact that $\sum_{i=1}^{m} p_i |\alpha_i|^2 = Tr(\rho P_(D_X^{\perp})$ is the sum of d largest eigenvalues and therefore it is larger than $1 - \varepsilon$. □

To finish the proof of Schumacher's theorem let us formulate another of its ingredients—a quantum modification of the classical result on typical sequences mentioned in Section 8.2.2.

Let $\lambda = \{\lambda_1, \ldots, \lambda_m\}$ be probabilities of X and let for any $k \in \mathbf{N}$, Λ_k be the set of probabilities (if elements of k-tuples are multipied and taken to form a multiset) of X^k. Let the probability of any subset of X_k be the sum of probabilities of its members. The result on typical sequences we will use has the form (Jozsa and Schumacher, 1994):

Lemma 8.2.12 *Let $\varepsilon, \delta > 0$. (1) For sufficiently large k, the set X_k may be partitioned into a subset L of "likely sequences" with at most $2^{k(QS(X)+\delta)}$ elements, which has probability greater than $1 - \varepsilon$, and into its complement of "unlikely sequences", which has the probability smaller than ε.*

(2) Any subset of X_k, which has less than $2^{k(QS(\rho)-\delta}$ elements, has probability smaller than $2^{k(QS(\rho)-\delta)}$.

Continuation of the proof of Schumacher's theorem. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of the density matrix ρ of X. $\lambda_1, \ldots, \lambda_n$ can be seen as forming a probability distribution Λ as considered above. In this case $QS(\rho) = S(p)$ for $p = (\lambda_1, \ldots, \lambda_n)$. The k-block version X^k has the density matrix $\rho_k = \bigotimes_{i=1}^{k} \rho$ and its eigenvalues are products of k-tuples of vectors from X^k.

Suppose now that $\varepsilon, \delta > 0$ are given. From the part (1) of Lemma 8.2.12, it follows that for all sufficient large k there is a set of $2^{k(QS(\rho)+\delta)}$ eigenvalues of ρ_k with the total sum greater than $1 - \frac{1}{2}\varepsilon$. Hence the sum of $2^{k(QS(X)+\delta)}$ largest eigenvalues is also greater than $1 - \frac{1}{2}\varepsilon$. By Lemma 8.2.11 there exists a coding scheme for X^k using $k(QS(\rho_k) + \delta)$-qubits per signal of X^k, i.e., $(QS(\rho) + \delta)$ qubits per signal of X, with fidelity $F > 1 - \varepsilon$, for k signals. This proves the part (a) of Schumacher's theorem.

From the point (2) of Lemma 8.2.12 it follows that for sufficiently large k any subset of X^k of the size less than $2^{k(QS(\rho)-\delta)}$ has probability less than ε; that is the sum of the $2^{k(QS(\rho)-\delta)}$ largest eigenvalues of ρ_k has the sum less than ε. By Lemma 8.2.10 every encoding/decoding scheme for X^k using $k(QS(\rho) - \delta)$ qubits per symbol of X^k, that is $QS(\rho) - \delta$ qubits per signal of X, will have fidelity less than ε for k signals. Since this holds for all sufficiently large k, this proves the part (b) of Schumacher's theorem. □

Example 8.2.13 *Let us consider again the source from Example 8.2.7 and let H_a be its 4-dimensional subspace spanned by the following set of three qubit states $\{|aaa\rangle, |aab\rangle, |aba\rangle, |baa\rangle\}$ having majority of as. Let us use the following encoding/decoding scheme:*

Encoding. *Let the states in H_a be encoded as follows:*

$$|aaaa\rangle \rightarrow |00\rangle, \quad |aab\rangle \rightarrow |01\rangle, \quad |aba\rangle \rightarrow |10\rangle, \quad |baa\rangle \rightarrow |11\rangle \tag{8.7}$$

and states not in H_a as $|00\rangle$.

Decoding. *Each two qubit state is decoded as the corresponding state in H_a, following encodings (8.7).*

Analysis. *Observe that due to our encoding methods none of the actual triples* $s_1s_2s_3$ *of signals is encoded perfectly because none of them lies in* H_a. *In spite of that we get high fidelity for the overall encoding/decoding scheme. Let* Γ *be the projection into* H_a. *The density matrix* ρ_a *has a diagonal form with elements*

$$(0.9^3, 0.9^2 \cdot 0.1, 0.9^2 \cdot 0.1, 0.9 \cdot 0.1^2, 0.9 \cdot 0.1^2, 0.9 \cdot 0.1^2, 0.9 \cdot 0.1^2, 0.1^3)$$

and therefore $Tr(\rho\Gamma) = 0.9^3 + 3(0.9)^2 0.1 = 0.972 = 1 - 0.18$. *By Lemma 8.2.10, this encoding/decoding scheme has fidelity at least* $0.964 = 1 - 2 \cdot 0.18$.

Remark 8.2.14 The first attempts to develop quantum information theory were based on applying classical information theory to probabilities derived from the representations of quantum mixed states. Schumacher's theorem has actually been the beginning of a new approach, in which quantum information theory is based on concepts, measures and codings that are inherently quantum.

Remark 8.2.15 Observe that quantum data compression has the following remarkable property: it allows to compress and expand each of the 2^n of distinct sequences of nonorthogonal states with fidelity approaching 1 even though the sequences cannot be reliably distinguished from one another by any measurement.

8.2.4 Dense quantum coding

There is a special way, due to Ambainis et al. (1998), presented in the following, how to encode m bits into $n < m$ qubits, for n not too much smaller than m, in such a way, which has no classical analogue, that each bit, but not all, can be retrieved with significant probability.

The Holevo theorem (see Holevo, 1973) says that no more than m bits can be transmitted (faithfully) by transferring m qubits. Codings described below are therefore quite surprising, even if they do not contradict Holevo theorem because even if each of m bits can be retrieved from $n < m$ qubits with significant probability, not all can be retrieved. These codings make use of special properties of quantum projection measurements. Namely, the fact that projection measurements are not commutative and a particular measurement can destroy information that could be revealed by other measurement.

Encodings discussed in the following have three parameters, n, m and p (notation $m \stackrel{p}{\to} n$). They consist of mapping m bits into n qubits and, in addition, in providing m measurements $\mathcal{O}_i$, $1 \leq i \leq m$, such that using the measurement $\mathcal{O}_i$ the ith bit can be retrieved from encoded qubits with probability p. Since such encodings use a random number as a parameter, we talk about random access encodings.

Definition 8.2.16 *A* $m \stackrel{p}{\to} n$ **random access encoding** *is a function* $f : \{0,1\}^m \times \mathcal{R}_{\{0,1\}} \to H_n$ *such that for any* $1 \leq i \leq m$ *there is a projection measurement* $\mathcal{O}_i$ *which when applied to the value-state of* f *returns* 0 *or* 1 *and such that for any* $b \in \{0,1\}^m$, $Pr(\mathcal{O}_i(f(b,r)) = b_i) \geq p$, *where* $\mathcal{R}$ *is a source of random bits (*f *is called the encoding function and* $\mathcal{O}_i$ *are decoding mappings).*

Example 8.2.17 *(A* $2 \stackrel{0.85}{\to} 1$ *encoding.) Let* $u_0 = |0\rangle$, $u_1 = |1\rangle$, $|v_0\rangle = |0'\rangle$, $|v_1\rangle = |1'\rangle$ *and* $f(x_1, x_2) = \frac{1}{\sqrt{2+\sqrt{2}}}(|u_{x_1}\rangle + |v_{x_2}\rangle)$. *The mapping* f *has the desirable properties provided the first (second) bit is measured with respect to the standard (dual) basis. (This follows from the fact that the distance between the codeword and the right subspace is* $\frac{\pi}{8}$ *and the access probability is therefore* $\cos^2(\frac{\pi}{8}) \approx 0.853$.)

A classical bit-to-bit $m \stackrel{p}{\to} n$ encoding can be defined similarly and it has been proved by Ambainis et al. (1998) that no $2 \stackrel{p}{\to} 1$ classical random access encoding exists if $p > \frac{1}{2}$. The potential of classical $m \stackrel{p}{\to} n$ encodings is already well understood.

Lemma 8.2.18 *If $\frac{1}{2} < p \leq 1$, then $n \geq (1 - H(p))m$ for any classical $m \stackrel{p}{\to} n$ encoding.*

Proof. Let f be a (possibly probabilistic) encoding mapping randomly and uniformly chosen $X = X_1 \dots X_m$ from $\{0,1\}^m$ into $Y = f(X) \in \{0,1\}^n$. In addition, let $Z = Z_1 \dots Z_m$ be random variables with values in $\{0,1\}^m$ obtained using the decoding. Now the claim of the lemma follows from two general series of inequalities concerning Shannon entropy and mutual information.

$$n \geq S(Y) \geq I(X:Y) = S(X) - S(X|Y) = m - S(X|Y),$$

$$S(X|Y) \leq S(X|Z) \leq \sum_{i=1}^{m} S(X_i|Z) \leq \sum_{i=1}^{m} S(X_i|Z_i)$$

and from the inequality $S(X_i|Z_i) \leq H(p)$. ∎

It has been shown by Ambainis et al. (1998) that the bound in Lemma 8.2.18 is pretty good. Namely, there is a classical $m \stackrel{p}{\to} n$ encoding with $n = (1-H(p))m + \mathcal{O}(\lg n)$ if $p > \frac{1}{2}$. In the quantum case it holds.

Theorem 8.2.19 *If there is a $m \stackrel{p}{\to} n$ quantum encoding where $p > \frac{1}{2}$, then $n > \Omega(\frac{m}{\lg m})$.*

Theorem 8.2.19 is a corollary of the following two lemmas:

Lemma 8.2.20 *If there is an $m \stackrel{p}{\to} n$ encoding for some $p > \frac{1}{2}$, then there is also an $m \stackrel{1-\varepsilon}{\to} \mathcal{O}(n \lg \frac{1}{\varepsilon})$ encoding for any $\varepsilon > 0$.*

Proof. Given an encoding $f : \{0,1\}^m \times \mathcal{R} \to H_n$ with decodings $\mathcal{O}_i$, $1 \leq i \leq m$ and the success probability $p > \frac{1}{2}$, let for $t \in \mathbf{N}$, $f_t\{0,1\}^m \times \mathcal{R}^t \to \bigoplus_{i=1}^n H_n$ be a new encoding defined by $f_t(x, r_1, \dots, r_t) = \bigoplus_{i=1}^n f(x, r_i)$—the tensor product of t independent copies of the original code. Let the new decoding $\mathcal{O}_i'$ consist of applying $\mathcal{O}_i$ to each of t copies of the code and taking the majority of the outcomes as the final 0/1 result. By applying one of Chernoff's bounds we get that the error probability decreases exponentially with t and it is therefore at most ε if $t = \mathcal{O}(\lg \frac{1}{\varepsilon})$. ∎

Lemma 8.2.21 *If there is an $m \stackrel{1-\varepsilon}{\to} n$ quantum encoding with $\varepsilon < \frac{1}{64m^2}$, then $n = \Omega(m)$.*

For the proof, which is quite technical, see Ambainis et al. (1998).

Open problem 8.2.22 *How good is lower bound of Theorem 8.2.19?*

The existence of $n \stackrel{0.85}{\to} n$ encoding follows from Example 8.2.17. Encoding $3n \stackrel{0.79}{\to} n$ is discussed in the following exercise.

Exercise 8.2.23 *Design a $3 \stackrel{0.79}{\to} 1$ encoding. (Hint: make eight vertices of a maximal cube embedded in the Bloch sphere to encode 3 bits. For decoding use: standard, dual and circular basis.)*[5]

Lemma 8.2.18 and Theorem 8.2.19 show the gap between the classical and quantum random access encodings.

8.2.5 Quantum Noisy Channel Transmissions

For quantum information to be useful, it must be transmitted and processed reliably. The fundamental limits on amount of quantum information that can be transmitted reliably along a noisy channel are therefore of large importance.

Classical channel capacity is the greatest number C such that for any $R < C$ and any $\varepsilon > 0$, there exist m and n and an encoding (error-correcting code mapping) $\mathcal{E}$ of any m-bit strings into n bit strings, where $\frac{m}{n} > R$ such that every m bit string x can be received, after decoding $\mathcal{D}$, with error probability less than ε at the receiving end of the channel, from its output y. In short:

$$C = \limsup\{c \,|\, \forall_{R<c}\forall_{\varepsilon>0}\exists_{\mathcal{E},\mathcal{D}}\exists_{m,n}(\frac{m}{n} > R \wedge Pr(\mathcal{D}(y) \neq x) < \varepsilon)\}.$$

The above result about the capacity of the noisy classical channel is called Shannon's bound for the capacity of noisy channels.

The basic definition of the capacity $QC(\mathcal{N})$ of quantum channels parallel that for classical channels. Namely, $QC(\mathcal{N})$ is the largest number R such that for any $\varepsilon > 0$ there exists n such that any state of n qubits can be transmitted with fidelity at least $1 - \varepsilon$ by fewer that $\frac{n}{R}$ (forward) uses of the channel $\mathcal{N}$. n qubits are encoded into m channel inputs, then transmitted and, finally decoded.

The basic definition of the capacity of a quantum noisy channel $\mathcal{N}$ can be concisely expressed as follows (Lloyd, 1997, Barnum et al. 1997, 1997a, Bennett, 1998):

$$QC(\mathcal{N}) = \lim_{\varepsilon\to 0}\limsup_{n\to\infty}\{\frac{n}{m} \,|\, \exists_{m,\mathcal{E},\mathcal{D}}\forall_{\psi\in H_{2^n}}(\langle\psi|\mathcal{D}\mathcal{N}^{\otimes m}\mathcal{E}(|\psi\rangle\langle\psi|)|\psi\rangle > 1-\varepsilon)\}.$$

where $\mathcal{E}$ is an encoding superoperator from n qubits to m quantum channels, see Figure 8.4, and $\mathcal{D}$ is a decoding superoperator from m channel outputs to n qubits.

A communication through a quantum channel can be accompanied by a communication through a classical channel. The above definition of the capacity of a quantum channel refers only to the transmission unassisted by classical communications. Two other measures of quantum channel capacity, $QC_1(\mathcal{N})$ and $QC_2(\mathcal{N})$, are defined as asymptotic quantum capacities of the quantum channel $\mathcal{N}$ assisted by one-way classical communication (from the sender to the receiver) or by two-way classical communication, between the sender and the receiver (see Figure 8.5), where the superoperator $\mathcal{E}$ stands for the encoding and $\mathcal{D}$ for the decoding processes. It has been shown, by Bennett et al. (1996a), that one-way classical communication does not help and therefore $QC_1(\mathcal{N}) = QC(\mathcal{N})$ for any quantum channel. On the other hand, two-way classical communication can increase the capacity of some noisy quantum channels.

The basic relations between these capacities can be summarized as follows (see Bennett et al. 1996a and Bennett, 1998b):

- $QC(\mathcal{N}) \leq C(\mathcal{N})$ for all quantum channels $\mathcal{N}$ and $QC(\mathcal{N}) < C(\theta)$ for some $\mathcal{N}$.
- $QC(\mathcal{N}) \leq QC_2(\mathcal{N})$ for all quantum channels $\mathcal{N}$ and $QC(\mathcal{N}) < QC_2(\mathcal{N})$ for some $\mathcal{N}$.

[5]Personal communication: I. Chuang $\rightarrow$ A. Ambainis, A. Nzyak, A. Ta Shma and U. Vazirani $\rightarrow$ the author.

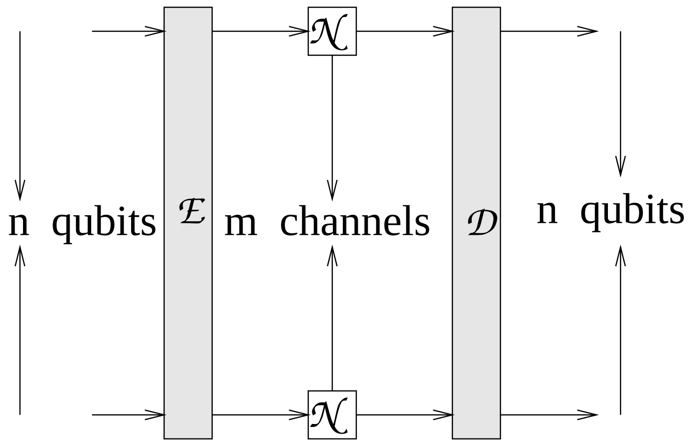

Figure 8.4: Channel transmission scheme

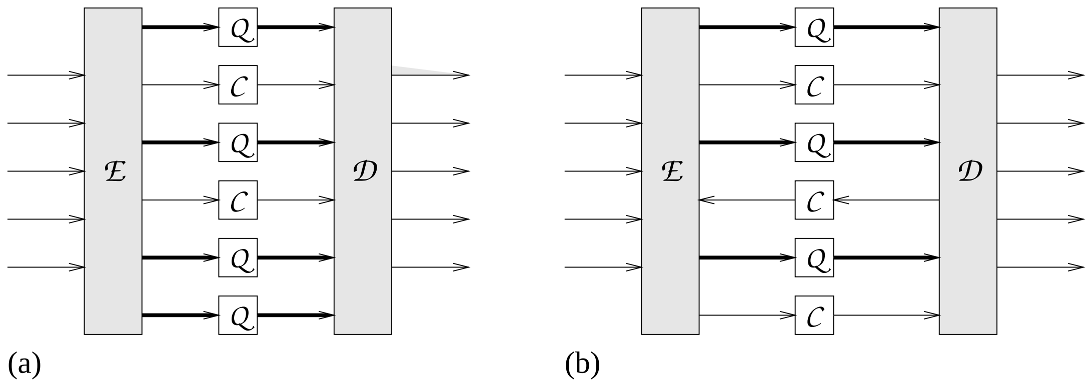

Figure 8.5: Communication through a quantum channel assisted by one-way and two-way classical communication. $\mathcal{Q}$ denotes a quantum channel and $\mathcal{C}$ a classical channel

- $QC_2(\mathcal{N}) \leq C(\mathcal{N})$ for all known quantum channels $\mathcal{N}$ and $QC_2(\mathcal{N}) < C(\mathcal{N})$ for some.

The inequality $QC(\mathcal{N}) \leq QC_2(\mathcal{N})$, follows from the definition. From the fact that orthogonal quantum states can be used to transmit bits, it follows that $QC(\mathcal{N}) \leq C(\mathcal{N})$. No channel $\mathcal{N}$ is known for which $QC_2(\mathcal{N}) > C(\mathcal{N})$.

8.2.6 Capacities of erasure and depolarizing channels

Quantum capacities are known well only for few types of quantum channels. With two such cases we deal now: the erasure channel and the depolarizing channel. For most of other channels at most some upper and lower bounds are known.

The quantum **erasure channel** (QECH, see Grassl et al. 1996), is the channel in which every qubit sent through the channel either gets through fine, with probability $1 - p$, or it gets totally random, with probability p and, in addition, it is known to the receiver, which case has occurred.

In the case of 2EPP, if Alice sends n EPR pairs through the erasure channel, then pn of them get destroyed and $(1-p)n$ will be intact. In addition, Bob knows which pair remain intact. He informs Alice and they discard the remaining pairs. The resulting rate is $1-p$ and clearly one cannot do better. Hence $QC_2(QECH) = 1-p$.

No cloning argument can be used to show (see Bennett et al. 1997a), that the erasure channel has the capacity at most $1-2p$ and therefore 0 if $p \geq \frac{1}{2}$. Hence $QC(QECH) = \max\{0, 1-2p\}$.

The quantum **depolarizing channel** (QDCH) is a very natural channel. Each qubit is left intact with probability $1-p$ and it is affected by each of the error operators $\sigma_x, \sigma_y, \sigma_z$ with the same probability $\frac{p}{3}$. The receiver does not know on which qubits randomization has been performed.

Another definition of the erasure channel is that the incoming qubit is, with probability p, replaced by an "erasure state" $|2\rangle$, orthogonal to the states $|0\rangle$ and $|1\rangle$ (thereby erasing the qubit), and informing the receiver that it has been erased.

For the depolarizing channel $\mathcal{D}$ the following bounds are known (see Bennett et al. (1996).

if $p < 0.25408$ **then** $QC(QDCH) > 0, QC_2(QDCH) > 0, C(QDCH) > 0$;

if $\frac{1}{3} < p < \frac{2}{3}$ **then** $QC_1(QDCH) = 0, QC_2(\mathcal{D}) > 0,\ C(QDCH) > 0$;

if $\frac{2}{3} \leq p \leq 1$ **then** : $QC_2(QDCH) = 0,\ C(QDCH) > 0$.

8.3 Quantum Entanglement

Quantum entanglement can be seen as the purest form of quantum information (see Bennett, 1998a). Quantitative and qualitative theory of entanglement as well as methods of creating, preserving, transmitting, and manipulating entanglement are therefore of great interest and importance for quantum information processing.

Entanglement is also an important communication resource. For example, for teleportation and dense coding. For that we mostly need pure entangled states (otherwise the EPR-channel gets noisy) and maximally entangled states—as for superdense coding.

Unfortunately, pure and maximally entangled states are not easy to create and to preserve. In practice we have therefore mainly noisy and week entanglement. An important problem is therefore whether there is an efficient way to purify noisy entanglement or, in other words, to distill pure entanglement out of the noisy one. Fortunately, the answer is positive.

8.3.1 Transformation and the partial order of entangled states.

The very basic questions concerning the existence and methods of transformations of one entangled state into another by local operations and classical communication are already quite well understood. An elegant necessary and sufficient condition for the existence of such a transformation is known, in the terms of the majorization of vectors of eigenvalues of the corresponding bipartite states, due to Nielsen (1998). The result implies the existence of different (incommensurate) types of entanglement. In order to formulate the main result the concept of "majorization" between vectors needs to be defined.

Definition 8.3.1 *A real-valued vector $X = (x_1, \ldots, x_d)$ is* **majorized** *by a real-valued vector $Y = (y_1, \ldots, y_d)$, notation $X \prec Y$, if, for $1 \leq k \leq d$, $\sum_{j=1}^{k} x_j^{\downarrow} \leq \sum_{j=1}^{k} y_j^{\downarrow}$ with equality holding for $k = d$, where $x_i^{\downarrow}$ denotes the ith largest element among $x_1, \ldots, x_d$.*

For a pure state $|\phi\rangle$ of a bipartite Hilbert space $H_A \otimes H_B$, let λ_ϕ denotes the vector of eigenvalues of the H_A-state $\rho_\phi = Tr_{H_B}(|\phi\rangle\langle\phi|)$.

Theorem 8.3.2 *(Nielsen (1998) If $|\psi\rangle$ and $|\phi\rangle$ are pure states of $H_A \otimes H_B$, then $|\psi\rangle$ can be transformed to $|\phi\rangle$ using local operations and classical communication if and only if $\lambda_\psi \prec \lambda_\phi$.*

Two states in $H_A \otimes H_b$ are called **incommensurate** if none can be obtained from another by local operations and classical communication.

Example 8.3.3 *Incommensurate are, for example, states*

$$|\psi\rangle = \sqrt{\frac{1}{2}}|11\rangle + \sqrt{\frac{2}{5}}|22\rangle + \sqrt{\frac{1}{10}}|33\rangle \text{ and } |\phi\rangle = \sqrt{\frac{3}{5}}|11\rangle + \sqrt{\frac{1}{5}}|22\rangle + \sqrt{\frac{1}{5}}|33\rangle$$

The proof of Theorem 8.3.2 also yields a method how to construct $|\psi\rangle$ out of $|\phi\rangle$ using at most $d-1$ bits of classical communication.

Open problem 8.3.4

1. *Find a necessary and sufficient condition that a mixed state ρ_1 can be transformed into a mixed state ρ_2 using only local operations and classical communication.*
2. *Find a necessary and sufficient condition to transform one three (or more) party entangled state into another one by local operations and classical communications of all involved parties.*

8.3.2 Entanglement purification/distillation

Starting with the pioneering paper by Bennett et al. (1996b), several purification protocols have been developed.

The idea behind the basic purification step is simple. Alice generates two pairs of entangled qubits (a_1, b_1) and (a_2, b_2) and sends the second of each pair to Bob through a noisy channel. Alice performs XOR on a_1 and a_2 and Bob on b_1 and b_2 and both measure target qubits—see Figure 8.6— a_2 and b_2, and communicate to each other the results of their

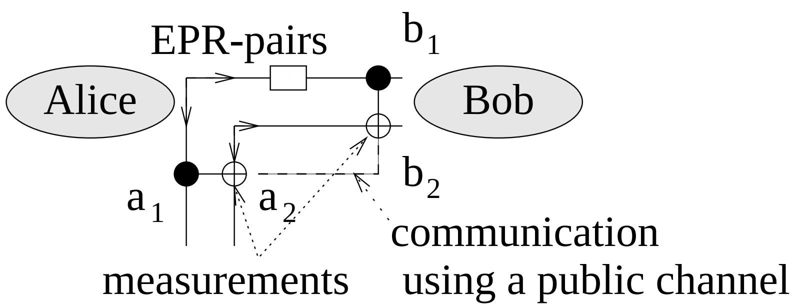

Figure 8.6: Basic step of the entanglement purification

measurements. If they agree they consider the first entangled pair (a_1, b_1) as good. If not they discard both pairs.

The overall purification process then goes as follows (see Figure 8.7). Alice generates pairs (a_1, b_1), (a_2, b_2), ... and sends the second particle of each pair to Bob (through a noisy channel). To each pair of particles (a_{2i-1}, b_{2i-1}), (a_{2i}, b_{2i}) they perform the above purification step at the end of which they either keep the pair (a_{2i-1}, b_{2i-1}) or discard both pairs.

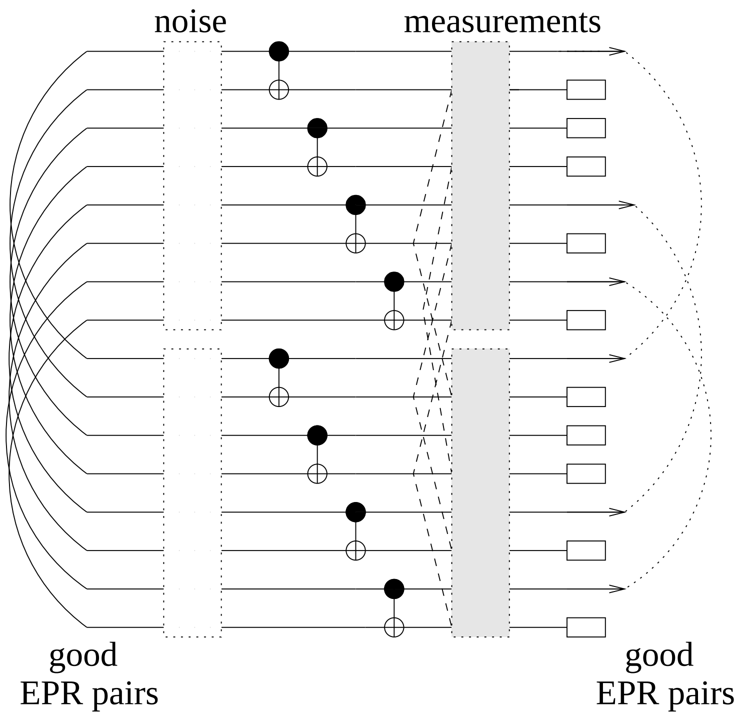

Figure 8.7: Entanglement purification with 2-way communication-80%

The input for the purification process do not have to be perfectly entangled pairs and such pairs can be produced by a source and get through the noisy channels on both ways to Alice and to Bob.

The above purification protocol can be modified and generalized by allowing both parties to perform first various local operations (superoperators) and to use some other post-selection procedures. Efficiency of purification is the main issue.

Example 8.3.5 *(Bennett et al. 1996b) Let us assume that a purification procedure is applied to a mixed state* ρ *and the aim is to obtain the singleton* $|\Psi^-\rangle$*. The purity of* ρ*, with respect to the goal* $|\Psi^-\rangle$*, can be expressed by the fidelity* $F = \langle\Psi^-|\rho|\Psi^-\rangle$*. In order to explore the impact of a given purification procedure instead of* ρ *we can consider the so-called* **Werner state**

$$W_F = F|\Psi^-\rangle\langle\Psi^-| + \frac{1-F}{3}(|\Psi^+\rangle\langle\Psi^+| + |\Phi^+\rangle\langle\Phi^+| + |\Phi^-\rangle\langle\Phi^-|), \tag{8.8}$$

which has the same fidelity as ρ*.*

In the simplified purification protocol given by Bennett et al. (1996b), at first Hadamard rotations are applied locally, on each of the two pairs and after the BXOR operation such rotations are applied again. They showed that in such a case new mixed state has with probability $\frac{1}{4}$ a better fidelity F', with respect to $|\Psi^-\rangle$, namely

$$F' = \frac{F^2 + \frac{1}{9}(1-F)^2}{F^2 + \frac{2}{3}F(1-F) + \frac{5}{9}(1-F)^2} > F.$$

Purification with one-way communication

The above model of purification is also called **purification with 2-way communication**, or **two-way entanglement purification protocol** (2EPP), because both parties have to communicate classically to reach the outcome.

Purification protocols with one-way classical communication (1EPP) have been developed by Bennett et al. (1996a). They have also proved that 2EPP are more powerful than 1EPP. For example, 1EPP cannot purify Werner states W_F if $F \leq \frac{5}{8}$, but there is a positive lower bound (0.00457) on the yield when $W_{\frac{5}{8}}$ is purified using 2EPP.

Entanglement purification and error correction

Entanglement purification protocols can be seen also as providing, in addition to QECC, another way of protecting quantum states from their interaction with the environment. Not only that, a close relation between quantum error-correcting codes and 1EPP can be shown.

The scenario we deal with is that Alice prepares a number of EPR pairs and sends one member of each pair to Bob. In a 1EPP Bob can receive quantum and classical information from Alice but cannot communicate to Alice.

At first we show how to make out of an quantum error correcting code a 1EPP. Alice generates EPR-pairs and encodes those qubits she wants to sent to Bob and sends them, encoded, to Bob. He performs error correction and decoding. The encoded qubits retain their entanglement with Alice's qubits. This way distillation works on 100%.

Secondly, let us have an 1EPP capable to distill k good EPR-pairs out of n noisy pairs. We show how to use such 1EPP to make a QECC. Alice will be the encoder and Bob the decoder.

Alice creates n EPR-pairs, sends them (the corresponding parts) to Bob and performs her part of 1EPP. In addition, she takes k qubits she wants to protect using her half of what will be the k good pairs and sends the classical information about the measurement results to Bob. Afterwards Bob completes the purification protocol to get k good pairs. Since they are good pairs he can complete the teleportation protocol. The whole process can be seen as encoding k qubits by n qubits.

8.3.3 Entanglement concentration and dilution

Entanglement purification requires classical public communication. Another technique, with similar aim, that requires no public communication is the so-called entanglement.

Entanglement concentration is a method to concentrate entanglement of a large number of partially entangled states into a smaller number of maximally entangled states. On the other hand, **entanglement dilution** is a method to dilute a small set of maximally entangled states into a larger set of partially entangled states (see Bennett et al. 1996).

8.3.4 Quantifying entanglement

As discussed in Section 2.2.2 the amount of entanglement $E(\phi)$ of a pure state $|\phi\rangle$ of a bipartite system $A \oplus B$ is defined, for $\rho = |\phi\rangle\langle\phi|$, through quantum entropy by

$$E(\phi) = -Tr\rho_A \lg \rho_A = -Tr\rho_B \lg \rho_B.$$

Using the Schmidt decomposition theorem, page 372, $|\phi\rangle$ can be expressed in the form

$$|\phi\rangle = \sum_{i=1}^{d} c_i |\alpha_i\rangle|\beta_i\rangle,$$

where c_i are positive and $\{\alpha_i\}_{i=1}^d$, $\{\beta_i\}_{i=1}^d$ are orthonormal sets of states in A and B. In such a case $\rho_A = T_B|\phi\rangle\langle\phi| = \sum_{i=1}^d c_i|\alpha_i\rangle\langle\alpha_i|$, and $E(\phi) = \sum_{i=1}^d -c_i^2 \lg c_i^2$.

A strong justification of the above definition of the entanglement of pure states is the following result (see Bennett et al. 1996b). Consider n entangled pairs of particles, each in the state $|\phi\rangle$. Let Alice hold one particle of each pair and Bob, spatially separated, holds the other. If $|\phi\rangle$ has E ebits of entanglement, then n pairs can be reversibly converted, by local operations and classical communications, into m singletons, where $\frac{m}{n}$ approaches E for large n and the fidelity approaches 1.

Example 8.3.6 *(Bennett et al. 1996a) If a state $|\phi\rangle$ has representation $|\phi\rangle = \sum_{i=1}^4 \alpha_i|m_i\rangle$, where $|m_i\rangle$ are states of the magic basis, then $E(\phi) = H(\frac{1}{2} + \frac{1}{2}\sqrt{1 - \sum_{i=1}^4 |\alpha_i|^2})$.*

Several measures of entanglement have been defined and investigated for mixed states and density matrices They are based on clear physical ideas. Each of them can be seen as a fundamental concept of quantum information theory.

Entanglement of formation (creation), E_f, (Bennett et al. 1996a), is defined for a density matrix ρ by

$$E_f(\rho) = \min\{\sum_i p_i E(\psi_i) \,|\, \rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|\},$$

where ρ is the density matrix for mixed states$\bigoplus_{i=1}^n (p_i, \psi_i)$. In other words, entanglement of formation, $E_f(\rho)$, is the least expected entanglement of any ensemble of pure states, with ρ as the density matrix.

Entanglement of formation is known exactly for many states, including all two-qubit states (see Hill and Wooters, 1997), having no more than two non-zero eigenvalues. In addition, Uhlmann (1997) has shown how to compute entanglement of formation for those states where E_f is not known to have a closed form.

Open problem 8.3.7 *Is entanglement of formation additive?*

Entanglement of distillation (distillable entanglement), E_d (see Bennett et al. 1996a), is intuitively, the maximum asymptotic yield of singleton states that can be produced from the given mixed state by local operations and classical communication.

Distillable entanglement is further classified by the type of classical communication allowed (no (E_{d0}), one-way (E_{d1}), and (E_{d2}) two-way classical communication) and depends also on the type of the local operations allowed.

Another way to define, informally, entanglement of distillation of a mixed state ρ is the maximum over all allowable protocols of the expected rate at which singletons can be obtained from a sequence of states ρ. For a rigorous treatment of the entanglement of distillation see Rains (1998).

Entanglement of assistance, E_a, (DiVincenzo et al. 1998), is in a sense dual to the entanglement of formation

$$E_(\rho) = \max\{\sum_i p_i E(\psi_i) \,|\, \rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|\},$$

Entanglement of relative-entropy, E_r**,** (Vedral and Plenio, 1997) defined by

$$E_r(\rho) = \min_{\rho' \in \mathcal{D}} (Tr\rho(\lg\rho - \lg\rho')),$$

where $\mathcal{D}$ denotes the set of unentangled mixed states.

For the entanglement of relative-entropy the following relation holds:

$$E_r(\rho_1 \otimes \rho_2) = E_r(\rho_1)E_r(\rho_2)$$

for all density matrices ρ_1 and ρ_2.

The following relations hold among the above entanglement measures (see Vedral and Plenio, 1997, and Bennett et al. 1996a):

$$E_{d0} \le E_{d1} \le E_{d2} \le E_r \le E_f.$$

All these entanglement measures are reduced to the entropy of entanglement for pure states.

By Vedral and Plenio (1997), a measure E of entanglement should satisfy the following conditions; (a) $E(\rho) = 0$ if and only if the state ρ is separable; (b) local unitary operations leave $E(\rho)$ invariant; (c) the expected entanglement cannot increase under local (general) measurements and classical communication.

8.3.5 Bound entanglement

A surprising discovery about the existence of two qualitatively different types of entanglement of mixed states was made by Horodecki et al. (1998). They showed that there is a

- "free", distillable entanglement;

and a

- "bound", enatnglement that is not distillable.

Horodecki et al. (1997) showed that any entangled mixed state of two qubits can be distilled to get the singleton. However, this is not true in general for entangled states. For example, there are entangled mixed states of two qutrits that cannot be distilled (and therefore they cannot be used for a quantum communication).

Example 8.3.8 *(Horodecki et al. 1998a) Two qutrit mixed states*

$$\sigma_+ = \frac{1}{3}(|0\rangle|1\rangle\langle 0|\langle 1| + |1\rangle|2\rangle\langle 1|\langle 2| + |2\rangle|0\rangle\langle 2|\langle 0|), \sigma_- = \frac{1}{3}(|1\rangle|0\rangle\langle 1|\langle 0| + |2\rangle|1\rangle\langle 2|\langle 1| + |0\rangle|2\rangle\langle 0|\langle 2|)$$

are separable. However, the state

$$\sigma_\alpha = \frac{2}{7}|\Psi^+\rangle\langle\Psi^+| + \frac{\alpha}{7}\sigma_+ + \frac{5-\alpha}{7}\sigma_-,$$

where $|\Psi^+\rangle = \frac{1}{\sqrt{3}}(|0\rangle|0\rangle + |1\rangle|1\rangle + |2\rangle|2\rangle)$, *is*

$$\begin{array}{rll} \textit{separable,} & \textit{if} & 2 \leq \alpha \leq 3 \\ \textit{bound entangled,} & \textit{if} & 3 < \alpha \leq 4 \\ \textit{free entangled,} & \textit{if} & 4 > \alpha \leq 5 \end{array}$$

A method to construct states with bound entanglement was developed by DiVincenzo et al. 1998a).

8.4 Quantum information processing principles and primitives

The development of concepts, methods, and results of importance for quantum information processing has to be accompanied by a study of the mosr fundamental principles and primitives of quantum information theory.

8.4.1 Search for quantum information principles

The thesis that information is physical and that the role of information in physics is analogous to that of energy in thermodynamics leads naturally to the search for informati on processing principles and laws. For example for principles and lawsındexlaw analogous to those in thermodynamics. It is only natural t hat quantum entanglement is expected to play the key role in suc h principles and laws. One such emerging principles seems to be **no-increasing of entanglement principle** (under local operations and communic ation).

Informally, this principle says that entanglement cannot inc rease under local operations and classical communication. More formally, it says that if E is a "proper" measure of entanglement and an input state ρ is transformed by quantum operations into the mixed state $\{(p_i, \rho_i)\}_{i=1}^k$, then $\sum_{i=1}^k p_i E(\rho_i) \leq E(\rho)$.

This principle has been proven to hold for entanglement of f ormation (see Bennett et al. 1997a). A form of this principle has been proven to be equivalent with no-cloning theorem (see M. Horodecki and R. Horodecki, 1997). They also formulated the following implications of the principle: "The entanglement of the com pound system does not change under unitary processes in one of the subsystems.

8.4.2 Quantum information processing primitives

Identification and comparison of quantum information processing primitives as well as the study of reducibilities among them is one of the most fundamental goals of quantum information processing theory.

Two such primitives are, for example, qubits and ebits. Let us therefore summarize some of their basic relations, and properties.

Qubits are, from a point of view of transmissions, a directed channel resource (Bennett, 1998a). They can be sent from a sender to a receiver in a particular direction. On the other hand ebits are undirected resource shared between a sender and a receiver.

Ebits are a weaker communication resource than qubits in the following sense. One can create one ebit by transmitting one qubit. On the other hand, sharing any number of ebits does not suffice by itself to transmit a qubit in either direction.

Both qubits and ebits, as communication primitives, have to be considered in two settings: in communication through a noiseless channel and through a noisy channel,

A detailed analysis of quantum information processing primitives and reducibilities has been done by Bennett (1998).